Proceedings

of the 6th International Symposium on the Acquisition of Second Language Speech, New Sounds 2010, Poznań, Poland, 1-3 May 2010.

Katarzyna Dziubalska-Kołaczyk,

Magdalena Wrembel and Małgorzata Kul (eds).

ISBN: 978-83-928167-9-9

Form-focused instruction in L2 pronunciation pedagogy: The effect of negotiation of form in a Japanese classroom

Hideki Abe

Tsuruoka National College of Technology/
Graduate School of Nagoya Gakuin University
habe@tsuruoka-nct.ac.jp

ABSTRACT

This study attempts to examine the effect of Form-focused Instruction (FFI) on the learning of connected speech in L2 pronunciation pedagogy. FFI treatment was designed to encourage learners to notice the error, negotiate its phonological form, and correct the error by themselves. Results of the treatment were compared with the non FFI treatment in this study involving sixty elementary to low intermediate level students. Progress of performance was measured with a pre-test and two post-tests, providing two major findings: (1) FFI had positive effects on the learning of English connected speech; (2) the subjects in the FFI treatment improved more significantly than the subjects of non-FFI treatment.

Keywords: connected speech, EFL, FFI, NoF, pronunciation pedagogy

1. INTRODUCTION

Recent studies in second language acquisition (SLA) have suggested that instruction taking psycholinguistic and cognitive factors into consideration is highly beneficial to second language teaching and learning (Doughty and Williams, 1998; Fotos and Nassaji, 2007; Long, 1991; Long and Robinson, 1998; Muranoi, 2006). To examine this issue in Instructed SLA, an approach called “focus on form” has been proposed. According to Long (1991), focus on form is defined as a type of instruction drawing “students’ attention to linguistic elements as they arise incidentally in lessons whose overriding focus is on meaning, or communication (p.46). Most studies investigating focus on form, however, have grammar instruction as their primary focus (Long and Robinson, 1998; Muranoi, 2006 to mention a few), while classroom pronunciation research has received only a little attention.

Although L2 pronunciation research has not been directly concerned with focus on form, several researchers appeared to claim that classroom instruction should involve systematic treatments to draw L2 learners’ attention to phonetic forms to develop a well-balanced phonological competence (Chan, 2006; Sicola, 2008; Tanabe and Koyama, 1998 to mention a few). To obtain significant data for this issue, the present study examined how a form-focussed instructional technique, more precisely the Negotiation of Form (NoF) in which a linguistic error is made explicit and ongoing negotiation (or interaction) help learners notice the error(s) and correct the error(s) by themselves (Lyster and Ranta, 1997). NoF was incorporated into a regular lesson to encourage learners to notice the gap in the target phonetic form and restructure their interlanguage phonology, because learners occasionally find it difficult to phonetic deviation in dyads performance in the implicit flow of communication.

Due to the nature of the research questions, however, the analyses were limited to Japanese college students and the acquisition of connected speech, namely, *rhythm, linking, assimilation* and *elision*. The connected speech was chosen because these aspects of phonology are considered to be critical communicative competence, and Japanese EFL learners tend to have difficultly in learning these features (Kohmoto, 1982; Watanabe, 1994). The following two major questions were investigated:

1. Does FFI, in which a teacher provides explicit instruction through NoF, affect EFL learners’ restructuring of their interlanguage phonology?
2. Do two types of instruction, differing in the manner of instruction have different effects on EFL learners’ acquisition of the L2 English connected speech?

2. METHOD

The present study was conducted in a regular classroom setting in Japan, and the participants were third-year students of high-school level enrolled in their intact EFL classes at a technical college. Their English levels at school were equivalent: low to intermediate. In this classroom-based study, the effects of FFI and the control treatment were compared quantitatively. The subjects in FFI received a negotiation of form treatment, which was comprised of noticing and interaction. In this treatment, the subjects listened to two different versions of oral readings of the same material (one spoken in a natural speed and one without connected speech processes). Then the teacher asked the subjects to compare the differences between the two in pairs. After pair-discussion of noticing, they shared their findings in class. Finally the treatment ended with a chorus reading. The control group received explanation of English connected speech and listen-and repeat exercises. Both pre-test and two post-tests were examined and scored by the investigator and a native speaker of American English, both of whom received phonetic and linguistic training as MA graduate students. A one-way analysis of variance (ANOVA) was run to determine if there were any statistically significant differences among the two groups' mean scores on the pre-test measuring ability to use English connected speech. No significant difference among the participants was revealed ($F(2/87)=3.10$, $p>.05$, ns). The pre- and post-test consisted of 20 questions, including the targeted prosodic features. Examples of test sentences are listed in Table 1,

Table 1: Examples of test sentences.

Aspect	Example
Rhythm	**Let's** in**vite** them to the **par**ty.
Linking	She ha**d a** sad expression.
Assimilation	They married las**t y**ear.
Deletion	He left jus**t n**ow.

The whole period of evaluation lasted over a period of three months; a period of five weeks for all eight treatments and one month between the first and second post tests.

3. RESULTS

3.1. Instructional effect

Table 2 indicates the means and standard deviations of the pre- and two post-tests. Figures 1 and 2 below graphically display the total scores respectively.

Table 2: Descriptive Statistics for the Total Scores of Perception and Production.

[Highest possible score = 20]

	Experimental Group (FFI)					
	Perception			Production		
	PR	P1	P2	PR	P1	P2
n	30	30	30	30	30	30
M	4.6	9.63	12.57	7.47	13.2	12.57
SD	2.2	2.76	3.59	2.56	2.76	3.6
	Control Group (NFI)					
	Perception			Production		
	PR	P1	P2	PR	P1	P2
n	30	30	30	30	30	30
M	3.67	7.36	9.27	6.5	8.07	6.7
SD	2.07	2.59	2.39	2.2	2.44	2.16

Note. PR= Pre-test, P1= Post-test 1, P2= Post-test 2

Figure 1: Improvement in Perception Abilities.

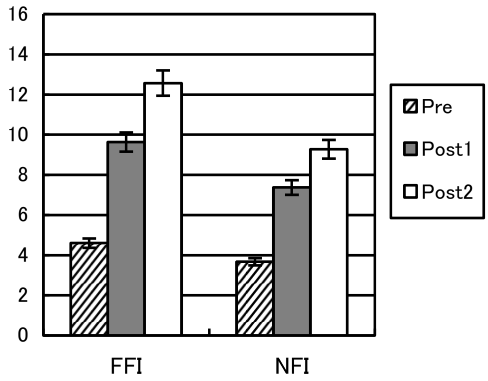
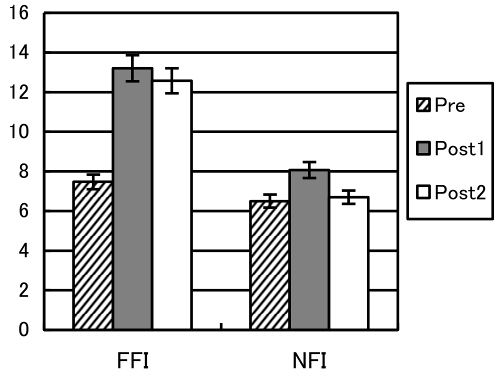


Figure 2: Improvement in Production Abilities.



As shown in table 3, results of the repeated measures of ANOVA for the perception and production scores revealed a significant main effect for instruction. The results, especially those from between-group comparisons, indicated that the FFI group receiving explicit instruction plus interaction through NoF performed significantly better than the NFI (Non Form–focused Instruction) group ($F(2,87)=3.10$, $p<.01$, $F(2,87)=3.10$, $p<.001$, respectively). Therefore, FFI positively affected the learning of connected speech. The level of improvement is indicated by the number of asterisks in the tables: the significance level of $p < 0.05$ = *, $p < 0.01$ = **, $p < 0.01$ = ***, and (ns) = ns.

Table 3: Comparisons between experimental groups.

Type of Test	Test Phase	p-value	Comparison
Perception	PR	p ＞ .05 (ns)	NFI ＜ FFI
	P1	p ＜ .01**	NFI ＜ FFI
	P2	p ＜ .001***	NFI ＜ FFI
Production	PR	p ＞ .05 (ns)	NFI ＜ FFI
	P1	p ＜ .001***	NFI ＜ FFI
	P2	p ＜ .001***	NFI ＜ FFI

Note. PR= Pre-test, P1= Post-test 1, P2= Post-test 2

Effect sizes for the between-test comparisons were calculated to examine the practical significance of between-test differences. First, Between-test comparisons for the FFI group revealed (1) that the FFI group did significantly better on the two post-tests for perception and production alike, and (2) the gain measured in the first post-test lasted for post-test periods. Second, Between-test comparisons for the NFI group indicated (1) that there was significant difference at a practical level for perception data, and (2) production data had just a small effect size.

Table 4: Summary of the results of between-test comparisons.

Perception		Production	
FFI		FFI	
	PR < P1 (.79) *		PR < P1 (.88) *
	PR < P2 (.64) *		PR < P2 (.81) *
NFI		NFI	
	PR < P1 (.74) *		PR < P1 (.37) *
	PR < P2 (.84) *		PR < P2 (.10)

Note. ES = Effect Sizes

* $p < 0.05$

3.2. Effects on perception and production

The performance for each aspect of phonology was then compared to investigate which aspects are most affected by explicit instruction. The results of the repeated measures of ANOVA in Table 5 suggest that for the perception data there was significant difference among the three groups for each aspect of phonology. The FFI group exhibited significant improvement in all four aspects in the first and the delayed post-test. The general pattern of the FFI group outperforming the control group did not change in the second post-tests, suggesting that the gains in the first post-test lasted for one month.

Table 5: Perception **c**omparison between groups according to aspect & test type.

Aspects	**Perception**		
	Pre-test	Post-test 1	Post-test 2
Rhythm	< .05*	< .05*	< .001***
Linking	> .05 (ns)	< .05*	> .05 (ns)
Assimilation	> .05 (ns)	< .01*	< .001***
Elision	> .05 (ns)	> .05 (ns)	< .05*

The results in the production data show that although there were no significant differences in the pre-test scores between groups regarding *rhythm, linking, assimilation*, and *elision*, the effect of NoF was robust and consistent (See Table 6). However, as indicated in Figure 2, there was a decrease between post-test 1 and post-test 2, though the participants did not fall back to their pre-test level. This implication is discussed in the next section.

Table 6: Production comparison between groups according to aspect & test type.

Aspects	**Production**		
	Pre-test	Post-test 1	Post-test 2
Rhythm	> .05 (ns)	< .001***	< .001***
Linking	> .05 (ns)	< .001***	< .001***
Assimilation	> .05 (ns)	< .001***	< .001***
Elision	> .05 (ns)	< .001***	< .001***

4. DISCUSSION

The results can be summarized as follows: (1) NoF had positive impacts on L2 learning of English connected speech, which lasted for one month; (2) NoF treatment had greater effects on learner performance than the control treatment on all aspects of L2 phonology.

This present study has investigated methodological difference and different effects they have on Japanese EFL learners' restructuring of their interlanguage phonology. Results indicated that the FFI group outperformed the NFI group on all tested items of post-tests. This suggests that NoF with teachers and students was more beneficial for L2 learning of connected speech than the NIF group, where NoF treatment was not provided.

The study has further examined whether the effect of instruction holds over the post-test period, if FFI indeed has some effect on learners' restructuring of their interlanguage phonology. This finding leads us to assume that instruction that appropriately incorporates NoF treatments can have a lasting positive effect on L2 acquisition. More specifically, the results of this study suggest that lasting instructional effects can be obtained through providing learners with opportunities to think of the target form through negotiation task. This, in consequence, had an effect which did not decline in the delayed post-test. However, results also indicated the possibility that gains in the treatment might slightly decrease without constant practice in production.

Regarding differential effects on learner performance of four aspects of English connected speech, results show that subjects in the FFI group improved significantly on all aspects more than NFI. With the NoF treatment, the teacher could promote activation of such cognitive processes as noticing, cognitive comparison as an option of pronunciation teaching in the EFL classroom setting.

5. CONCLUSION

Several factors including methodological limitations may have affected the argument and outcome, however, this study to some extent has been successful in exploring the relationship between an instructional approach and phonological acquisition, and in proposing that the FFI in pronunciation pedagogy could be more effective than the traditional approach. Further research should also consider whether focus-on-form treatments involving both implicit and explicit formal instruction can help learners improve their performance.

6. REFERENCES

Chan, M. 2006. *Pronunciation instruction, learner awareness, and development.* PhD dissertation, OISE.

Doughty, C. 2003 Instructed SLA: Constraints, compensation, and enhancement. In: Doughty, C., Long, M. (eds.), 2003. *The handbook of second language acquisition.* Oxford: Blackwell. 256-310.

Doughty, C., Williams, J. (eds.). 1998. *Focus on form in classroom second language acquisition.* Cambridge: Cambridge University Press.

Fotos, S., Nassaji, H. 2007. *Form-focused instruction and teacher education: Studies in honour of Rod Ellis.* Oxford: Oxford University Press.

Kohmoto, S. 1982. *New English phonology: A contrastive study of English and Japanese pronunciation.* Tokyo: Nan'un-do.

Long, M. 1991. Focus on Form: A design feature in language teaching methodology. In: de Bot, K., Ginsberg, R., Kramsch, C. (eds.), *Foreign language research in cross-cultural perspective*. Amsterdam: John Benjamin. 39-52.

Long, M., Robinson, P. 1998. Focus on form and practice. In: Doughty, C., Williams, J. (eds.). 15-41.

Lyster, R., Ranta, L. 1997. Corrective feedback and learner uptake: negotiation of form in communicative classrooms. *Studies in Second Language Acquisition*, 19(2). 37-66.

Muranoi, H. 2006. Dainigengo Shuutoku kenkyuu kara mita koukatekina Eigo gakusyuu kenkyuu (SLA research and second language teaching and learning). Tokyo: Taishuu-kan.

Sicola, L. 2008. *No, they won't 'just sound like each other': NNS-NNS negotiated interaction and attention to phonological form on targeted L2 pronunciation tasks.* Bern: Peter Lang.

Tanabe, Y., Koyama, K. 1998. An experimental study on teaching English pronunciation: From input to intake. *English Phonetics* 2. 353-376.

Watanabe, K. 1994. *Eigo no rizumu ·intoneeshon nosShidou* (Teaching English rhythm and intonation). Tokyo: Taishuu-kan.

Wrembel, M. 2005. Metacompetence-oriented model of phonological acquisition: Implications for the teaching and learning of second language pronunciation. *Phonetic Teaching and Learning Conference 2005.* <http://www.phon. ucl. ac.uk/ ptlc 2005. html/> Last access 10 Sep 2007.

Production and Perception of SSBE Vowels by Syrian Arabic Speakers

Rana Alhussein Almbark

York University
raa502@york.ac.uk

Usually, studies of second language acquisition focus on examining the productions of advanced learners who have experience in the L2 country. The present study examined the productions of L2 learners of English who had learnt English for at least 10 years, and had no experience in an English speaking country . The main interest of the present study was to examine the predictions of PAM in the productions of three Standard Southern British English (SSBE) vocalic contrasts /ɪ/- /ɛ/, /ɒ/- /ʌ/, and /æ/- /ɑː/, by Syrian Arabic (SA) learners of English compared to native SSBE speakers. Five SA and four SSBE female speakers were recorded producing the target vowels within /hVd/ context, SA speakers were also recorded producing comparable Arabic data. Vowel duration and the first two formants of ten automatically inserted points were extracted. The results showed that PAM predictions stand for some of the production patterns of SA learners of English. Also, accurate perception of L2 contrasts do not necessarily lead to accurate production of these contrasts.

Keywords: SSBE, SA, Vowels, Production, PAM.

1. INTRODUCTION

Second language learners, particularly in their early stages, tend to perceive and produce L2 sounds in terms of their L1 sounds. The Perceptual Assimilation Model (PAM), which was proposed by Best (1995; 1999), assumes a relationship between the perception and production of L2 sounds. Best claims that the perception of sounds in a foreign language is determined by their similarities to, or discrepancies from, L1 sounds. Additionally, perceptual limitations determine the kind of difficulty that L2 learners may encounter in learning L2 sounds. Thus, PAM postulates that listening to L2 sounds is not a straightforward process of mapping L2 sounds to their similar cognates in L1 sounds or deciding the sounds that differ from L1 sounds. It involves discriminating two L2 sounds from one another as well as distinguishing the L2 sounds from L1 sounds.

Relying on the degree of similarity between L2 and L1 sounds, Best (1995, 1999) predicted the discrimination of four assimilation categories. First, *Two-Category (TC)* members of the L2 contrast assimilate to two different L1 categories. This category yields a very good discrimination. Second, *Category Goodness (CG)*, in which each member of the L2 contrast assimilates to one native category with one of the members being more deviant from the native sound than the other. Thus, this category yields a good discrimination. Third, *Single Category (SC)*, in which both L2 sounds assimilate to one phoneme in the native category and both are equally deviant from the native sound, and this yields poor discrimination. Finally, if the non-native sounds are *Non-Assimilable* (NA) to any native category, they will be perceived with good discrimination as non-speech sounds. A number of studies have examined the perception of non-native contrasts based on the predictions of PAM such as Pilus(2005), Best et al. (2001), and Ingram and Park (1997).

The present study examines the productions of three SSBE vocalic contrasts, /ɪ/- /ɛ/, /ɒ/- /ʌ/ and /æ/- /ɑː/ by SA learners of English in terms of PAM predictions. A pilot study consisting of two experiments was conducted to examine the perception of SSBE vowels by SA learners. First, the 'classification experiment' which examined the perception of SSBE vowels in terms of SA L1 phonemic vowel inventory which included (/iː/, /eː/, /aː/, /oː/, /uː/, /i/, /e/, /a/, /o/, and /u/). The results showed that /ɪ/- /ɛ/ were classified as SA /e/, /ɒ/- /ʌ/ as SA /a/, and /æ/- /ɑː/ as SA /aː/. The first member of these contrasts was more deviant from the SA sound. Second, the 'identification' experiment in which SA speakers were asked to identify SSBE vowels. The results showed that the second member of /ɪ/- /ɛ/, /ɒ/- /ʌ/ contrasts, which were closer to the SA vowels, were fully identified, whereas the first members were not. Unexpectedly, the two members of /æ/- /ɑː/ contrasts were fully identified by SA speakers. A preliminary explanation for this result is that SA speakers use the characteristics of pharyngealisation from their L1, i.e. tongue body retraction and constriction in the upper pharynx, to identify /ɑː/. According to PAM perceptual categories, /ɪ/- /ɛ/ and /ɒ/- /ʌ/ contrasts were considered as *CG* while /æ/- /ɑː/ as *TC*. This study examines the accuracy of the predictions of this classification in terms of SA speakers' productions of SSBE contrasts.

2. METHODS

2.1. Speakers

Five SA female speakers were chosen to participate in the production task. In order to avoid dialectal variation in Arabic, only SA speakers of Damascene variety were recorded. Following Strange (1998), SA subjects were chosen to have no exposure to English in an English speaking community. All SA subjects had English formal education during school and university for at least 10 years. To guarantee a good level of proficiency in English, all subjects were chosen to be university students in the English department at Damascus University or medium/high English level students at Asia Institute for Languages. A questionnaire was used to extract information about the speakers, in general, and their language background, in particular.

Table 1: SA speakers details and the scores of their English skills

Speaker	Age	Formal education	Speaking	Understanding	Reading	Writing	Average score
S1	30	10	6	6	6	6	**6**
S2	23	12	7	6	7	7	**6.7**
S3	28	12	5	5	5	3	**4.5**
S4	38	10	7	7	7	7	**7**
S5	23	16	3	5	6	6	**5**
Average	**28.4**	**12**	**5.6**	**5.8**	**6.2**	**5.8**	**5.8**

Table 1 above shows that the average age of SA speakers is 28.4 years old, and the average of their formal education is 12 years. The speakers were asked to rate their language skills abilities on a 7 points scale. The rate for reading was higher than the rates for the other skills; this is an expected result since their English input relies mostly on formal Education which focuses on reading more than any other skills. On the other hand, the rate for speaking was the least which is also expected since this skill is not developed efficiently through formal education. Depending on these ratings, speakers are expected to have difficulty producing what matches the native English productions. As a control group, four SSBE female speakers were also recorded, their average age was 34.75.

2.2. Recordings

Syrian Arabic speakers were recorded at Asia Institute for Languages in Damascus. They were recorded in a quiet computer room using Marantz Professional Portable Solid State Recorder, PMD660 model. The microphone was Shure Professional unidirectional, which was a head-worn dynamic microphone. Audio files were recorded at 44.1 Khz sampling rate as .wav files on a compact flash TM memory card.

SSBE speakers were recorded in acoustically treated recording area in the linguistic lab at the University of York. The recordings were carried out using Behringer C1 large diaphragm condenser microphone, TAC Scorpion-Mixing desk, M-Audio LT1010 PCI Audio card, and Adobe audition v1.0 on PC software with 16bit and 44.1Khz sampling rate.

2.3. Stimulus Material

2.3.1. SSBE Stimuli

SSBE and SA speakers produced five randomized blocks of /hVd/ keywords among which the target words for this study, (hid, head, hod, hudd, had, and hard) representing the 6 SSBE vowels (/ɪ/- /ɛ/, /ɒ/- /ʌ/, and /æ/- /ɑː/) respectively. /hVd/ context was found to have negligible coarticulatory effect on the vowel (Stevens and House 1963). The target words were inserted in the phrase "say ____ again". The best three productions of each vowel for each speaker, which did not contain any noise, hesitation or mispronunciation, were extracted and analyzed acoustically. Together the SSBE stimuli included 3 repetitions × 6 vowels × 9 speakers = 162 items.

2.3.2. Syrian Arabic Stimuli

In order to test the effect of L1 on L2 productions, SA speakers were asked to produce five randomized blocks of the SA vowels (/aː/, /aːˤ/, /ɪ/, /e/, /o/, /a/). /aːˤ/ was added to the analysis in order to compare it with SSBE /ɑː/. However, this vowel occurs after pharyngealised segments, therefore, a different context was used for it, as can be seen in table 2 below. SA target vowels were put in the same monosyllabic /hVd/ context used for SSBE stimuli, and the target words were put in the phrase /ktoːb ___ marteːn/ "Write _____ twice". The best three productions of each vowel for each speaker were extracted similarly to SSBE data. Together SA stimuli included 3 repetitions × 6 vowels × 5 speakers = 90 items.

It was difficult to have real /hVd/ words in SA for all target vowels since SA short vowels do not occur in monosyllabic words (Cowell 1964) . Thus, di-syllabic words were presented simultaneously with the /hVd/ words, and the speakers were asked to produce the target word in a similar way to the second part of the disyllabic word as illustrated for /e/ and /o/ in table 2 below.

Table 2: SA vowels which best match the SSBE vowels under investigation

Vowel	Target word	Similar to target	Arabic word	English gloss
/aː/	/haːd/	/haːd/	هاد	this one
/aːˤ/	/tˤaːˤr/	/tˤaːˤr/	طار	flew
/i/	/hɪd/	/hɪdd/	هِد	destroy !
/e/	/hed/	/naːhed/	ناهد	proper name
/o/	/hod/	/hidhod/	هدهد	kind of bird
/a/	/had/	/hadd/	هَد	he destroyed

2.3.3. Procedures

The stimulus material was presented to the speakers using flash cards. Each card contained two phrases in a random order. The English stimulus material was presented in conventional English orthography, and the Arabic stimulus material was presented in conventional Standard Arabic orthography. Nevertheless, the speakers were asked to read the Arabic phrases in their local SA and not as standard Arabic. SA and SSBE speakers were acquainted with the target words in advance, and when having a difficulty with a particular word, another familiar word with the same vowel was presented.

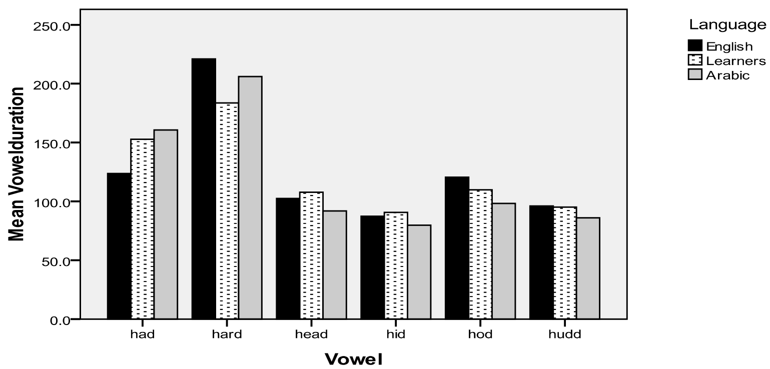
2.4. Data Analysis

The target vowels were extracted and analysed acoustically using Praat 5.1.14 (Boersma and Weenink 2009). The analysis procedures were automated using a Praat script. The acoustic landmarks being investigated were identified as time points using TextGrid files. Spectrograms were used for segmentation and waveforms for more fine-grained segmentation decisions (Turk et al. 2006). I placed the boundaries carefully by hand using waveforms and wide-band spectrograms in addition to auditory verification. For this study, the first two formants F1-F2 and vowel duration were extracted. Following Yang (1996), the beginning of the vowel was identified manually as the end of noise for /h/ and beginning of the periodic waveform, whereas the end of the vowel was identified as the end of the periodic waveform and the beginning of the closure period of /d/. Following McDougall and Nolan (2007), frequency measurements were extracted from +10% automatically selected steps, which is useful to show formants dynamics. However, for the purpose of this study, only midpoint frequency measurements were statistically examined.

3. Results and Discussion

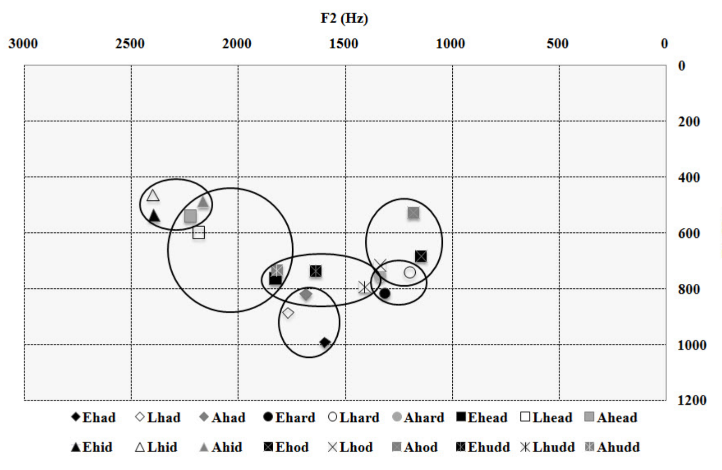
A multivariate ANOVA was used to test for significant differences for six vowels /ɪ/- /ɛ/, /ɒ/- /ʌ/, and /æ/- /ɑː/. For each test, Language (English, Arabic, and Learners) was used as the fixed factor, and F1, F2, and vowel duration as the dependent variables. SA vowels (/aː/, /aːˀ/, /e/, /i/, /o/, and /a/), which were mapped to the SSBE vowels, respectively, were added to the test in order to examine the effect of L1.

Figure 1: Mean vowel duration of six English vowels as they were produced by SSBE and SA speakers, compared to six Arabic vowels (/aː/, /aːˀ/, /e/, /i/, /o/, and /a/) produced by the same SA speakers.



The results showed that there was a significant effect for Language on vowel duration of [had] only *F*(2, 42) = 4.622, *p* < .01. A Tukey Post hoc test showed that the language effect was significant between English and Arabic vowels only. Thus, duration had no effect on the English productions of SA learners, as can be seen in Fig. 1.

Figure 2: F1 and F2 frequencies of the six vowels as they were produced in English by SSBE speakers (black), SA learners (transparent), and in Arabic (gray).



There was also a significant effect for language on F1 for [had] F(2, 42) = 16.072, p < .000, [hard] F(2, 42) = 4.329, p < .02, [head] F(2, 42) = 57.201, p < .000, [hid] F(2, 42) = 3.980, p < .02, and [hod] F(2, 42) = 18.963, p < .000, but not for [hudd] . F(2, 42) = 1.620, p < .2. As for F2, language had a significant effect for [hard] F(2, 42) = 5.977, p < .005, [head] F(2, 42) = 11.001, p < .000, [hod] F(2, 42) = 6.856, p = .003, and

[hudd] $F(2, 42) = 35.778$, $p < .000$, but not for [had] $F(2, 42) = 2.942$, $p <= .06$ or [hid] $F(2, 42) = 2.728$, $p < .07$.

Post hoc comparisons with the Tukey procedure revealed that for [had] there were significant differences between SA English productions and native SSBE productions in F1($p = .004$) and F2 ($p = .05$). However, there were no differences between SA English and Arabic productions. As for hard, F2 of SA English productions were significantly different from SSBE F2 ($p = .03$) and SA F2 ($p = .007$). Figure 2 shows that the effect of SA /aː/, /aːˀ/is still prevailing, particularly for [had]. Nevertheless, SA speakers are approximating SSBE native productions, particularly for [hard], where they used more F2 lowering which accompanies pharyngealisation to approximate the English vowel /ɑː/. Although /æ/- /ɑː/ contrast was accurately identified by SA speakers, it was not produced in the same accuracy. Thus, Best's *TC* classification which was applied to this contrast yields very good discrimination but not necessarily very good production because the members of this contrast were both acoustically deviant from their closest L1 vowels.

On the other hand, Post hoc comparisons with the Tukey procedure for [head] revealed that SA English productions were significantly different from SSBE in F1($p = .000$) and F2 ($p = .02$). However, SA English productions of [head] were significantly different from SA in F1($p = .01$) but not in F2 ($p = .8$). Furthermore, there were no significant differences in F1 and F2 of [hid] between SA, SSBE, and SA English productions. Figure 2 shows that SA /e/ and /i/ are very close to each other, and SA English production of /ɛ/ is close to both Arabic vowels. Having the three productions close to each other might indicate that SA /e/ and /i/ represent a single vowel in SA rather than two, which accounts for perceiving both members of the SSBE /ɪ/- /ɛ/ contrast as SA /e/. Thus, considering them as *CG* in terms of Best's perceptual categories was accurate but with SSBE /ɪ/ closer to SA vowel.

The results of the Post hoc comparisons for [hod] and [hudd] showed that SA English productions were significantly different from SSBE only in F2 ($p = .005$) and ($p < .000$), respectively. They were also different from SA in F2 ($p = .01$), and ($p <.000$), respectively. There were no differences in F1, which indicates that SA learners were able to figure the height of SSBE /ɒ/- /ʌ/ but not the place, and this might be due to the lack of central vowels in SA. According to Best's categories, /ɒ/- /ʌ/ contrast was classified as CG, however, this classification did not apply to SA productions of /ɒ/- /ʌ/. Figure 2 shows that, acoustically, the members of this contrast are both deviant from SA /a/, and they should be considered as SC which accounts for any poor discrimination.

In sum, the perceptual categories of Best (1995) do not necessarily apply to the learners' production patterns. This also determines the kind of relationship between production and perception of L2 sounds and their relationship with L1 sounds. Accurate perception of particular contrasts do not require accurate production, and this is reasonable because changing the articulatory configurations used in L1needs more time, and for some sounds learners might not succeed amending their articulatory configurations for L2 sounds due to perceptual assimilation with L1 sounds.

Further analysis for SA vowels, particularly mid vowels, needs to be conducted to determine the acoustic characteristics of the vocalic inventory of SA. Additionally, SA learners with experience in SSBE community needs to be recorded and compared to the inexperienced speakers of this study, who had good level of English, in general.

4. Acknowledgments

I would like to thank the University of York and Aleppo University for funding this research. I also wish to express my sincere gratitude to my supervisor Dr. Sam Hellmuth for her unconditional support and guidance, and to Dr. Paul Foulks and Dr. Danijela Trenkic for their valuable comments. Many thanks goes to the participants and to Asia institute for languages in Damascus for their help in the recordings and in recruiting some participants from their students. I also wish to thank my husband for his love and support.

5. References

Best, C. T. 1995. A direct realist view of cross-language speech perception. Speech perception and linguistic experience: Issues in cross-language research W. Strange. 171–204. Baltimore: York Press.

Best, C. T. 1999. Development of Language-Specific Influences on Speech Perception and Production in Pre-verbal Infancy. *International Congress of Phonetic Sciences*. Department of Linguistics, University of California at Berkeley.

Boersma, P. and D. Weenink. 2009. "Praat: doing phonetics by computer (Version 5.1.14) [Computer program]." Praat homepage. <http://www.praat.org/>. Retrieved 27 Aug 2009.

Cowell, M. 1964. *A Reference Grammar of Syrian Arabic*. ed. Georgetown University Press.

Ingram, J. and S. Park. 1997. Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics* 25. 343 – 370.

McDougall, K., Nolan, F. 2007. Discrimination of speakers using the formant dynamics of /uː/ in British English. *Proceedings of the 16th ICPhS, Saarbrücken*. 1825-1828.

Pilus, Z. 2005. The perception of voicing in English word-final consonants: Examining the perceptual assimilation model. *Malaysian Journal of ELT Research.* 1.

Strange, W., R. Yamada, R. Kubo, S. Trent, K. Nishi and J. Jenkins. 1998. Perceptual assimilation of American English vowels by Japanese listeners. *Journal of Phonetics* 26. 311-344.

Stevens, K. and A. House. 1963. Perturbation of Vowel Articulations by Consonantal Context: an Acoustical Study. *Journal of Speech and Hearing Research* 6. 111-128.

Turk, A., Nakai, S., Sugahara, M. 2006. Acoustic Segment Durations in Prosodic Research: A Practical Guide. In: *Methods in Empirical Prosody Research.* Sudhoff, S., D. Leneratova, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter and J. Schlieβer (eds.). Berlin: Walter de Gruyter. 1-27.

Yang, B. 1996. A comparative study of American English and Korean vowels produced by male and female speakers. *Journal of Phonetics* 24. 245 – 261.

Measuring perceptual cue weighting after training: A comparison of auditory vs. articulatory training methods

Cristina Aliaga-Garcia

Universitat de Barcelona
cristinaaliaga@ub.edu[1]

ABSTRACT

Language learners' difficulties with second-language (L2) sounds often lie in the recognition and weighting of the most reliable cues to vowel categorization. A good example is Catalan/Spanish learners' poor perception of English vowels, which tends to focus on durational rather than spectral information. Several perception training studies have shown that L2 learners can learn to integrate multiple relevant acoustic cues in L2 perception. This study addresses the question of whether native-like acoustic cue weighting is best promoted by identification (ID) or articulatory (ART) audiovisual High Variability Phonetic Training (HVPT). Separate groups of bilingual Catalan/Spanish learners of English (*N*=64) received ID and ART training on the 11 English RP monophthongs (/i ɪ e æ ʌ ɑ ɜ ɒ ɔ ʊ u/), and were pre- and post-tested through a 4-choice categorization task with natural CVC words and a forced-choice categorization task based on 8 synthesized /hVd/ continua (/iː/-/ɪ/, /æ/-/ʌ/, /ʌ/-/ɑː/, /ʊ/-/uː/). The study revealed significant training effects on L2 vowel categorization and a more reliable use of cue weighting, depending less on duration. The effect of *training type* was found to significantly interact with *vowel condition* (natural vs. synthesized), *vowel type* and degree of *dispersion* of errors per vowel type examined.

Keywords: auditory vs. articulatory training, audiovisual training, cue weighting, duration.

1. INTRODUCTION

It is well established that the L1 vowel system exerts a tremendous influence on L2 speech learning and, more specifically, in the degree of success with which L2 speech sounds are acquired. Languages vary in the size of their vowel inventories, ranging from three-vowel to fifteen-vowel systems, and therefore posing different degrees of difficulty in perceiving and producing L2 vowel categories (Iverson and Evans 2007). Previous work (Escudero 2000, Cebrian 2006) has generally attributed such learning constraints to language learners' difficulties in recognizing and weighting the most reliable cues to L2 vowel categorization. The multiplicity of cues to phonetic contrasts differing systematically across languages is also well documented in the literature (McAllister et all 2002; Bohn 1995;). Different L1 vowel inventories often use different acoustic cues in phonological distinction or rather weight the same acoustic cue differently, giving higher weight to the critical or most reliable cues for sound recognition (Holt and Lotto 2006). It is often the case that learners with relative simple "uncrowded" L1 vowel systems (i.e. 5-vowel systems of Spanish and Greek), and therefore at initial disadvantage, may undergo an arduous work when learning more complex L2 vowel systems requiring the use of multiple cues. Given the difficulty discerning differences between L2 vowel categories, learners can mistakenly apply L1 cues to L2 speech recognition, or simply resort to L2 secondary cues as a compensatory strategy, whether they are phonologically contrastive in their L1 or not (Bohn and Flege 1990). If secondary cues are weighted higher than primary cues in L2 perception, this will probably translate into learners producing L2 vowels "coloured" by properties of their L1. On the contrary, it has been assumed that having a large and complex L1 vowel system (i.e. 11- and 13- vowel systems of English and German, respectively) may *speed up* the process of L2 learning, as individuals may more successfully recognize the salient L2 cues by rapidly resorting to their native vowel inventory and applying L1 cues easily available to them. However, this advantage for speakers of large vowel inventories has been limited to initial stages of L2 learning and called into question in the literature, as it seems that learning new L2 categories requires changing or readjusting so many existing categories in the L1 (Flege 1995, 2003; Munro et al. 1996).

Adult L2 learners' difficulties learning L2 vowel categories have also been explained in the light of current L2 speech perception models. Both Best's Perceptual Assimilation Model (PAM) (Best 1995) and Flege's Speech Learning Model SLM (Flege 1995) have observed L2 learning difficulties in terms of the (dis)similarity between L1 and L2 systems. According to PAM, the most challenging L2 vowels, acoustically similar to L1 vowel categories, can be easily (wrongly) assimilated into one single L1 category. Alternatively, SLM predicts that learning new L2 categories become somehow easier when these fall into "uncrowded" areas of the L1 vowel space and therefore do not overlap with similar confusable L1 categories. Regardless of the difficulties faced by adult L2 learners, the SLM has lent support to the positive effect of L1/L2 experience –operationalized as length of residence (LOR), amount of L1/L2 use, and age of onset of L2 learning (AOL)– on L2 speech perception and production, leading to less accented speech (Piske et al. 2001) and more successful development of new phonetic categories for L2 sounds.

Alternatively, phonetic training studies constitute another promising area of research in the light of arguments in favour of ongoing plasticity in L2 speech perception. Regardless of the challenge that L2 speech learning poses for L2 learners, different training techniques and studies conducted so far have established that it is possible to approach native-like perception and production of L2 sounds (i.e. Jamieson and Morosan 1989; Logan and Pruitt 1995; Iverson & Evans 2007; Nishi and Kewley-Port 2008; Ylinen et al. 2009). Perceptual training studies conducted so far have mainly investigated both consonant and vowel contrasts, with the exception of a few suprasegmental studies (i.e. Wang et al. 2003), but the case of Japanese adults learning the English /r/-/l/ contrast has been one of the main focus of research within the area (i.e. Logan et al. 1991), aiming at reallocating the learners' attention on the L2 relevant cues by means of the Perceptual Fading technique, which made use of (maximally contrastive) enhanced stimuli, as well as techniques based on cue manipulation, such as All Enhancement and Secondary Cue Variability types of training (Iverson et al. 2005). High-variability phonetic training (HVPT) has, alternatively, proved particularly effective in promoting 'robust' L2 categories, by means of a high variety of natural stimuli and multiple talkers, especially when the difficulties with L2 sounds lie in the use of L2 phonetic cues which are unused or weighted differently in the L1, or are rather based on the integration of multiple cues (). Ultimately, audiovisual (AV) types of HVPT (Hardison 1999; Ortega-Llebaria et al. 2001) have proved more effective than only-auditory training, leading to long-lasting improvement in L2 sound categorization, even in environmental degradation, and improving L2 production (Bradlow et al. 1997). However, research on the assessment of methods other than perceptual training for non-native vowels is still scarce (i.e. production training, Hattori and Iverson 2009) and, to my knowledge, none of the previous vowel studies have compared the impact of auditory vs. production-based training on the relative weighting of acoustic cues.

The present study explores this issue further by comparing the effectiveness of two types of HVPT training, namely auditory vs. articulatory training, on L2 vowel perception by Spanish/Catalan advanced learners of EFL. It is probably the first training study which addresses the question of whether native-like acoustic cue weighting is best promoted by auditory or articulatory training for native speakers of smaller vowel systems than English. The fact that Catalan/Spanish native speakers identify English vowels poorly lies in the nature of their L1 system: the Spanish and Catalan vowel systems consists of 5 and 7 vowels, respectively, and, unlike English, there are no tense-lax or temporal contrasts in either of these languages. These EFL learners consequently tend to focus on duration rather than spectral information for L2 vowel categorization, despite duration being non-contrastive in their L1, possibly because single-category assimilation may have desensitized them for the perception of spectral cues (Escudero and Boersma 2004; Cebrian 2006).

The first aim of the study was to assess whether higher accuracy in L2 perception could be best promoted through identification or articulatory HVTP procedures, as assessed by a 4-choice categorization task with natural CVC words and a forced-choice categorization task based on 8 synthesized /hVd/ continua. The second aim was to explore learners' categorization of natural stimuli (with unmodified duration) and synthesized stimuli (with manipulated duration; i.e. when /i/ is shortened and /ɪ/ lengthened) following each of the training procedures.

2. METHOD

The general design of this study has three phases: (1) a pre-test phase, (2) an identification training phase parallel to an articulatory training phase, and (3) a post-test phase.

2.1. Participants

The subjects in the present study were a group of adult Spanish-Catalan bilinguals (*N*=84; 12 males, 72 females) in their second and third year of a degree in English studies at the University of Barcelona, and a group of southern British English (SBE) speakers (*N*=10; 5 male, 5 female) used for baseline data, both of which completed pre- and post-tests. Only a subset (*N*=64) of advanced EFL learners (*mean age:* 22.3; *range:* 19-50; 8 males and 56 females) enrolled in a phonetic training program and were given course credit for their participation. All the experimental subjects were Catalan-Spanish bilinguals living in Barcelona and neighbouring areas. All had at least one prior semester of formal study of English Phonetics but had not studied English abroad (for longer than 2 months). Specific exclusion criteria for the study verified further that students were not from families in which at least one of the parents was a native speaker of English. All reported having normal hearing or having no speech-related dysfunctions. The SBE control tested at University College London (UCL) lived in London except for one male speaker of the same Southern variety of English residing in Barcelona but having spent most of his life in the London area.

2.2. Stimuli and apparatus

The audiovisual training corpus comprised 163 different CVC real words containing the whole set of English monophthongal vowels (/i ɪ e æ ʌ ɑ ɜ ɒ ɔ ʊ u/), in high contextual variation (b_d, b_t, d_d, d_n, f_l, h_l, k_n, l_n, m_d, p_k, p_t, s_d, t_k, t_n, w_t) as pronounced by 10 different talkers (5 male, 5 female) representing typical SBE pronunciation. The stimuli words were carefully read at normal speed and falling intonation and videorecorded using a Canon XL-1 DV camrecorder, and a Bruel and Kjaer type 4165 microphone, in an acoustically attenuated chamber in the Phonetics Laboratory at UCL (London), following a short training phase during which talkers were instructed how to read the words one at a time on a computer screen to avoid list-reading intonation. The talker's face was set against a blue-background and illuminated with a key and a fill light, and the resulting videoclips were edited so that the start and end frames of each token, lasting 3 seconds, showed a neutral facial expression.

The testing corpus originally consisted of 66 CVC stimuli (6 tokens x 11 vowels) containing the 11 English vowels in 14 different contexts (b_n, b_s, d_l, f_b, f_d, g_d, h_d, h_m, k_p, l_d, l_k, m_l, p_m, r_t, s_d, s_k, s_ts, ʃ_d, ʃ_t, t_g, t_t, w_d, w_f, w_l) as pronounced by 2 novel SBE speakers (1 male, 1 female). 50% of the tokens (N=33) converged in the two sets of stimuli but differed in talker. The reason for this coincidence was that pre- and post-tests aimed at comparing the learners' performance on trained and untrained stimuli. The testing stimuli were recorded, segmented and edited following exactly the same procedures used for the training stimuli. Finally, the 66 video files were converted to sound, creating a new subset of 66 CVC stimuli so that each testing stimulus could be presented auditorily and audiovisually. (/i ɪ e æ ʌ ɑ ɜ ɒ ɔ ʊ u/),

2.3. High-variability Phonetic Training

Separate groups of Catalan/Spanish learners of English (*N*=64) received different types of AV High Variability Phonetic Training (HVPT) (natural CVC words from multiple talkers), namely identification (ID) (*N*=32) and articulatory (ART) training (*N*=32), on the 11 English RP monophthongal vowels. The HVPT training comprised 10 computer-based sessions over 5 weeks, designed and run using DMDX software, which guaranteed exposure to a minimum of 176 trials (4 tokens x 11 vowels x 4 repetitions or blocks) and included the three subsets of vowels: high front vowels /i ɪ e/, high back vowels /ɔ ʊ u/, and low vowels /æ ʌ ɑ/ as pronounced by 2 different talkers (i.e. NS1 blocks 1 and 3; NS2 blocks 2 and 4). Both ID and ART training sessions were administered separately in a quiet computer room, with an omnidirectional headset microphone Soyntec550. ID participants listened to the stimuli and gave their responses with a mouse click by selecting buttons appearing on the computer screen. ART training required participants to listen to the stimuli over headphones, speak into the microphone and use the keyboard to press buttons appearing on the computer screen for word repetition or going to the next item.

2.3.1. Identification training

In the ID training the participants heard the whole set of vowels within a varied CVC context and watched short video-clips showing the articulation of words. On each trial, subjects were required to focus on both AV and acoustic cues for vowel recognition and then choose the correct response among 3 or 4 alternatives. After errors, they heard the wrong response immediately followed by the correct word and they could try again. They were showed the mean percent of correct identification at the end of the session.

2.3.2. Articulatory training

The ART training consisted of one imitation training task in which learners were presented words audiovisually so that they could focus both on acoustic properties of sound and articulatory gestures for more accurate vowel articulation. At the end of each section, they heard their own responses (‘*This is you…*’), compared them with the native speakers (‘*This is the native speaker*…’) and then corrected themselves as many times as necessary (‘*Try again.*’).

2.4. Pre- and Post-test

Both ID and ART groups were pre- and post- tested using a 4-choice categorization task with natural CVC words including the whole vowel set and a forced-choice categorization task based on 8 synthesized /hVd/ continua (/iː/-/ɪ/, /æ/-/ʌ/, /ʌ/-/ɑː/, /ʊ/-/uː/). Ten native speakers of SBE served as baseline data. The perceptual data were analyzed by calculating mean percent correct identification of natural vs. synthesized vowels (with vs. without manipulated duration), percentage of gains, mean slope coefficients and index of dispersion of wrong responses per vowel (natural vs. synthesized), word type (trained vs. untrained) and traning type.

The two perceptual tasks were administered through DMDX software in a quiet computer.The subjects listened to the stimuli individually over headphones and gave their responses by selecting buttons appearing on the computer screen. Each perception tasks was preceded by a short training phase to familiarize participants with the range of possible identification responses. On average it took subjects about 45 minutes to complete the two tasks, but they were allowed to stop and rest if necessary.

2.4.1. Identification task I

Subjects first completed a 30-minute multiple-choice identification (ID) task with 264 CVC natural words distributed into 2 blocks or conditions (6 tokens x 11 vowels x 2 talkers x 2 blocks). In the ID task participants were presented two types of stimuli, trained vs. untrained, in two conditions: AV vs. only-audiroty (A) stimulus presentation (2 orders: A-AV, AV-A). The subjects were required to watch a 3-second videoclip of the talkers’ mouth, or rather hear the stimulus, and then click on the correct response among 3-4 possible responses displayed horizontally (i.e. *feel-fill-fell-furl, cat-cut-cart-cot, Paul-pull-pool*). (/iː/-/ɪ/, /æ/-/ʌ/, /ʌ/-/ɑː/, /ʊ/-/uː/).

2.4.2. Identification task II

Following identification testing for natural stimuli, subjects completed a 15-minute forced-choice ID task with 280 synthesized /hVd/ stimuli (8 vowels x 7 steps x 5 repetitions) based on 8 seven-step duration continua (80, 116.7, 153.3, 190, 226.7, 263.3, 300 ms) which were created from the vowel pairs /hiːd/-/hɪd/, /hæd/-/hʌd/, /hʌd/-/hɑːd/ and /hʊd/-/huːd/, respectively, using Praat software (Boersma & Weenink 2004). The main aim was to assess the learners’ vowel perception when the duration cue was ambiguous (i.e. when /iː/ vowels were shortened and /ɪ/ lengthened). Participants heard one cue-manipulated stimuli at a time and had to identify one of the two responses on the screen (one on the left, another on the right: i.e. *he’d* vs. *hid*).

3. RESULTS AND CONCLUSION

The participants' perceptual phonological competence was assessed by computing mean percent correct vowel identification scores for each subject, *training type* (ID vs. ART), *vowel type*, *token type* (trained vs. untrained) and *vowel condition* (natural vs. synthesized) at pre- and post-test. Changes in natural vowel ID after training were measured through the *degree of dispersion* of errors (score from 1 to 3) and mean percent gains obtained by subtracting pre-test from post-test scores. In order to further assess improvement in synthesized vowel ID, and in particular L2 cue (re)weigthing after training, changes in *slope values* of the vowel ID functions were explored and a *duration effect* score was computed for each of the synthesized vowels by subtracting correct ID at the last step of the continuum with „manipulated duration" from that at the first step with „normal duration".

Table 1: Means for vowel perception measures: identification (ID) of natural and synthesized vowels at pre- and post-test.

Vowel ID	Experimental group (N=64)								Control group (N=20)				NS (N=10)
	ID training (N=32)				**ART training** (N=32)								
	T1	T2	T1vsT2	gains	T1	T2	T1vsT2	gains	T1	T2	T1vsT2	gains	T1
Nat. Vs	**61.55**	**79.18**	***.000**	***17.63***	**67.28**	**80.16**	***.002**	***12.88***	63.16	68.73	.125	*5.57*	**98.11**
Nat /iː/	60.55	76.17	***.000**	*15.62*	62.89	75.00	***.003**	*12.11*	56.04	61.88	.190	*5.84*	91.92
Nat /ɪ/	58.72	80.07	***.000**	*21.35*	66.93	80.08	***.007**	*13.15*	62.92	65.00	.940	*2.08*	99.48
Nat /æ/	50.11	76.17	***.001**	*26.06*	71.22	78.52	***.131**	*7.30*	59.17	64.58	.214	*5.41*	100.0
Nat /ʌ/	65.56	78.64	***.000**	*13.08*	62.50	80.60	***.000**	*18.10*	48.54	48.13	.899	*-0.41*	98.96
Nat /ɑː/	67.97	83.72	***.001**	*15.75*	72.13	83.07	***.008**	*10.94*	57.08	62.92	.350	*5.84*	99.48
Nat /ʊ/	45.96	57.16	***.001**	*11.20*	48.83	55.34	***.030**	*6.51*	55.00	56.46	.852	*1.46*	82.50
Nat /uː/	44.66	67.06	***.000**	*22.40*	48.83	66.15	***.000**	*17.32*	53.54	65.21	.064	*11.67*	93.44
Synt. Vs	**67.66**	**76.72**	***.027**	**9.06**	**66.18**	**78.14**	***.013**	**11.95**	57.98	62.96	.192	4.98	**97.39**
Syn /iː/	68.39	70.80	.248	2.41	70.27	80.45	.538	10.18	55.43	65.86	.197	10.43	98.93
Syn /ɪ/	67.86	72.59	.578	4.73	69.73	79.02	.123	9.29	55.00	63.72	.085	8.72	99.28
Syn /æ/	93.48	95.36	***.000**	1.88	89.20	94.38	***.042**	5.18	61.00	63.00	.088	2.00	99.28
Syn /ʌ/	65.89	72.41	***.038**	6.52	58.48	69.28	***.010**	10.80	66.71	66.86	.200	0.15	99.64
Syn /ɑː/	56.96	67.68	***.005**	10.72	57.77	65.54	***.008**	7.77	55.00	55.71	.272	0.71	87.50
Syn /ʊ/	61.16	79.91	.469	18.75	59.64	78.30	.144	18.66	58.28	54.00	.271	-4.28	98.93
Syn /uː/	59.91	78.30	.926	18.39	58.21	80	***.041**	21.79	54.43	61.57	.131	7.14	98.21

Table 1 illustrates mean percent correct ID of natural vs. synthesized vowels per training group before and after training. The main result revealed by ANOVA was an overall significant main effect of time of testing for percent correct ID of vowel with and without manipulated duration. Each of the groups showed similar significant improvement in their ID accuracy at post-test, regardless of vowel condition, token type and vowel type. These results were very encouraging in terms of the effectiveness of training, as improvement also generalized to untrained tokens and any token type from novel talkers, even when the duration cue was ambiguous and could not be relied upon for L2 vowel recognition.

To our surprise, ANOVA did not reveal a significant main effect of training type, meaning that ID and ART types of training did not produce significantly different effects on L2 vowel categorization. However, statistical results showed a significant training type x token type interaction, the ID group outperforming the ART one for trained tokens. This result indicated that the words used in the training were far easier than the those from the pre- and post-tests. It also suggests that the ID training was more effective than the ART training as regards the development of accurate (long-term) representations of sounds for "familiar" tokens administered during the training. Between-group comparisons of the dispersion index scores indicated diverging patterns of vowel (mis)identification for training type: the ID group obtained lesser degree of dispersion of wrong responses (T1: 1.89; T2: 1.59) than the ART group (T1: 1.71; 1.70), esp. for vowels /ɪ/, /ʌ/, /ɑ/ and /uː./

The main significant finding with regard to the effect of training type on cue weighting was that the percent correct for duration-modified /æ/, /ʌ/, /ɑ/ was significantly higher for both ID and ART training, but only the ART improved significantly for /ʊ/ with manipulated duration. With the duration-modified /iː/ and

/ɪ/, the improvement observed from pre- to post-test did not reach significance, however. On the other hand, within-group analysis showed significantly lower slope coefficients and steeper categorization functions for both groups at post-test, meaning higher precision and (slighlty) nore categorical perception of vowels with manipulated duration at post-test. Despite the overall improvement observed for synthesized vowels, esp. the significant improvement observed for low and back vowels –which indicates a positive trend towards more native-like use of cue weighting–, percent natural vowels were generally better identified than their cue-manipulated counterparts, pointing out a still pervasive duration effect after training. Importantly, a significant effect of the manipulation of duration was found for shortened shortened /iː/ (steps 1-3 on the continuum: 80-153,3ms) and lengthened /ɪ/ (steps 4-7 on the continuum: 190-300ms), meaning that learners very often relied on the duration cue after training. Finally, paired-samples t-tests revealed a significant overall decrease in the duration effect scores for the ID groups' accurate identification of synthesized /i/, /ɪ/, /ʌ/ and /ɑ/, and for the ART's identification of /ɪ/, /æ/, /ʌ/ and /ʊ/, indicating a smaller distorting effect of duration modification for some vowels.

The present study aimed to train Catalan/Spanish learners of EFL to weight L2 phonetic cues differently by means of two types of AV training, namely identification and articulatory training. Repeated-measures t-test and mixed ANOVA analyses revealed significant improvement in the learners' ability to identify vowels with normal and modified duration at post-test. Following the 10-session high-variability phonetic training, native speakers of Catalan/Spanish obtained 13.34% gains in correct ID of natural vowels and 14.41% gains in correct ID of synthesized vowels. Although the ID and ART training groups slighltly differed as regards some measures of L2 vowel perception, it can be concluded that, after ID and ART training, Catalan/Spanish learners of English were more reliably able to distinguish two similar vowels based on spectral cues, without further significant differences found in the performance of the two language groups.

4. REFERENCES

Best. 1995. A direct realist view of cross-language speech perception. Speech Perception and Language Experience: Issues in Cross-Language research. Strange, W. (ed.). 171-204.

Bohn, O. S. 1995. Cross-language speech perception in adults: First language transfer doesn't tell it all. In: Strange, W. (ed.). *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press. 279-302.

Bohn, O., Flege, J. E. 1990. Interlingual identification and the role of foreign language experience in L2 speech perception. *Applied Psycholinguistics 11. 303-328.*

Bradlow, A. R., D. B. Pisoni, R. A. Yamada, and Y. Tohkura. 1997. Training Japanese listeners to identify English /r/ and /l/: II. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America* 101. 2299-2310.

Cebrian, J. 2006. Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics 34*. 371-387.

Escudero, P. 2000. Developmental patterns in the adult L2 acquisition of new contrasts: The acoustic cue weighting in the perception of Scottish tense/lax vowels in Spanish speakers. Unpublished M. Sc. thesis, University of Edinburgh.

Escudero, P. 2000. Developmental patterns in the adult L2 acquisition of new contrasts: The acoustic cue weighting in the perception of Scottish tense/lax vowels in Spanish speakers. Unpublished M. Sc. thesis, University of Edinburgh.

Escudero, P., Boersma, P. 2004. Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition* 26. 551–585.

Flege, J. E. 1995. Second language speech learning: Theory, findings and problems. In: Strange, W. (ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press. 233-277.

Flege, J., Schirru, C., Mackay, J. 2003. Interaction between the native and second language phonetic subsystems. *Speech Communication* 40. 467-491.

Hardison, D. 2003. Acquisition of second-language speech: effects of visual cues, context and talker variability. Applies *Psycholinguistics* 24. 495-522.

Hattori, K., Iverson, P. 2009. English /r/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy. *Journal of the Acoustical Society of America* 125. 469-479.

Hazan, V., Sennema, A., Iba, M., and Faulkner, A. 2005. Effect of audiovisual perceptual training on the perception and production of consonant by Japanese learners of English. *Speech Communication* 47. 360-378.

Holt, L. L., Lotto, A. J. 2006. Cue weighting in auditory categorization: Implications for first and second language acquisition. *Journal of the Acoustical Society of America* 119. 3059-3071.

Iverson, P., Evans, B. G. 2007. Auditory training of English vowels for first-language speakers of Spanish and German. *Proceedings of the 16th International Congress of Phonetic Sciences*. Saarbrücken, Germany. 1625-1628.

Iverson, P., Hazan, V., Bannister, K. 2005. Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *Journal of the Acoustical Society of America 118*. 3267-3278.

Jamieson, D.G., Morosan, D.E. 1989. Training new, nonnative speech contrasts: A comparison of the prototype and perceptual fading techniques. *Canadian Journal of Psychology* 43. 88-96.

Logan, J. S., Lively, S. E., Pisoni, D. B. 1991. Training Japanese learners to identify English /r/ and /l/: a first report. *Journal of the Acoustical Society of America 89*. 874-886.

Logan, J. S., Pruitt, J. S. 1995. Methodological issues in training listeners to perceive non-native phonemes. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues*, Timonium, MD: York Press. 351-378.

McAllister, R., Flege, J.E., Piske, T. 2002. The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics* 30 (2). 229-258.

Munro, M. J., Flege, J. E., and MacKay, I. R. A.1996. The effects if age of second language learning on the production of English vowels. *Applied Psycholinguistics* 17. 313-334.

Nishi, K., Kewley-Port, D. 2008. Nonnative speech perception training using vowel subsets: Effects of vowels in sets and order of training, *Journal of Speech, Language, and Hearing Research* 51. 1480-1493.

Ortega-Llebaria, M., Faulker, A., Hazan, V. 2001. Auditory vs. Visual L2 speech perception: effects of visual cues and acoustic-phonetic context for Spanish learners of English. *Proceedings AVSP-2001*. 149-154.

Piske, T., MacKay, I. R. A., Flege, J. E. 2001. Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29. 191-215.

Wang, Y., Jongman, A., Sereno, J. A. 2003. Acoustic and perceptual evaluations of Mandarine tone productions before and after training. Journal of the Acoustical Society of America 113. 1033-1043.

Ylinen,S., Uther,M., Latvala,A., Vepsäläinen,S., Iverson,P., Akahane-Yamada,R.,Näätänen,R. 2009. Training the brain to weight speech cues differently: A study of Finnish second-language users of English. *Journal of Cognitive Neuroscience.* 1-14.

Phonological short-term memory and L2 speech learning in adulthood

Cristina Aliaga-Garcia[1], Joan C. Mora[2], Eva Cerviño-Povedano[3]

Universitat de Barcelona
cristinaaliaga@ub.edu[1], mora@ub.edu[2], ecervino@ub.edu[3]

ABSTRACT

The aim of the present study is to extend research on individual differences in cognitive ability and second language acquisition by exploring the relationship between PSTM and L2 speech learning in advanced EFL learners. The participants (*N*=40) were a subset of a larger group of EFL learners participating in 10 60-minute high-variability phonetic training sessions on the perception of the 11 British English monophthongs. Perceptual accuracy was assessed at pre- and post-test through categorization and discrimination tasks based on 5 monophtongal vowels (/iː ɪ æ ʌ ɑː/). PSTM capacity measures were obtained at post-test using a SNWR task consisting of CVC nonwords in the subjects' L1 (Catalan). The participants were assigned to two PSTM capacity groups according to the scores obtained in the SNWR task (percent correct recognition): LowPSTM vs. HighPSTM, and the perceptual scores and gains these groups obtained were compared. The HighPSTM group was found to obtain higher accuracy scores and greater perceptual accuracy gains than the LowPSTM group. These results suggest that PSTM may play a role in L2 speech learning and may be involved in learners' ability to successfully form phonetic categories for L2 sounds.

Keywords: Phonological short-term memory (PSTM), vowel perception, cue weighting, vowel duration.

1. INTRODUCTION

Phonological short-term memory (PSTM) is one of the subcomponents, together with a sub-vocal articulatory rehearsal system, of the phonological loop, the component of working memory responsible for the processing of verbal-acoustic information (Baddeley 1986, 2003). PSTM is a storage subcomponent that allows speakers to store spoken utterances for a short period of time (2 seconds, approximately) or longer if refreshed through articulatory rehearsal. The acoustic information temporarily stored in PSTM is coded phonologically and has been shown to be sensitive to language-specific distributional properties of sounds, such as syllable frequency (Nimmo and Roodenrys 2002), and overall linguistic development (French and O'Brien 2008). More recent developments of Baddeley's working memory model include an episodic buffer that enables information from long-term memory to be integrated into the PSTM storage (Baddeley 2000).

The role of PSTM in language acquisition is well-established. It has been shown to predict vocabulary acquisition in children (Baddeley et al. 1998) and is related to other aspects of L1 development such as semantics and syntax (Adams and Gathercole 2000). In second language acquisition PSTM predicts children's level of development in the L2 as regards vocabulary (Massoura and Gathercole 1999) and grammar (French and O'Brien 2008). Overall competence in the L2 and oral production skills in children and adults have also been shown to be related to PSTM (French 2006; Kormos and Sáfár 2008). For example, French and O'Brien (2008) examined the role of PSTM in L2 grammar development in Francophone children in a 5-month intensive English programme. PSTM, assessed by Arabic and English nonword repetition tasks, was found to explain 27.9% of the variance in the grammatical knowledge obtained between pre- and post-test. Similarly, O'Brien et al. (2007) investigated the relationship between PSTM, assessed by means of a serial noword recognition (SNWR) task, and the oral fluency development of adult English learners of Spanish spending a semester abroad. PSTM was found to be significantly correlated to oral fluency gains obtained in their stay abroad, explaining 4.5-9.7% of unique variance in their oral fluency gain scores. In general, greater PSTM capacity, as measured by a variety of nonword repetition (NWR) or SNWR tasks, is assumed to grant language learners greater aptitude in speech processing tasks, and it is consequently found to predict greater gains in several areas of linguistic competence. To the best of our knowledge, however, no studies so far have directly examined the relationship between PSTM and L2 speech learning in adults. On the basis of previous research and the role of speech processing mechanisms in speech perception and production, we hypothesized that L2 learners with greater PSTM capacity may be capable of developing more accurate representations for L2 vowels and consonants than L2 learners with poorer PSTM abilities, and as a consequence the former might be found to outperform the latter in terms of the L2 phonological development and L2 pronunciation. The aim of the present study is to extend research on individual differences in cognitive ability and second language acquisition by exploring the relationship between PSTM and L2 speech learning in advanced EFL learners. The study focuses on the relationship between PSTM and the development of L2 perceptual phonological competence.

Cross-language speech perception research within Flege's (1995) Speech Learning Model (SLM) has shown that L1/L2 experience, operationalized as length of residence (LOR) and amount of L1/L2 use, as well as age of onset of L2 learning (AOL), often indexed by age of arrival in the L2 speaking country, affect both L2 speech perception and production, younger AOLs and longer LORs leading to less accented speech (see Piske et al. 2001 for a review). Research findings in immersion settings have generally supported Flege's SLM and one of its main tenets, namely that increased experience may lead late learners to discern phonetic differences between L1 and L2 sounds, thus making the development of new phonetic categories for L2 sounds possible, which will eventually lead to greater accuracy in the production of L2 sounds (Flege et al. 1997). The development of accurate long-term representations for L2 sounds (phonetic categories) may also be affected by learners' individual differences in their ability to process acoustic information. One may speculate that learners with better PSTM abilities, having a larger capacity store for retaining sound sequences in memory, would be more apt than learners with lower PSTM capacity at attending to L2 acoustic cues that are not used contrastively in the L1. For example, the difficulty many Spanish learners of English have in attending to both temporal and spectral cues in the perception of the English tense-lax vowel contrasts and their over-reliance on duration cues, possibly because single-category assimilation may have desensitized them for the perception of spectral cues (Bohn 1995; Escudero and Boersma 2004; Cebrian 2006), may be explained at least in part by learners' differences in PSTM capacity. The only one study we are aware of that has examined the relationship between PSTM and L2 speech perception is Isaachs and Trofimovich (*in press*), who investigated the effect of individual differences in PSTM, attention control and musical ability and listeners' judgements of L2 speech for accentedness, comprehensibility and fluency. Neither attention control nor PSTM significantly affected listeners' ratings. In the present study we investigated the effect of individual differences in PSTM on EFL learners' perceptual phonological competence development through phonetic training.

2. METHOD

Regardless of the challenge that L2 speech learning poses for L2 learners, particulary in the mastery of L2 vowels, several studies conducted so far have established that it is possible to achieve improvement in the perception and production of L2 sounds though phonetic training (Jamieson and Morosan 1989; Logan and Pruitt 1995; Iverson and Evans 2007; Ylinen et al. 2009), which has been shown to be most effective when including high variability stimuli in the training set (Hazan et al. 2005; Iverson et al. 2005), especially when the difficulties with L2 sounds lie in the use of L2 phonetic cues which are not used or are weighted differently in the L1 (Cebrian 2006; Escudero and Boersma 2004; Holt and Lotto 2006). The aim of the study was twofold. First, we investigated whether high-variability training resulted in higher accuracy in L2 vowel identification and discrimination. Secondly, we explored the relationship between the phonological short-term memory (PSTM) abilities of the participants and their gains obtained through vowel training. This was done to confirm our hypotheses that PSTM would be positively related to the development of L2 vowel perception.

Measures of perceptual phonological competence in the participants' L2 (English) were obtained before (pre-test) and after (post-test) 10 sessions of high-variability phonetic training on the perception of British English monophthongs. The perception tasks consisted of a vowel identification (labelling) task with CVC word stimuli, a forced-choice vowel categorization task in a /h_d/ context based on 7-step duration continua, and an AX discrimination task. A SNWR task was used to obtain a measure of the participants' PSTM capacity. In order to test whether PSTM was related to the perceptual phonological measures, the participants were assigned to Low vs. High PSTM groups and differences between the two groups were examined for the perception scores obtained at pre- and post-test. The relationship between PSTM capacity and the perceptual gains obtained through the training was explored by calculating residualized gain scores through regression of the pre-test perceptual accuracy scores on the post-test scores and comparing these between the Low and the High PSTM groups.

2.1. Participants

The participants for this study were a subset (*N*=40) of a larger group of advanced EFL learners from the University of Barcelona enrolled in a Phonetic Training program (*mean age:* 23; *range:* 20-50; 4 males and 36 females). All participants were Catalan-Spanish bilinguals who had learned English in a formal instruction setting in Catalonia. They were taking a degree in English studies and all had at least one prior semester of formal study of English Phonetics, but had not studied English abroad for longer than 2 months. All reported having normal hearing and having no speech-related dysfunctions.

2.2. High-variability Phonetic Training

The audiovisual training corpus comprised 163 CVC real words containing the whole set of English monophthongal vowels (/iː ɪ e ɜː æ ʌ ɑː ɒ ɔː ʊ uː/) in high contextual variation (b_d, b_t, p_k, p_t, m_d, f_l, t_n, t_k, d_n, d_d, s_d, l_n, k_n, w_t, h_l) pronounced by 10 different talkers (5 male, 5 female) of Southern British English pronunciation. The stimuli words were read at normal speed on a falling intonation and videorecorded using a Canon XL-1 DV camrecorder, and a Bruel and Kjaer type 4165 microphone, in an acoustically attenuated chamber in the Phonetics Laboratory at University College of London. the words were read one ata time from a computer screen to avoid list-reading intonation. The talker's face was set against a blue-background and illuminated with a key and a fill light, and the resulting videoclips were edited so that the start and end frames of each token, lasting 3 seconds, showed a neutral facial expression.

The audiovisual vowel training was run using DmDx display presentation software, which guaranteed exposure to a minimum of 176 trials (4 tokens x 11 vowels x 4 repetitions or blocks) and included three subsets of vowels (high front vowels /iː ɪ e/, high back vowels /ɔː ʊ uː/, and low vowels /æ ʌ ɑː ɒ/) as pronounced by 2 different talkers (i.e. NS1 blocks 1 and 3; NS2 blocks 2 and 4). Participants received immediate feedback over headphones after each response until they corrected themselves (maximum 4 trials). Learners participated in two 60-minute training sessions per week, during which they heard and watched short video-clips showing the native speaker's mouth region and chose one among 3 or 4 alternatives. After errors, they heard the wrong response immediately followed by the correct word so that they could try again. They could also focus on the articulatory gestures for more accurate vowel articulation and compare their own responses with native speakers' as many times as necessary.

2.3. Pre- and Post-test

Subjects were pre- and post-tested on their identification of natural and cue-manipulated stimuli, and on natural vowel discrimination. The three perceptual tasks were computer-administered (DmDx), in a quiet computer room in groups. The subjects listened to the stimuli individually over headphones and gave their responses by selecting buttons appearing on the computer screen. Each perception task was preceded by a short training phase to familiarize participants with the range of possible identification and discrimination responses. On average it took subjects about 50 minutes to complete the three tasks, but they were allowed to stop and rest if necessary. A total of 107 CVC testing stimuli, 50% of which had been included in the training, were recorded by 10 novel SBE talkers (5 male, 5 female), different from the ones appearing in the training, following the same recording procedures used for the training stimuli (cf. Section 2.2.) and the same segmentation and video-removal edition process.

2.3.1. Identification task I

Subjects first completed a multiple-choice identification task with 120 CVC natural words distributed into 2 blocks (6 tokens x 5 vowels x 2 repetitions x 2 blocks/conditions). There were 2 conditions: (1) audiovisual

(AV) vs. only-auditory (A) stimuli presentation (2 orders: A-AV, AV-A), and (2) trained vs.new words. The task consisted of 15 trained and 15 untrained tokens in varied contexts produced by 2 SBE native speakers (1 male, 1 female) different from the ones in the training. The video data was converted into sound files and edited for the only-auditory condition, so that each stimulus had two versions, auditory vs. audiovisual. Subjects were required to watch a 3-second videoclip of the talkers' mouth, or hear the stimuli, and then click on one response button to select one word out of 2-3 options displayed horizontally (i.e. *cat-cut-cart*).

2.3.2. Identification task II

Following identification testing for natural stimuli, subjects completed a forced-choice identification task with 210 synthesized /hVd/ stimuli (6 vowels x 7 steps x 5 repetitions) based on 6 7-step duration continua (80, 116.7, 153.3, 190, 226.7, 263.3, 300 ms) which were created from the vowel pairs /hiːd/-/hɪd/, /hæd/-/hʌd/ and /hʌd/-/hɑːd/, respectively, using Praat software in an attempt to assess the learners' vowel perception when the duration cue was ambiguous (i.e. shortened /i/ vs. lengthened /ɪ/). In this identification task participants heard one cue-manipulated stimuli at a time and had to identify one of the two responses on the screen (one on the left, another on the right: i.e. *he'd* vs. *hid*).

2.3.3. Discrimination task

The AX discrimination task required subjects to discriminate 128 CVC vowel pairs (25% false-alarm trials were included) based on four natural vowel contrasts (/iː/-/ɪ/, /æ/-/ʌ/, /ʌ/-/ɑː/, /æ/-/ɑː/) and distributed in two blocks (2 orders: AB vs. BA) with different testing conditions (fixed CVC context vs. context variability) and untrained talker variability (8 different "new" SBE talkers: 4 male, 4 female). Participants heard one minimal pair at a time and labelled it as *same* or *different.*

2.4. SNWR task

A serial nonword recognition (SNWR) task was employed to obtain a measure of PSTM. The reason for choosing a SNWR task over other methods widely used in the literature, such as serial nonword recall and nonword recognition (e.g. French and O'Brien 2008), is that it avoids the effect of the articulatory component on the PSTM scores (Snowling et al. 1991) and minimizes lexical knowledge effects (Gathercole et al. 2001). SNWR tasks have also been employed as a measure of PSTM in recent research on individual differences in cognitive ability and adult L2 oral production (O'Brien et al. 2006, 2007) and perception (Isaachs and Trofimovich *in press*).

This task contained 144 pronounceable CVC nonwords conforming to Catalan syllable structure phonotactics. Catalan nonwords were used to avoid possible effects of lexical knowledge on the PSTM measure. All subjects used Catalan on a daily basis. Therefore, it was assumed that differences in the PSTM measures could not be attributed to differences in their knowledge of Catalan (O'Brien et al. 2006). Several realizations of the nonwords, embedded in carrier phrases (e.g. '*Rima amb* ***s/ɛ/t****, ara dic* ***f/ɛ/k****. Ara dic* ***f/ɛ/k*** *un cop*') were elicited from a female native speaker of Catalan, who read the carrier phrases at normal speed in a sound-proof booth. The phrases were digitally recorded with a Marantz PMD660 recorder and a unidirectional dynamic microphone (ShureSM58) and were subsequently segmented and edited. The best token of each nonword was chosen for the SNWR task, which contained 144 nonwords distributed in 24 nonword sequence pairs (8 sequence pairs each at 5-, 6- and 7-nonword lengths). DmDx was used to run the SNWR task and record subjects' responses. Nonwords were presented via headphones at a rate of 750ms, sequences within a pair were separated by a 1500ms silence and sequence pairs were presented after a 500ms delay following the subject's response. The subjects were instructed to decide whether the nonwords in one sequence were presented in the same or a different order in the following sequence in the pair and responded by pressing one of two keyboard keys. The task was preceded by a short familiarization phase of 2 same-order and 2 different-order sequence pairs. Nonword sequences were built so that in a given sequence all CVC nonwords contained a different vowel and as many different consonants as possible. In different-order pairs, the first and last nonword never changed position. Transposed nonwords randomly varied in position. Half of the nonword sequence pairs, which were randomly presented within 5-, 6- and 7-nonword length blocks, were different-order sequence pairs. The number of correctly identified same/different nonword sequence pairs was used as a measure of PSTM (Isaachs and Trofimovich *in press*).

3. RESULTS AND CONCLUSION

The participants' perceptual phonological competence at pre- and post-tests was assessed by computing mean percent correct vowel identification and discrimination scores for each subject, vowel type (/iː/, /ɪ/, /æ/, /ʌ/, /ɑː/), vowel condition (natural vs. synthesized) and vowel contrast (/iː/-/ɪ/, /æ/-/ʌ/, /ʌ/-/ɑː/, /æ/-/ɑː/), at pre- and post-test. For each vowel identification (ID) and discrimination (DIS) measure, two types of gains were computed: (1) gains obtained by substracting pre-test from post-test performance, and (2) residualized gain scores by regressing post-test performance on pre-test performance. We next wished to see whether PSTM performance was related to the development of perceptual phonological competence, following a 10-session training period. The participants were assigned to two PSTM capacity groups through median split, LowPSTM (mean: 54%, range: 37.5-62.5; N=20) vs. High PSTM (mean: 78.95%, range: 66.67-95.83; N=20), and both pre-test and post-test scores and gains obtained by each of the groups were compared. Participants with higher levels of PSTM ability were predicted to identify and discriminate vowels with higher accuracy, especially at post-test, and were expected to obtain greater gains after training than those with poorer PSTM ability.

Paired-samples t-tests on T1-T2 scores revealed a significant improvement in the learners' ability to identify natural and synthesized vowels, as well as significantly higher scores in vowel discrimination (see Table 1), which confirmed that Catalan/Spanish subjects improved significantly following the 10-session high-variability phonetic training. They obtained 10.61% gains in mean correct ID of natural vowels, 6.20% in mean correct ID of synthesized vowels and 3.09% gains in mean vowel contrast DIS. Independent-samples t-tests revealed differences in overall L2 vowel perception following training as a function of PSTM capacity for some of the perception measures. High PSTM learners outperformed Low PSTM learners consistently for all ID and DIS measures, suggesting that perceptual ability may be partly explained by learners' differences in PSTM.

Table 1: Mean vowel ID and DIS scores of natural (N) and synthesized (S) stimuli at pre- (T1) and post-test (T2).

Vowel perception measures	Entire cohort (N=40)			Low PSTM (N=20)			High PSTM (N=20)			Low vs. High		
	T1	T2	diff.	T1	T2	gains	T1	T2	gains	T1	T2	gains
N_Vowel ID	**67.95**	**78,58**	***.000**	65.93	85.75	10.32	69.87	81.23	10.18	-	-	-
nat. /iː/, /ɪ/	**66.39**	**77.36**	***.000**	64.12	75.33	11.23	68.53	79.28	10.75	-	-	-
nat. /æ/, /ʌ/, /ɑː/	**68.99**	**79.39**	***.000**	67.13	76.02	8.95	70.76	82.53	11.77	-	-	-
S_Vowel ID	**76.28**	**82.48**	***.000**	**72.88**	**78.25**	**5.36**	**79.67**	**86.72**	**7.04**	***.036**	***.004**	-
syn. /iː/, /ɪ/	71.99	80.30	.069	**65.71**	74.21	8.50	**78.27**	86.39	8.12	***.051**	-	-
syn. /æ/, /ʌ/, /ɑː/	**78.42**	**83.57**	***.000**	76.47	**80.26**	3.80	80.38	**86.88**	6.50	-	***.034**	-
Vowel DIS	87.45	90.54	.095	89.20	90.75	1.20	85.7	90.97	4.99		-	-
/iː/-/ɪ/	**73.74**	**76.47**	***.001**	**69.05**	76.32	5.26	**78.43**	76.85	7.42	***.029**	-	-
/æ/-/ʌ/, /ʌ/-/ɑː/, /æ/-/ɑː/	92.02	95.23	-	92.78	95.56	2.24	91.25	95.68	4.17	-	-	-

Learners showed overall improvement on all the vowel perception measures between T1 and T2 and t-tests revealed that high PSTM outperformed low PSTM on these measures, both at T1 and T2 (see Table 1), suggesting that PSTM may be associated with different improvement rates. Both at pre- and post-test, the High PSTM group identified natural and synthesized vowels at higher accuracy rates than the Low PSTM group, even for synthesized vowels with extremely short/long durations at the two ends of duration continua. The High PSTM group's better perceptual ability may be explained by their higher memory ability granting them an advantage in the development of accurate long-term representations for L2 vowels. Independent-samples t-tests failed to reveal overall robust significant differences between residualized change scores obtained for the Low and High PSTM groups. However, some significant and near-significant differences in favour of the high PSTM group for vowel ID (synthesized lengthened /ʌ/, p=.02; synthesized shortened /iː/,

p=.07; /æ/ p=.07 and /ɑː/ p=.06, emerged. Although the groups did not statistically differ on residualized gains for all vowel perception measures, the high PSTM group showed a general trend toward greater *gains* operationalized as both *mean subtracted T2-T1 differences* and *residualized change scores*, as compared to the low PSTM group.

The present study demonstrated that PSTM is implicated in perceptual phonological competence development through phonetic training in adulthood. The Low PSTM group scored lower than the High PSTM group before training, which points to the potential role of PSTM in L2 phonological acquisition. After training, the low and high PSTM groups showed increasingly different patterns of results, suggesting that PSTM may contribute significantly to the development of L2 speech perception.

4. REFERENCES

Adams, A. M., Gathercole, S. E. 2000. Limitations in working memory: Implications for language development. *International Journal of Language and Communication Disorders* 35. 95–116.

Baddeley, A. 1986. *Working memory*. Oxford: Clarendon Press.

Baddeley, A. D. 2000. The episodic buffer: A new component of working memory? *Trends in cognitive sciences* 4. 417-423.

Baddeley, A. D. 2003. Working memory and language: An overview. *Journal of Communication Disorders* 36. 189-208.

Bohn, O. S. 1995. Cross-language speech perception in adults: First language transfer doesn't tell it all. In: Strange, W. (ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press. 279-302.

Cebrian, J. 2006. Experience and the use of duration in the categorization of L2 vowels. *Journal of Phonetics* 34. 372-387.

Escudero, P., Boersma, P. 2004. Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition* 26. 551–585.

Flege, J. E. 1995. Second language speech learning: Theory, findings and problems. In: Strange, W. (ed.), *speech perception and linguistic experience: issues in cross-language research*. Timonium, MD: York Press. 233-277.

Flege, J. E., Bohn, O. S., & Jang, S. 1997. Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics* 25. 437-470.

French, L. M. 2006. *Phonological working memory and L2 acquisition: a developmental study of Quebec francophone children learning English*. New York: Edwin Mellen Press.

French, L., O'Brien 2008. Phonological memory and children's second language grammar learning. *Applied Psycholinguistics* 29. 463-487.

Gathercole, S. E., Pickering, S. J., Hall, M., Peaker, S. M. 2001. Dissociable lexical and phonological influences on serial recognition and serial recall. *Quarterly Journal of Experimental Psychology* 54(A). 1-30.

Hazan, V., Sennema, A., Iba, M., Faulkner, A. 2005. Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication* 47. 360-378.

Holt, L. L., Lotto, A. J. 2006. Cue weighting in auditory categorization: Implications for first and second language acquisition. *Journal of the Acoustical Society of America* 119. 3059-3071.

Isaacs, T., Trofimovich, P. (in press). Phonological memory, attention control, and musical ability: Effects of individual differences on rater judgments of L2 speech. *Applied Psycholinguistics*.

Iverson, P., Evans, B. G. 2007. Auditory training of English vowels for first-language speakers of Spanish and German. *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1625-1628). Saarbrücken, Germany.

Iverson, P., Hazan, V., Bannister, K. 2005. Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *Journal of the Acoustical Society of America118*. 3267-3278.

Kormos, J., Sáfár, A. 2008. Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition* 11. 261-271

Logan, J. S., Pruitt, J. S. 1995. Methodological issues in training listeners to perceive non-native phonemes. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues*, Timonium, MD: York Press 351-378

Nimmo, L. M, Roodenrys, S. 2002. Syllable frequency effects on phonological short-term memory tasks. *Applied Psycholinguistics* 23. 643-659.

O'Brien, I., Segalowitz, N., Collentine, J., Freed, B. 2006. Phonological memory and lexical, narrative, and grammatical skills in second language oral production by adult learners. *Applied Psycholinguistics* 27. 377-402.

O'Brien, I., Segalowitz, N., Freed, B., Collentine, J. 2007. Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition* 29. 577-582.

Piske, T., MacKay, I., Flege, J. E. 2001. Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29. 191-215.

Snowling, M., Chiat, S., Hulme, S. 1991. Words, nonwords, and phonological processes: Some comments on Gathercole, Willis, Emslie & Baddeley. *Applied Psycholinguistics* 12. 369-373.

Ylinen,S., Uther, M., Latvala, A., Vepsäläinen, S.m Iverson, P., Akahane-Yamada, R., Näätänen, R. 2009. Training the brain to weight speech cues differently: a study of Finnish second-language users of English. *Journal of Cognitive Neuroscience* 1-14.

Influential Factors on the production of English /θ/ by Japanese Learners of English

Rika Aoki

University of Tokyo
galangalanga@gmail.com

ABSTRACT

This study investigates the linguistic and sociolinguistic factors which trigger high frequency of /s/-substitution (i.e. the substitution of /s/ for /θ/) by Japanese-speaking learners of English. From the production experiment participated by 8 Japanese learners, it was found that /s/-substitution was more likely to occur before high vowels in a longer word with an initial syllable stress, with /s/ in proximity, and in more spontaneous speech. The results of this study could be applied to pedagogical strategies to help Japanese learners overcome their pronunciation difficulties more efficiently.

Keywords: /s/-substitution, L2 production, adjacent segments, stress, task type.

1. INTRODUCTION

It has been suggested that English /θ/ is highly problematic for Japanese-speaking learners of English, because /θ/ is absent from their first language (L1). Due to this fact, it is often the case that Japanese learners assimilate /θ/ to their native category, /s/ (/s/-substitution), which means that minimal pairs such as *think-sink, thick-sick,* and *path-pass* become homophones (Takebayashi et al, 1991: 90). Although a large number of previous studies have shown this fact, the exact factors which influence /s/-substitution have not fully investigated. Therefore, with the aim of finding the contexts which trigger /s/-substitution, this study is intended to examine the production of English /θ/ in various contexts by native speakers of Japanese.

2. FACTS ABOUT ENGLISH /θ/

English /θ/ is a voiceless dental fricative produced by the tip of the tongue against the upper teeth. According to Ladefoged (2006), while British people tend to pronounce it as a dental sound, putting the tongue "close behind the upper front teeth" (Ladefoged 2006: 10), Americans often produce it as an interdental sound by having "the tip of the tongue protruding between the upper and lower front teeth" (Ladefoged 2006: 10).

Acoustically, it is often claimed that /θ/ is similar to /f/. In fact, according to Takebayashi (1996), there are some native speakers of English (especially children) who use [f] instead of /θ/ (e.g. *I had only free* [=*three*]) (Takebayashi 1996: 200). Black English speakers and Cockneys are also known to pronounce /θ/ as [f] (Wells, 1982). Moreover, the pronunciation of this sound tends to be greatly varied with individuals (Ladefoged, 2006).

As in many languages other than English, /θ/ does not exist in Japanese phonemic inventory. This leads Japanese learners to replace English /θ/ with [s], and to have difficulty in distinguishing /θ/ from /s/ (Takebayashi and Saito, 1994: p.94), in spite of the facts that in Tosa dialect, *Tosa* (a geographical name) is pronounced as [toθa] (Hattori 1984: 79), and also the increasing number of Japanese have pronounced /s/ as [θ] or its variants (Kazama et al, 2007: 226).

Several previous studies such as Ritchie (1968) and Takebayashi (1996) suggest that there are two typical phones used for substitution of /θ/: [s] and [t]. While Japanese and French speakers often substitute [s] for /θ/, German, Russian, Thai, Tagalog, and Turkish speakers tend to replace /θ/ with [t]. There could be two possible explanations for these different types of substitution, as Flege et al (1995) stated. First, it may be because of the difference in prominence of a certain feature of segments in one's second language (L2). For example, for Japanese, the feature of non-stridency would be more prominent than that of continuancy in assimilating /θ/ to their L1 category. Second, the frequency of a certain feature would decide the type of substitution.

According to Jenkins (1996), the substitutions of /θ/ with [t] or [s] are so common to non-native speakers of English that they do not cause a serious unintelligibility in interactions not only between native and non-native speakers of English, but also between non-native speakers. However, there also exists a report from Gimson and Cruttenden (1994), suggesting that the replacement of /θ/ with [t] may impede the intelligibility.

3. PREVIOUS STUDIES

In this chapter, the previous studies on linguistic and sociolinguistic factors which are claimed to have an impact on the pronunciation of /θ/ by English learners are introduced.

3.1. Linguistic factors

According to Tarone (2007), linguistic contexts would include stress placement, segment position, and cognitive status of the referential form. For example, Lambacher et al (2001) investigated the effect of surrounding vowels on the English consonant perception by Japanese learners. They found that among the five consonants, /f s ʃ θ h/, the correct response rate was the lowest for /θ/ (55%), and the confusability rate was the greatest between /θ/ and /s/. In addition, "the learners had the most difficulty identifying /θ/ when spoken with the /e/" (p.336). It was also worth noting that Japanese and native speakers of English were common in having difficulty distinguishing the /f/-/θ/ contrast before /o/ or /e/ in the context of CV (consonants followed by vowels). Therefore, they concluded that the consonant perception by Japanese learners is influenced "by both vowel environment and consonant position" (p.342).

As for the study on non-Japanese learners of English, Wester et al (2007) investigated the substitution of English /θ/ by Dutch learners. In their study, the authors found that instead of /θ/, [t] was used most frequently in syllable-initial position, while [s] occurred most in syllable-final position. Thus, as the results of Lambacher (2001) introduced above, it can be concluded that the substitution of /θ/ by Dutch speakers is affected by consonant position.

3.2. Sociolinguistic factors

Various kinds of sociolinguistic factors have been examined by several researchers. For example, Purcell and Suter (1980) concluded that the reliable predictors of L2 learners' accents were L1, aptitude for oral mimicry, residency and attitude. In addition, Moyer (2004) considered the following factors as important: L2 experience, L1, aptitude, motivation, frequency of L1 and L2 use, the amount of L2 input, social identity and the kinds of instruction.

Among these several factors, L2 experience has been examined most frequently and profoundly by many studies. Usually, L2 experience is described by two criteria: Age of Learning (AOL) and Length of Residence (LOR). The former refers to learners' age of first exposure to the L2, while the latter "specifies the number of years spent in a community where the L2 is the predominant language" (Piske et al, 2001, p.192). As for AOL, many studies such as Piske et al (2001) and Munro (1996) found that the later the L2 learners began to learn English, the less accurately they produce or perceive English sounds, which means that AOL has a significant impact on the pronunciation of L2 learners. On the other hand, as concerns LOR, Flege et al (2006) stated that L2 learners whose LOR was rather long were able to perceive and produce spectral difference of L2 sounds in the same way as native speakers while those whose LOR was short received lower ratings. This result suggests that LOR could also be a predictor of the degree of L2 accent. Comparing AOL and LOR, however, Piske et al (2001) argued that "LOR was not identified as a significant predicator of degree of L2 foreign accent" (p.198) in the multiple regression analysis of several studies.

The results of other studies found the effect of the sociolinguistic factors than L2 experience: task type (Dickerson, 1975; Beebe, 1980; Strange et al, 2001), frequency of L1 or L2 use (Piske et al, 2001). Besides, Wenk (1979), who studied the production of /θ/ by French, observed that the substitution type depended on one's proficiency and style. As another example, Gatbonton (1977) insisted that the production of /θ/ by French Canadians changed with the strength of "ethnic identification," in other words, the degree of which one's political attitude is nationalistic.

Thus, it has been suggested that both linguistic and sociolinguistic factors can change the production and the perception of learners of English. However, most studies focus on a specific factor, and therefore a comprehensive investigation has not been conducted yet. Therefore, in this study, a research question which will be found in the next section was approached.

4. THE EXPERIMENT

In this chapter, a research question of the current study and its related variables are discussed, followed by the method used in the experiment.

4.1. Research question

A research question of this study is following: what are the linguistic (and sociolinguistic) factors that influence the frequency of /s/-substitution of English syllable-initial /θ/ produced by Japanese-speaking learners of English?

As Lambacher (1999) suggests, in the pronunciation education and research of recent years, the more attention has been paid to the suprasegmental features compared to the segmental ones. While it is indeed that the suprasegmental features are essential for intelligibility, the segmental features (i.e. segmental contrasts) can also interfere intelligibility (Lambacher, 1999: 138). Therefore, this study contributes to decrease communicative problems caused by learners' failure in contrasting L2 sounds.

Here, among many English consonants, /θ/ has been chosen as a dependent variable. This was because the production /θ/ has not been systematically investigated compared to other consonants such as /r/ and /l/, although /s/-substitution by Japanese learners of English has been indicated for a long time. In addition, it is widely known that many Japanese learners often take considerable troubles in acquiring an accurate /θ/ pronunciation. Therefore, to explore and specify influential factors on the production of /θ/ could be a help for Japanese to produce /θ/ correctly, because it might enable not only learners but also teachers to concentrate in some specific features which are necessary for an accurate pronunciation. Thus, this study may contribute to the better pronunciation of /θ/ by Japanese-speaking learners of English.

As independent variables, 4 linguistic and 1 sociolinguistic factors were chosen. The first linguistic factor was Stress Placement: if /θ/ is located in a syllable-initial σ (e.g. *thesis*) or other σ (e.g. *thesaurus*). Second, Following Environment was considered: if the following phonetic status of /θ/ was a high vowel (e.g. *think*), a non-high vowel (e.g. *thought*), or a consonant /r/ (e.g. *thrill*). The third linguistic factor was Word Size: if a target word consisted of 1 syllable (e.g. *three*) or 2 and more syllables (e.g. *theory*). Finally, Proximity of Similar Sounds was taken into account, that is, if /s/ existed within a word (e.g. *thirsty*), or not (e.g. *through*). As for a sociolinguistic factor, Style was introduced, which was represented by a word list reading and a short interview.

4.2. Method

In this section, participants, a measure to collect and analyse data are briefly explained.

4.2.1. Participants

There were eight Japanese-speaking learners of English who took part in the present study. All participants were undergraduate or graduate students studying in Tokyo, for the purpose of controlling their age and length of English education at schools. Although their language backgrounds were varied and their English proficiency was not uniform, all participants were considered that their academic interests were related to English. Their LOR was fixed as 0.

4.2.2. Data collection

As mentioned in the last section, there were two kinds of methods of collecting learners' production of /θ/: a read-aloud task of words in carrier sentences, *"I say ___ on the tape."*, a question-and-answer session on various kinds of topics (See Ch.8). The purpose of the recording was concealed until the end of the recording session, in order to exclude influences on the pronunciation caused by the subjects' consciousness of the project aim.

The recording was carried on in the soundproof booth of the University of Tokyo. A digital audio recorder (Cool Edit Pro version 1.2) and a dynamic microphone (SONY ECM-MS957) were used, which were the accessories of the booth. The audio signals from each talker were digitized at 44.10 kHz with 16 bits of amplitude resolution.

4.2.3. Data analysis

The existence/absence of /s/-substitution was judged by a trained phonetician, with an aid of the spectrograms, produced by sound analysis program Praat version 5.0.35 (Boersma & Weenink, 2008). After the decision of /s/ existence, a statistic program GoldVarb X (Robinson et al, 2001) was used in order to determine the weight which a given factor contributes to the probability of /θ/ being substituted with [s] by variable rule analysis.

5. RESULT

The result of the present study suggested that the factors considered were influential on /s/-substitution by Japanese learners of English, and it was more likely to occur before high vowels in a longer word with an initial syllable stress, with /s/ in proximity, and in more spontaneous speech (Table 1).

Table 1: Independent variables and their weights in this experiment (weight: 0~1).

Factor		Factor weight	
Stress Placement	syllable-initial σ	0.934	☜
	other σ	0.066	
Following Environment	high vowel	0.623	☜
	non-high vowel	0.115	
	consonant /r/	0.262	
Word Size	1 syllable	0.262	
	2 + more syllables	0.738	☜
Proximity of /s/	no /s/ in proximity	0.148	
	/s/ in proximity	0.852	☜
Style	word list	0.443	
	interview	0.557	☜

Table 1 shows each independent factor adopted in the present study, and their factor weights, that is, the degree to which a factor was considered to have an impact on the appearance of /s/-substitution of English syllable-initial /θ/. The weight value can vary from 0 (no effect) to 1 (strong effect), and empirically the values more than 0.5 are interpreted as heavy enough to affect dependent variables. For example, as for Stress Placement, the factor weight of syllable-initial syllable was much heavier than that of other syllable, which means that /θ/ was substituted with /s/ more often syllable-initially than in non-syllable-initially. Similarly, among the three Following Environment factors, a high vowel was judged as the most influential factor on /s/-substitution. The difference of the weight was large in Word Size and Proximity of /s/: in the former factor class, disyllables and longer words elicited /s/ more often; in the latter factor group, more /s/ were observed in the words containing /s/. Finally, although small the difference is, Japanese learners produced /θ/ with /s/ more frequently in a question-and-answer session than in wordlist reading.

6. DISCUSSION

In the result of the present study, all the factors considered in this study were found to be as significant predictors of /s/-substitution of Japanese learners of English, and the specific contexts which are likely to elicit /s/ -substitution were identified. Based on the result of this study, pedagogical application might be possible: in teaching the pronunciation of /θ/, activities should target problematic contexts where /s/-substitution is more frequent.

Phonological explanations can be added in order to justify the linguistic factors. For example, as for Stress Placement, stressed (i.e. more prominent) /θ/ tends to be replaced by more unmarked candidate /s/, which agrees with the theory of positional markedness. Besides, with regard to Following Environment, /θ/ followed by higher (i.e. more prominent) vowels is substituted by more unmarked candidate, which also accords with positional markedness. Moreover, for Proximity of /s/, it can be said that the adjacency of two similar but not same sounds tends to be avoided, which represents the feature of the Obligatory Contour Principle. A sociolinguistic factor (i.e. Style) might be able to be accounted for by the different amount of attention or memory toward pronunciation, because it has an impact on the process of language in terms of how much attention it demands, as Kormos (1999) suggests.

There still remain limitations due to the small number of participants and the immaturity of the experimental material. First, English proficiency of some participants was so high that they seldom produce /s/-substitution. Second, the carrier sentence in the wordlist task occasionally avoided from producing /θ/ correctly (e.g. in the speech of one participant, /s/ in *say* in carrier sentence and /θ/ in the target words were metathesized frequently), which suggests that the words used in the carrier sentences should have been chosen more carefully. It could also affect the result that the fact that all target words contained /θ/ enabled the participants to notice the aim of this experiment. Finally, in future studies, perception test should be conducted in order to ascertain if the pattern of the effectiveness of each factor would be the same as in production.

7. REFERENCES

A Beebe, L. (1980). “Sociolinguistic variation and style-shifting in second language acquisition.” In *Language Learning, 30: 433-447.*

Best, C. T. (1995). “A direct-realist view of cross-language speech perception.” In W. Strange (ed.), *Speech perception and linguistic experience: Theoretical and methodological issues: 171-204*.

Dickerson, L. (1975). “The learner’s interlanguage as a system of variable rules.” In *TESOL Quarterly, 9: 401-407.*

Flege, J. E. (1995). “Second-Language Speech Learning: Theory, Findings and Problems.” In W. Strange (ed.) *Speech Perception and Linguistic Experience: Issues in Cross-Language Research: 233-273.*

Segalowitz, N. & Gatbonton, E. (1977), 'Studies of the non-fluent bilingual', in P.A. Hornby (ed). Bilingualism: Psychological, Social, and Educational Implications. New York: Academic Press, Inc. (pp 77-89).

Gimson, A. C., and A. Cruttenden. (1994). *Gimson’s Pronunciation of English*. Edward Arnold: London.

Hattori, S. (1984). *Onseigaku [Phonetics]*. Iwanami Shoten : Tokyo.

Jenkins, J. (1996). “Changing pronunciation priorities for successful communication in international contexts.” In *Speak out! No. 17: 15-22.*

Kazama, K., Z. Uwano, K. Matsumura, and K. Machida. (2007). *Gengogaku [Linguistics] 2nd ed*. Tokyo Daigaku Shuppankai : Tokyo.

Kormos. J. (1999). “The effect of speaker variables on the self-correction behaviour of L2 learners.” In *System, Volume 27, Issue 2: 207-221.*

Ladefoged, P. (2006). *A Course in Phonetics 5th ed.* Heinle.

Lambacher, S. (1999). “A CALL Tool for Improving Second Language Acquisition of English Consonants by Japanese Learners.” In *Computer Assisted Language Learning, Vol. 12, Issue 2: 137-156.*

Lambacher, S., W. Martens, B. Nelson, and J. Berman. (2001). Identification of English voiceless fricatives by Japanese listeners: "The influence of vowel context on sensitivity and response bias." In *Journal of Acoustic Science and Technology, Vol. 22, Issue 5: 334-343*.

Moyer, A. (2004). *Age, accent and experience in second language acquisition.* Multilingual Matters: Clevedon.

Munro, M., J. E. Flege, and I. R. A. MacKay. (1996). “The effect of age of second-language learning on the production of English vowels.” In *Applied Psycholinguistics, 17: 313-334*.

Piske, T., I. R. A. MacKay, and J. E. Flege. (2001). “Factors affecting degree of foreign accent in an L2: a review.” In *Journal of Phonetics, 29: 191-215*.

Purcell, E. and R. Suter. (1980). “Predictors of pronunciation accuracy: A reexamination.” In *Language Learning, 30: 271-344*.

Ritchie, W. C. (1968). "On the explanation of phonic interference." In *Language Learning, Vol. 18: 449-464*.

Strange, W., R. Akahane-Yamada, R. Kubo, S. A. Trent, and K. Nishi. (2001). “Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners.” In *Journal of the Acoustical Society of America, 109: 1691-1704*.

Takebayashi, S. (1996). *Eigo Onseigaku [English Phonetics]*. Kenkyusha: Tokyo.

Takebayashi, S. and H. Saito. (1998). *Eigo Onseigaku Nyuumon [Introduction to English Phonetics]*. Taishukan Shoten: Tokyo.

Takebayashi, S., M. Watanabe, A. Shimizu and H. Saito. (1991). *Shokyuu Eigo Onseigaku [Basic English Phonetics]*. Taishukan Shoten: Tokyo.

Tarone, E. (2007). “Sociolinguistic Approaches to Second Language Acquisition Research – 1997-2007.” In *Modern Language Journal, Vol. 91, Issue 1: 835-848*.

Wells, J C. (1982). *Accent of English 1: An Introduction.* Cambridge University Press: Cambridge.
Wenk, B. (1979). "Articulatory setting and de-fossilization." In *Interlanguage Studies Bulletin 4: 202-20.*
Wester, F., D. Gilberts and W. Lowie. (2007). "Substitution of dental fricatives in English by Dutch L2 speakers." In *Language Science Vol. 29: 477-491.*

8. APPENDIX

Questions and figures used in the question-and-answer session.
Q1. How many laughing ducklings are there in this picture?
Q2. What does this remind you of?
Q3. How much is this lobster?
Q4. What is the boy with the mask doing now?
Q5. What do you think about this woman?
Q6. What do you think about mathematics?
Q7. Have you ever experienced gender discrimination?
Q8. What would you do if you were in a poor, developed country and discriminated?

Conditions for over-riding the L1 phonological filter

John Archibald

Department of Linguistics, University of Calgary
John.archibald@ucalgary.ca

ABSTRACT

In this paper, I argue that the deficit model of second language acquisition as argued by Brown (2000) and, for syntax/morphology, Hawkins & Hattori (2006) is not empirically supported when we look at a range of data concerning the acquisition of L2 contrasts which are based on phonological features not active in the L1. Drawing on the notion of robust phonetic cues (Wright (2001) we can demonstrate that certain input elements will be become intake to the phonological processor before other elements and will, hence, be acquired earlier. Absence of a particular phonological feature in the L1 does not result in impaired L2 phonological representations. Furthermore, phonological structures which are absent from the L1 *can* be acquired in the L2.

Keywords: L1 phonological filter, acquiring new structure, L2 laryngeal features.

1. INTRODUCTION

The field of cross-linguistic speech perception attempts to answer the question of why some second language (L2) sounds are easier to acquire than others. For example, the difficulties of Japanese learners of the English [r] have oft been noted (e.g., Brown, 2000), and yet Japanese learners seem to have less difficulty acquiring a Russian [r] (Larson Hall, 2004). Why should this be the case? Regardless of the model adopted (SLM; Flege 1995 or PAM; Best 1993), however, certain facts are agreed upon: (1) the first language grammar can act as a kind of phonological filter for L2 sounds; (2) L2 sounds which are quite similar to L1 categories may be difficult to perceive accurately (unless there are robust phonetic cues); (3) L2 sounds which are quite distinct from L1 categories may be accurately perceived more easily.

In this paper, I will present evidence from several language combinations to demonstrate that the L1 phonological filter can be overridden, and new contrasts acquired, when there are robust phonetic cues (in the sense of Wright, 2001) found in the L2 speech stream. While the learners are exposed to much evidence for new contrasts in the speech stream, not all of that input becomes available at once. The more perceptible elements become intake to the processor before the less perceptible elements. Thus, developmental sequences can be explained. In this way, the robustness of the cue determines when input becomes intake.

2. ACQUISITION OF NEW STRUCTURE

Obviously, second language learners must be exposed to input in the target language to acquire the relevant structures. However, it is not the case that learners merely acquire what they are exposed to. We know that learners can acquire some things that are not directly encoded in the speech stream (e.g., traces, moraic consonants, syllable weight, extrametricality, features). And, as plausible as it might seem, even features are not directly read off the input signal. The way in which a language chooses to implement a feature such as [voice] may vary considerably. For example, the feature [+voice] in English is cued very reliably by the lengthening of the preceding vowel, (e.g., *bead/beat*). Final glottal vibration (actual voicing) can be suppressed entirely (Keyser & Stevens, 2006). Chen (1970) showed that vowels before voiced obstruents are about twice as long as those before voiceless consonants. [+voice] can also be cued by high tone on an adjacent vowel in some languages. It is worth noting that these are language-specific effects. Consonant voicing has no significant effect on vowel duration in Polish or Czech (Keating. 1985). Japanese vowels are *shorter* when they are followed by voiced consonants (Campbell. 1992). So, when we are acquiring new features, we need to be aware of what properties cue those features phonetically in the L2. This kind of cross-linguistic variation makes it clear that the acquisition of L2 features is a *learning* problem, not just a matter of *noticing* aspects of the input stream.

2.1 The Deficit Hypothesis

Elsewhere in the field of SLA, we have witnessed a debate between those who argue that certain linguistic properties (e.g. some functional category features) may be unable to be acquired by adult speakers (e.g., Hawkins & Chan, 1997; Hawkins & Hattori, 2006) and those (such as White, 2003) who argue that adult learners *are* able to acquire these features. The first line of thought is what we can call the deficit hypothesis. The deficit hypothesis holds that if element *x* is not found in the first language then it will be unlearnable in adult second language acquisition. So, from a deficit perspective it would be argued if a speaker's L1 lacks a [tense] feature then it will be impossible for that learner to acquire the feature [tense] in an L2. The opposing view would hold that the lack of surface inflection in production does not entail the lack of the appropriate linguistic feature in the grammar. Lardiere (2006) argued that a Chinese L1 subject who was consistently omitting tense markers in her English L2 production also showed evidence of having acquired the abstract feature related to finiteness in her grammar.

Let us return to the field of L2 phonology. In many second language learning scenarios, we may find that someone from a given L1 is attempting to acquire an L2 which has some different phonological properties. Perhaps a feature may be lacking, or the onsets don't branch, or the codas don't project moras, or the feet are iambic rather than trochaic. The empirical question is: will second language learners be able to acquire structures which are not found in their first language? A classic treatment of this question can be found in the work of Brown (2000) who argued that second language learners could not acquire contrasts based on features which were absent from their L1.

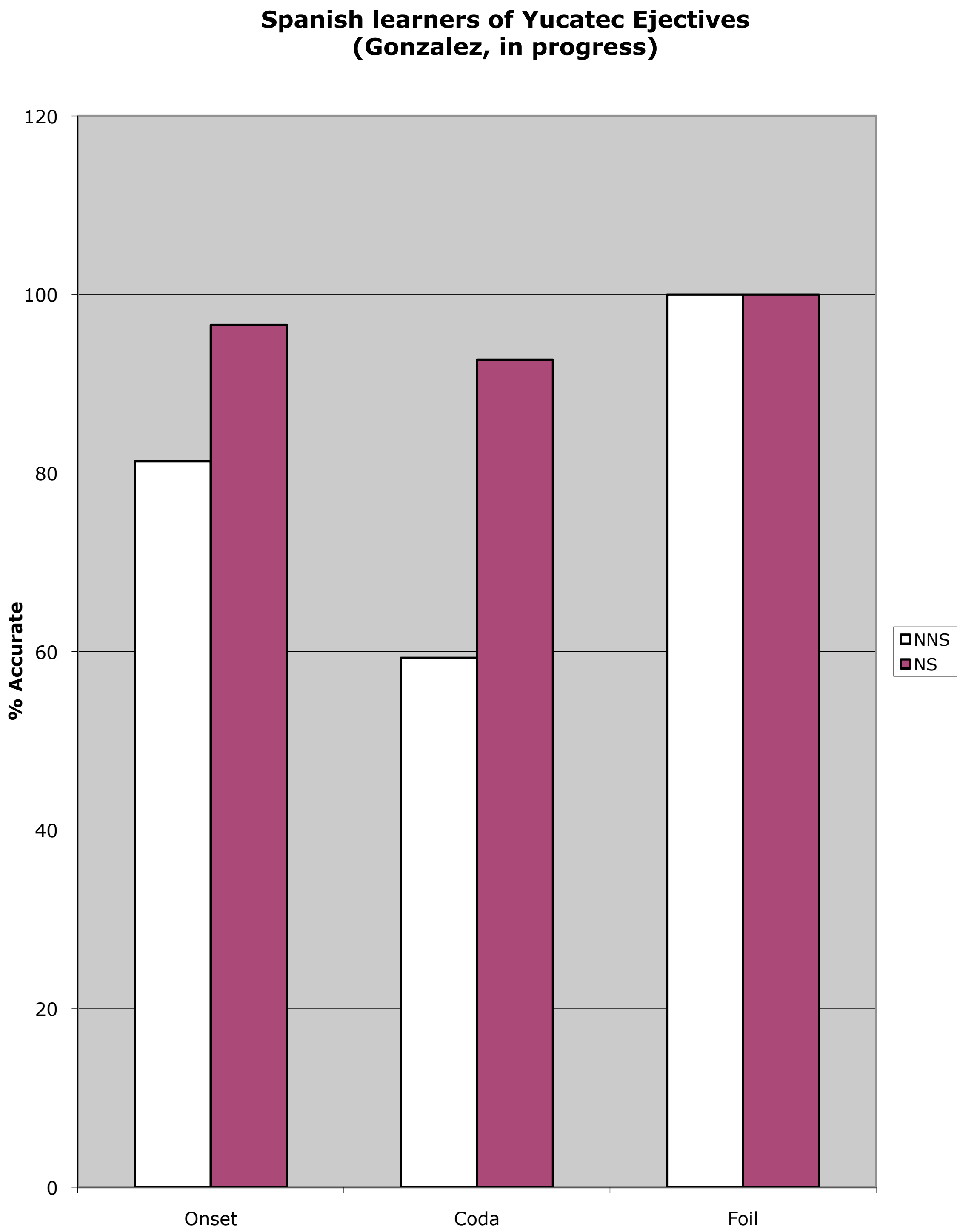
2.2 Robust Cues

There is, however, reason to believe that this deficit model is too strong. There are a number of studies that suggest that circumstances exist where adult second language learners can acquire phonological contrasts even when the relevant feature is inactive in their L1. I will suggest here that one of the conditions under which the L1 filter can be over-ridden is when the input provides robust phonetic cues to what needs to be acquired. Larson-Hall (2004) looks at the perceptual abilities of Japanese speakers learning Russian. Brown argued that the Japanese subjects were unable to acquire the English [l]/[r] contrast in onsets because they lacked the relevant phonological feature in their L1. Larson-Hall's data clearly shows that the Japanese learners of Russian were able to perceive the contrast successfully. Even the beginners were accurate more than 70% of the time (contrasted with the 30% accuracy of Brown's learners of English [ɹ].One possible explanation for this is that the Russian [r] is a trilled sound which makes it very salient in the input to the L2 learners. When the phonetic cues are robust, it is possible to override the effects of the L1 filter (see Wright 2004 for a discussion or robust cues).

3. ACQUISITION OF [CONSTRICTED GLOTTIS]

Gonzalez (in progress) provides evidence of a situation where L2 learners are able to acquire a contrast based on a feature absent from their L1. He looks at the acquisition of Yucatec Maya ejectives by Spanish speakers. Spanish lacks the [constricted glottis] feature required for the phonological structure of ejectives. He conducted both an auditory discrimination task (AX) and a forced-choice picture selection task on 12 non-native speakers and 3 native speaker controls. Thirty items contained plain versus ejectives in singleton onset position (e.g., /ka:n/ 'snake' vs. /k'a:n/ 'land measure'); 24 items contained plain versus ejective voiceless stops and affricates in singleton coda position (e.g., /i:k/ 'hot pepper' vs. /i:k'/ 'wind'); 9 foil pairs of identical stimuli (e.g., /i:k'/ vs. /i:k'/); 57 items consisted of minimal pairs involving contrasts other than ejectives (also involving features present in the first language). The results of the auditory discrimination task are shown in Figure 1.

Figure 1: Spanish Speakers Discrimination of Yucatec Mayan Ejectives



In onset position, the Spanish speakers were not performing significantly differently from the native Yucatec Maya speakers; they were able to acquire the contrast. In the coda position, however, they were not behaving in a nativelike range. One explanation for this goes back to the notion of robust phonetic cues. The transitional cue from the ejective in onset position to the vowel is much more robust that the phonetic cue found when an ejective is at the end of a word. Learners appear to be sensitive to such distinctions. A word-final ejective displays much subtler acoustic cues and it is much more difficult to recover the place and manner of the final consonant. This is also true of final palatal stops in Czech (Atkey, 2001) and final palatalized consonants in Russian (Kulikov, 2007)).

4. ACQUISITION OF [SPREAD GLOTTIS] and [VOICE]

Again, we see that it is the nature of the phonetic properties of the segment which account for its different developmental path even when the L1 feature is absent. Jackson (2009) looked at the discrimination (in an ABX task) of Hindi stops by native speakers of English and French. Her results are given in Table 1 (scores indicate percentages of correct responses).

Table 1. English and French discrimination scores by feature (Jackson, 2009).

	[voice]	[spread glottis]	both features
English	68.9	83.9	85.9
French	79.6	63.5	78.8

French subjects performed significantly better than English subjects on contrasts which differed by [voice] alone. English subjects performed significantly better than French subjects on contrasts which differed by [spread glottis] alone. The subjects' performance on Hindi contrasts which rely on *both* [voice] and [spread glottis] were telling. Regardless of their L1 feature inventories (which she argued to be based on [voice] for French, and [spread glottis] for English), subjects were as good at discriminating the Hindi voiced aspirated stops as they were discriminating contrasts which invoked only their L1 features. Again, I think we can make the argument that the robust phonetic cues evident in the voiced aspirated stop allows the L1 filter to be over-ridden. In spite of lacking the [voice] feature in their L1, the robust phonetic cue which is present allows the English speakers to discriminate [b^h] from [p^h]. The high scores for both English and French subjects on the discrimination of the voiced, aspirated stops clearly shows that they can acquire contrasts which are not found in their L1 something not predicted by deficit models.

5. NON-ROBUST CUES

Mah, Goad and Steinhauer (2007) provides data which show the problematic nature of such cues which are *not* robust to the listener. They looked at native speakers of French acquiring an English /h/ and showed that subjects accurately perceive [h] in non-linguistic tasks but fail to perceive it accurately in auditory discrimination tasks with lexical items (as determined by the Mis-Matched Negativity (MMN) paradigm in Event-Related Potential (ERP) studies). Under this paradigm, subjects are exposed to a majority of a given stimulus (e.g., [p]) and somewhere in the experimental trial an oddball sound (e.g., [b]) is introduced. If a subject detects a difference in the input tokens a particular electrical signal is generated. However, if the distinction is not perceived then this can also be read off the electroencephalogram. Thus, Mah's work shows that it is not that the subjects fail to perceive the [h] at an auditory level but rather that they fail to process it when linguistic representations are invoked.

Furthermore, ERP data show that the NNS subjects fail to invoke an N400 response to lexical items like 'hair' and 'air'. The N400 response in ERP studies is triggered during the processing of a semantic anomaly. If a native English-speaking subject is exposed to sentences such as "The pizza is too hot to eat" and "The pizza is too hot to drink", the latter sentence will trigger a negative electrical pattern 400 milliseconds after the onset of the anomaly (i.e., *drink*). This suggests that the L1 French subjects' lexical representations are impoverished as their inaccurate response is on a lexical task not a discrimination task.

I would argue that these results demonstrate that it is the subtle acoustic properties of the [h]/[∅] contrast which make it difficult for French learners of English to acquire English /h/. It is not that the contrast is impossible to acquire but merely that the input cues in the speech signal are more difficult for the learners to perceive and hence more difficult to process. Following Vanderweide (2005) the more robust input strings will be processed and parsed before the less robust strings. As a result, the more robust strings will be grammatically encoded before the less-robust strings.

6. L2 PHONOLOGY IN A DEAF POPULATION

I would finally like to report on a very different type of study we have been engaged in. Sagae (2007) investigated the role of locality in parsing (using both online (eye-tracking) and offline tasks (questionnaire) in both deaf and hearing populations. She found that relative clause length (illustrated below) did not affect the deaf and hearing populations significantly differently. Clause length had been proposed as a mechanism to explain differences between low-attachment preferences in relative clause attachment and high-attachment preferences. For example, in sentences such as, "Someone shot the servant of the actress on the balcony" there is ambiguity as to whether the servant or the actress was on the balcony. English speakers (Cuetos & Mitchell, 1988) prefer low attachment (the actress) while Spanish speakers (Fernandez, 1988) prefer high attachment (the servant). Fodor (2002) showed that attachment preferences could be biased by controlling for the length of the relative clause which she argued to be a phonological phenomenon. Consider the differences between the sentences:

1. Someone shot the servant of the actress [who was on the balcony]. – short relative clause
2. Someone shot the servant of the actress [who was on the balcony with her husband]. –long clause

She argued that short relative clauses have a low-attachment preference while long relative clauses have high-attachment preference. Her claim is that there is a preference for phonological and syntactic structures to be congruent. While, of course, it is acknowledged that signed languages have phonological structures (e.g., sign lengthening before certain prosodic units), I would argue here that the evidence shows that these deaf learners of English as a second language had acquired processing abilities which crucially relied on phonological structures not found in their L1, as they did not differ from the native speakers in the study.

7. CONCLUSION

All of these data taken together show that second language learners can acquire phonological structures and processing procedures which are not based on their L1 grammars or parsers; the deficit hypothesis cannot be maintained. Even under conditions where the L1 (a signed language) is utlilzing fundamentally different features to encode linguistic structure, we see that the L2 Deaf subjects are able to acquire target-like abilities. We argue here that the phonetic properties of certain robust input cues make certain stimuli available to the L2 processor before others. SLA data concerning such phenomena demonstrate that L2 learners can, indeed, override the L1 phonological filter.

References

Atkey, S. (2001). The Acquisition of L2 Segmental Contrasts: English Speakers' Perception and Production of Czech Palatal Stops. M.A. Thesis. University of Calgary.

Best, C. (1993). A direct realist view of cross-language speech perception. In *Speech Perception and Linguistic Experience: Issues in Cross Language Research*, W. Strange, ed. York Press.

Brown, C. (2000). The interrelation between speech perception and phonological acquisition from infant to adult. In J. Archibald, ed. *Second Language Acquisition and Linguistic Theory*. Blackwell. Pp. 4-63.

Campbell, N. (1992). Speech timing in English and Japanese. *Proceedings for International Symposium on Japanese Prosody,* 207-216. Nara, Japan.

Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica* 22, 129-159.

Cuetos, F. & Mitchell, D. (1988). Cross-linguistic differences in parsing: restrictions on the use of the late closure strategy n Spanish. *Cognition* 30: 73-105.

Fernandez, E.M. (2003). *Bilingual Sentence Processing: Relative Clause Attachment in English and Spanish*. Amsterdam: John Benjamins.

Flege, J. E. (1995). Second language speech learning: theory and findings and problems. In W. Strange, ed. *Speech Perception and Linguistic Experience: Issues in Cross-Linguistic Research*. York Press. pp. 233-277.

Flege, J. E. and R. F. Port. 1981. Cross-language phonetic interference: Arabic to English. *Language and Speech 24:125-146.*

Fodor, J.D. (2002) Prosodic disambiguation in silent reading. In M. Hirotanii (ed.). *Proceedings of the North East Linguistic Society* 32, GSLA, University of Massachusetts, Amherst, 112-132.

Gonzalez, A. (in preparation). *The Acquisition of Yucatec Maya Ejectives by Native Spanish Speakers*. Ph. D. Dissertation. University of Calgary.

Hawkins, R. & Hattori, H. (2006). Interpretation of multiple *wh*-questions by Japanese speakers: a missing uninterpretable account. *Second Language Research* 22, 269-301.

Hawkins, R. &. Chan, Y.-H.C (1997). The partial availability of Universal Grammar in second language acquisition: the 'failed functional features hypothesis.' *Second Language Research* 13, 187-226.

Jackson. S. (2009). *Non-native perception of laryngeal features*. MA thesis, University of Calgary.

Keating, P. (1985). Universal phonetics and the organization of grammars. In V. Fromkin, Ed. *Phonetic linguistics: Essays in honor of Peter Ladefoged*, 115-132. Orlando, FL.: Academic Press.

Keyser, S.J. & Stevens, K. (2006). Enhancement and overlap in the speech chain. *Language* 82(1), 33-63.

Kulikov, V. (2007). Acquisition of [± back] phonological contrast by English learners of Russian. Manuscript, University of Iowa.

Lardiere, D. (2006). *Ultimate Attainment in Second Language Acquisition: A Case Study*. Mahwah, NJ: Lawrence Erlbaum Associates.

Larson-Hall, J. (2004). Predicting perceptual success with segments: a test of Japanese speakers of Russian. *Second Language Research* 20(1): 32-76.

Mah, Jennifer, Heather Goad, and Karsten Steinhauer. 2007. Francophones and English /h/: an acoustic problem? Evidence from event-related brain potentials. Boston University Conference on Language Development 32. Boston, MA.

Sagae, S. (2007). Sentence Processing and Processing. M.A. Thesis. University of Calgary

Wright, R. (2001). Perceptual cues in contrast maintenance. In E. Hume and K. Johnson, eds. *The Role of Speech Perception in Phonology*. Academic Press. pp. 252-277.

Transfer of Greek palatals in L2 English

E. Babatsouli and I. Kappa

Division of Linguistics, University of Crete, Greece
ebabatsouli@yahoo.com,kappa@phl.uoc.gr

ABSTRACT

While extensive literature has appeared on L2 English segmental phonology with respect to many L1 tongues, there is little work on L1 Greek-L2 English interlanguage phonology (e.g. Coutsougera, 2007). Our pilot study reports that the palatals of L1 Greek language are erroneously transferred into L2 English; native Greek speakers of L2 English mispronounce velar stops (/k/, /g/) as palatal ones ([c], [ɟ]) and the voiceless glottal fricative (/h/) as palatal fricative ([ç]) in the environment where high front vowels [i, ɪ, e] follow. On a related note, the English cluster /kj/ is realized as either [c] or [k] when [u, ʊ] follow. The results are supported by a pilot study in which native Greek informants of various proficiency levels in English read a wordlist including the sounds in question. Our study showed that a distinct majority of the informants adhered to the phonological patterns described above with phonological awareness proving imperative for near-native articulations. The phonological output is biased by articulatory habits and lack of perceptual distinctiveness leading to L1 sounds being transferred to L2 interlanguage. We find that *gradience*, characteristic of the palatalization process is a phenomenon that explains inter-language variation and L2 interlanguage inconsistency in terms of both production and perception.

Keywords: L2 transfer, palatalization, production, perception, gradience

1. INTRODUCTION

In this study, we document for the first time the transfer of L1 Greek palatals into L2 English. Experts and laymen alike agree that foreign accent is the quintessence of second language speech and transfer is rudimentary to its understanding. A ‘global foreign accent’ (Major, 2001: 18) results from inconsistency in the production of individual segments, their combinations and of prosodic features whilst lack of mastery of all three phonological levels precludes near-native production (ibid). Mispronunciations of L2 segments are classified as phonemic, phonetic, allophonic and distributional (e.g. Moulton, 1962). Transfer has been documented in numerous phonological studies as on segments (e.g. Archibald and Young-Scholten, 2003) and loan phonology (e.g. Paradis, 2006). Substantively, it is explained in terms of both production and perception. The surface and underlying forms of L1 sounds affect the L2 speaker’s perception of L2 sounds. In a top down manner, L2 learners unknowingly replace new, similar or identical foreign sounds with L1 ones as they are ‘forcing square pegs into round holes’ (Flege, 1991:151). Perceived similarity of L1 and L2 sounds determines the level of L2 sound assimilation (Best, 1995). The phonetic decoding that adjusts representations into language specific ones is facilitated by acoustic proximity (Kuhl, 2000) and proximity of subtle articulator gestures (Best and Strange, 1992). Such substantive factors account for phonological bias (Wilson, 2006) in favour of the L1 phonological system. In short, the transfer of Greek palatal sounds into L2 English is a direct result of both co-articulatory processes of habit and economy, as well as, perceptual assimilation.

The stratified nature of phonological systems (Jakobson, 1968) cannot be more evident than in the palatalization process which affects the primary articulation of consonants and necessitates secondary articulations in the context of high fronted vowels or of the palatal approximant. Palatalization is a phonological process occurring in languages with variable distribution and phonetic implementation; it is phonemic in some (e.g. Baltic and Slavic languages) and allophonic in others (e.g. Greek and English). Articulatory and acoustic variation, however slight, creates distinct phonemic and phonetic categories that spread across languages and within individual languages. With respect to Czech, Hungarian, English and Russian, the phonetic outcomes of velar palatalization are distinct and can be both phonemic and contextual in varying gradience levels (Keating & Lahiri, 1993). A review of four attested patterns of velar palatalization (i.e. k→k^j, k →c, k→tɕ, k →t) and their languages is given by Lee (2000:416). A great deal of variation and style-shifting is attested in interlanguage phonology by many researchers (Tarone, 1979, among others) but it seems that such variation also exhibits gradience. Phonological awareness resulting from perceptual differentiation of the underlying and surface forms leads the way to increasing articulatory practice as the L2 phonological system is built up through intermittent stages of variable, gradient interlanguage articulations.

2. THE PILOT STUDY

2.1. Informants and Experiment

We have conducted a pilot study that included one hundred adult native Greek informants of various degrees of proficiency in English, the minimum being the University of Cambridge LESIE Certificate of Proficiency and the maximum a PhD degree from a USA or UK university and/or residence experience (RE). The group was diverse in including both L2-educated and L2-experienced speakers as e.g. a 5% of informants with some RE and formal education in an L2 native environment in childhood and a 10% sample of immigrants with very long RE but minimal formal L2 education. It is of significance that the majority of informants (90%) have not or no longer use the L2 in daily interactions with native speakers of the L2. The same six written words were read aloud by each informant separately for each of the seven sounds under question for a total of forty two words. The words were chosen with the relevant sound appearing in word-initial and word-medial positions. The word list used is given in Table 1 below with L1 English transcriptions and an example output of interlanguage transfer from our data.

Table 1: Word list of relevant sounds used in the study

/ki/ → [ci] (e.g.)	**/ke/ → [ce]** (e.g.)	**/gi/ → [ɟi]** (e.g.)	**/ge/ → [ɟe]** (e.g.)
skill /skɪl/ → [scil] *baking* /'beɪkiŋ/ → ['beɪciŋ] *napkin* /'næpkɪn/→['napcɪn] *keep* /ki:p/ → [cip] *folkish* /ˈfəʊkɪʃ/ → [folkiʃ] *kitty* /ˈkɪtɪ/ → [citi]	*encase* /ɪnˈkeis/ → [ɛnˈceis] *occasion* /ə'keiʒ(ə)n/ → [oˈceiʒon] *cake* /'keɪk/ → [ˈceik] *kept* /kept/ → [cept] *hacker* /hækə(r)/ → [ˈxɑcer] *decay* /dɪ'keɪ/ → [diˈcei]	*begin* /bɪˈgɪn/ → [biˈɟin] *give* /gɪv/ → [ɟiv] *leggings* /ˈlegɪ:ŋs/ → [ˈleɟiŋs] *giggle* /ˈgɪg(ə)l/ → [ˈɟigəl] *giddy* /ˈgɪdɪ / → [ˈɟidi] *hagiography* /ˌhægɪˈɒgrəphɪ/→ [ˌɑɟɪˈogrɑphɪ]	*girl* /gɜ:l/ → [ɟe^{r}l] *game* /'geɪm/ → [ˈɟeim] *again* /əˈgen/ → [eˈɟen] *forget* /fə:'get/ → [forˈɟet] *beget* /bɪˈget/ → [biˈɟet] *guess* /ges/ → [ɟes]
/he/ → [çe] (e.g.)	**/hi/ → [çi]** (e.g.)	**/kju/ → [cu]** (e.g.)	
hen /hen/ → [çen] *ahead* /əˈhed/ → [ɛˈçed] *herd* /hɜ:d/ → [çerd] *helmet* /ˈhelmɪt/ → [ˈçelmet] *upheld* /ʌpˈheld/ → [ʌpˈçeld] *behave* /bɪˈheɪv/ → [biˈçeiv]	*Uphill* /ʌpˈhɪl/ → [ʌpˈçil] *hippie* /ˈhɪpɪ/ → [ˈçipi] *here* /ˈhɪə(r)/ → [ˈçiɑ^{r}] *reheat* /ri:ˈhi:t/ → [riˈçit] *prohibit* /prəˈhɪbɪt/ → [proˈçibit] *hinder* /ˈhɪndə(r)/ → [ˈçinde^{r}]	*particular* /pəˈrtɪkjʊlə(r)/ → [pɑrˈticulə^{r}] *articulate* /ɑ:ˈtɪkjʊlət/ → [ɑrˈticulet] *thank you* /ˈθæŋkju:/ → [ˈθɛŋcu] *excuse* /ɪkˈskju:z/ → [ɛkˈscuz] *cucumber* /ˈkju:kʌmbə(r) / → [ˈcukɑmbe^{r}] *cute* /ˈkju:t/ → [cut]	

2.2. The Results

The results of the pilot study are depicted in Figures 1 and 2 below. Figure 1 shows the frequency with which informants transfer L1 Greek palatal sounds into L2 English. Out of the seven relevant sounds, the data document: circa 80% transfer for four of the sounds, i.e. /ke/→[ce] (80%), /ki/→[ci] (78%), /gi/→[ɟi] (81%), /kju/→[cu] (81%); the highest percentage is documented at 84% for /ge/ → [ɟe] transfer while the lowest percentages are documented at 52% transfer of /hi/→[çi] and 69% of /he/→[çe]. For some of the sounds, the study additionally showed further variable interlanguage outputs of a characteristically gradient nature, gradating between L1 Greek sounds, the attested palatal sounds in transfer and near-native (nN) outputs. Such gradience is evidenced for the /ge/, /gi/, /he/, /hi/ and /kju/ sounds but not for /ke/ and /ki/. The output of these *gradient* variables is illustrated in Figure 2 showing the frequency of informants' gradient realizations. For each phoneme, the frequency shown includes more than one gradient sound, as discussed in detail in section 3.3 below; only the *dominant* gradient variable is written in the figure.

Figure 1: Frequency of informant transfer of L1 Greek palatal sounds into L2 English

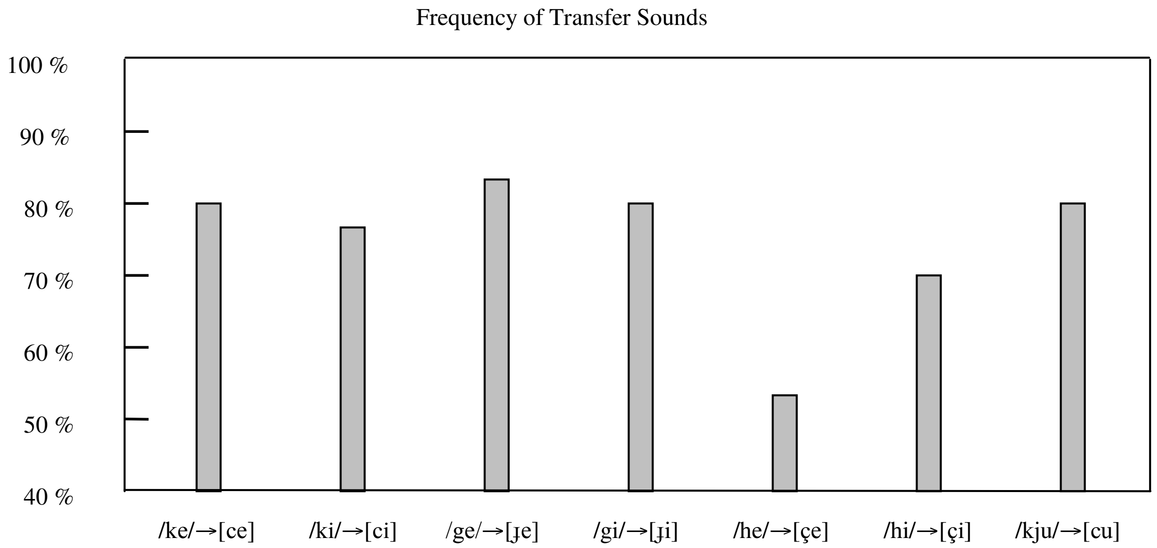
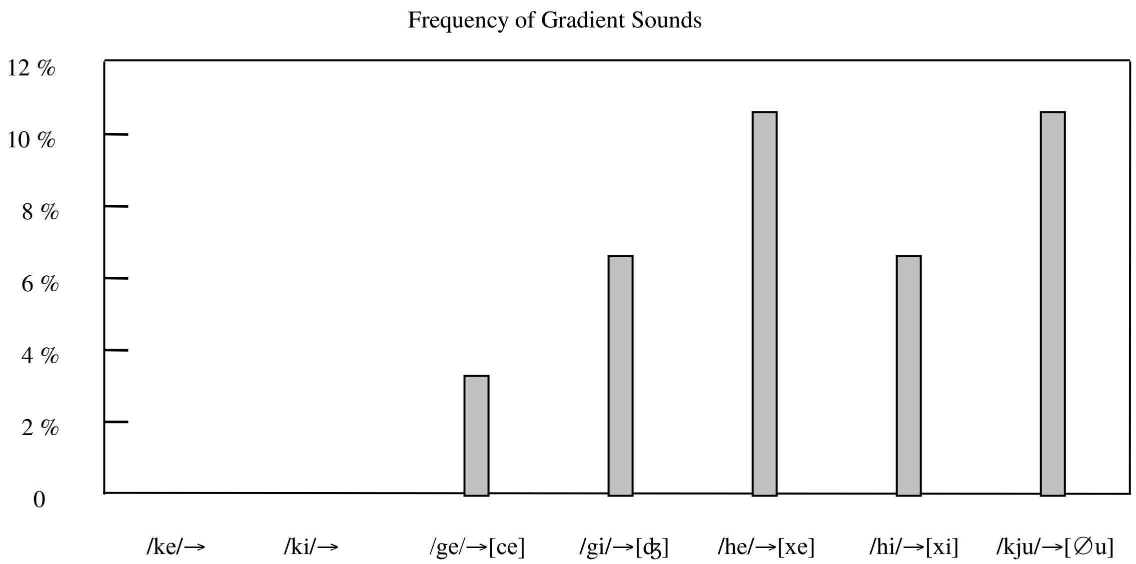


Figure 2: Frequency of informants' gradient interlanguage sounds



Specifically, in Figure 2 there are 10% of gradient productions for /he/ and /kju/, 6% for /gi/ and /hi/ and 3% of gradient articulations for /ge/. The informants' frequency of nN productions is evidently obtained by subtracting both the transfer and the gradient variable articulations from the total number of productions, resulting in: 15% nN outputs for /ke/, 22% for /ki/, 13% for /ge/, 13% for /gi/, 38% for /he/, 25 % for /hi/ and 9% for /kju/. The remaining 5% of /ke/ interlanguage realizations are affected by erroneous vowel reading and is, thus, not included in either transfer or gradient frequencies.

3. DISCUSSION AND DATA ANALYSIS

3.1. Relevant Palatalization in Greek and English

Since transfer of palatalization is the main focus of our study, an overview of the relevant processes in L1 Greek and L1 English will facilitate an understanding of the underlying forms involved in the respective languages, of the transfer phenomenon as well as of the reported inter-language and L2 interlanguage gradience. Palatalization is a cover term in the literature and the 'palatalization rule' for Greek and English arbitrarily refers to two different palatalization processes. Specifically:

*The Greek Palatalization Rule***:** In Greek, the phonemic velar stops /k, g/ and the voiceless velar fricative /x/ are assigned each a pair of allophones in complementary distribution, i.e. /k/→[k], [c], /g/→[g], [ɟ], and /x/→[x], [ç]. Both sets of allophones are systematically realized in the Greek language in specific obligatory environments. The palatalization rule applies in the presence of high front vowels /i, e/. Then the underlyingly back consonants /k, g, x/ are fronted in the oral cavity and realized as palatal allophones [c, ɟ, ç] respectively, e.g. *and* [ce] , *I look* [ci'tao], *bridle* /'ɟemi/, *bad luck* /'ɟiɲa/, *hand* ['çeri], *I pour* ['çino]. Lee (2000:416) misrepresents Greek velar palatalization by assigning it to the first of his four attested velar palatalization outputs: /k/→[k^j] rather than the second /k/→[c]; the example provided is [k^jino], an abbreviation of *that one* [e'cino]. The Greek velar palatalization outcomes are the IPA palatal sounds. They are articulated at variable points of constriction and with 'different types of gestural organization for the same underlying units' (Nikolaidis, 2001:73). In line with Keating & Lahiri, (1993), the Greek palatals [c, ɟ, ç] are distinct in articulatory and perceptual terms from the gradient fronted or palatalized velars articulated closer to the soft palate border rather than the main hard palate region. Similarly, velar palatalization occurs when /i/ is followed by a stressed /u/ or any other vowel, e.g. *earthen cask* ['cupi] *κιούπι*, *snow* ['çoni] *χιόνι*, *humour* ['çumor] *χιούμορ*, *type of bird* ['ɟonis] *γκιόνης*, etc. Although /i/ is assimilated i.e. /k + i + u/ → [cu], it is orthographically transparent.

The English Palatalization Rule: Velar palatalization in English is under-represented in the literature. The palatalization rule refers to the affrication of alveolars to palatoalveolars when a palatal glide /j/ follows; affrication of /g/→[ʤ]/– [i, e] is another allophonic palatalization process. Outcomes in either case are not distinct palatals. According to Keating & Lahiri (1993:76), the English voiceless velar /k/ is contextually fronted or palatalized but not palatal although the IPA palatal symbol is occasionally used in transcriptions. The /k, g/ closures can be near-palatal when /i:/ has a very front and tense articulation (e.g. Gimson, 1989). The process is non-obligatory, allophonic in free variance necessitated by coarticulation rather than an underlying rule. Native English speakers are not markedly aware of the distinction between velars and their allophonic free variants; mental representations of the sounds are predominantly velar. Consequently, it is to no surprise that Lee (op. cit.) does not include the English language in his overview of velar palatalization. On a related note, the palatal glide [j] is articulated after /k/→[kj]/-[u:, ʊə], e.g. *document* ['dɒkjʊmənt] and a palatal fricative [ç] after /k/ in the same context in a stressed syllable e.g. *cute* ['kçut]. A near-palatal [c] is sometimes articulated coalescing the constriction of [kç]. The function of /j/ here is obscure producing complex consonants. This last process in English is orthographically opaque.

3.2. Data Transfer Sounds

We note that all palatal transfer instances in the study are irrespective of level of education and length of RE. Even among the small groups of nN articulations produced by people that satisfied primarily phonological awareness and, secondarily, high educational level and/or long RE, palatal transfer instances sprung up in non-coincidentally the same words, that is, easy words in vocabulary terms. For example, in the case of **/ke/, /ki/→[ce], [ci]**, instances include *kept*→[cept], *keep*→[ci:p], *cake*→['ceɪk] and *skill*→[scɪl]. The *keep*→[ci:p] outputs may be justified as resulting from the L1 English allophonic process (→[k^ji:p]) but the remaining interlanguage articulations are clearly L1 Greek transfer persisting. Arguably, these words' ease in vocabulary terms relaxed the L2 speakers' guard in a way that words like *encase* → e.g. [ən'keis] and *occasion* → e.g. [ɒ'keiʒən], [o'keiʒon] wouldn't. This is also the case for **/ge/, /gi/→[ɟe], [ɟi]** where lapses of transfer are also evidenced among the group of nN speakers in familiar words e.g. *guess* [ɟess], *give* [ɟiv] but not in *beget* and *giddy*. nN articulations *are* documented among informants with no RE and/or advanced education. We further note that the informants' phonological awareness is centered on individual words in the sense that nN articulations of the sounds discussed are produced in the same, specific words but not in the remaining instances. For example, /k/ in *folkish* was near-natively articulated by 40% of the informants when these same informants transferred the [c] palatal in all other instances. A similar practice is observed for the words *girl* and *hacker*. This indicates that it is *individual words* in the L2 lexicon that are more likely to stand out as phonetically distinct rather than *individual sounds*. In the case of **/he/ /hi/ → [çe] [çi]**, we note the smallest percentages of palatalization transfer in the study. Of all individual sounds discussed, the glottal fricative /h/ does not exist in the L1 Greek phonological system. Interlanguage instances are articulated with the transfer of the L1 Greek palatal [ç]. These results further support the 'Speech Learning Model': new sounds are more readily perceived and acquired in the L2 than similar ones (Flege, op. cit.). Lastly, **/kju/→[cu]** also exhibits a high transfer average. Exceptional in the study is 100% of nN articulations of /kju/→[kju] by 9% of all the informants that satisfy the factors: some native formal education in childhood, graduate/post-graduate education in the L2, RE and continuing interaction with native speakers of the L2.

3.3. Data Gradient Sounds

This group of interlanguage productions unambiguously shows a gradience of articulations that includes palatal transfers, other L1 Greek sound transfers, L1 English palatalization sounds, nN productions and deletion of relevant sounds. Specifically, in the case of **/ge/**, **/gi/** gradient variables follow two phonological processes: **1.** devoicing of /g/→[c]/ - [e, i], e.g. *forget* [ˌfor'cet], *gift* [cift], *beget* ['becet], *giggle* [cigɫ]. This is evidenced in all informant levels including nN speakers suggesting a universal tendency for articulatory economy and ease. Interestingly, devoicing does not deter palatal transfer with /g/→[ɟ]→[c], not /g/→[k]; **2.** affrication of /g/→[ʤ] mostly before [i] but also before [e] and a single instance of affrication of /g/→[ʧ]/- [i], e.g. *giggle* [ʤigɫ], *beget* [be'ʤet]; *giggle* [ˈʧigɫ]. This may be explained as deriving from knowledge of the English allophonic rule in complementary distribution: /g/→[ʤ] and/or the L2 speaker's unfamiliarity with the word; **3.** for **/gi/**, gradient outputs include sound deletion, e.g. *hagiography* [haiˈografi] and an instance of /gi/→ [xi], i.e. *hagiography* [xaxioˈgrafi]. For the **/he/**, **/hi/** sounds, only one gradient output is documented: L1 Greek [x] transfer, e.g. *upheld* [ʌpˈxeld] and *hippie* ['xipi]. The [x] realization indicates informant awareness of /he/ as a new sound but due to articulatory inexperience, articulation is backed only as far as the acquired L1 velar fricative. The **/kju/** sound also shows gradience: **1.** /kju/ → [∅] deletion e.g. *excuse* [eksˈuz] denoting the articulatory difficulty of the two-cluster sequence /ks + kj/; **2.** affrication of /kju/→[kʃ], [ʧ] e.g. *excuse* [ekʃˈuz], *cute* [tʃut]; and **3.** /kju/→[ka, ku] showing phonological and/or orthographic word unfamiliarity which consequently leads to L1 orthographically instigated realizations e.g. *cucumber* [ku'kamber]. This individual word exhibited most variable outputs by all informant levels punning once more on the importance of individual words in the L2 lexicon rather than sounds. **4.** /kju/→[k.c], [g.j] in *thank you* [ˈθenk.cu], [ˈθeng.ju]. *Thank you* exhibited a high average of nN realizations being the only example in the list with a relevant sound at word boundaries. Finally, for **/ke/, /ki/**, no noticeable intermediate gradient articulations are documented which translates into either that L2 speakers cannot perceive and articulate the phonetic differences of such intermediate gradient sounds in line with the native English speakers' practice, or that such gradient variables might show up in a palatographic study.

4. CONCLUSIVE REMARKS

Paramount in our findings is that a combination of both general phonological awareness and acoustic familiarity with the word facilitates nN articulations irrespective of native formal education past childhood and length of RE. The position of the sounds in the word seems non-detrimental to the L2 realization in this study. Exceptions arise when a sound appears at word-boundaries forcing speculations that purposeful word-boundary inclusion of the relevant sounds may have yielded different results. Phonemic and phonetic gradience underlines palatalization and differentiates the process in Greek and English. Intra- and inter-language palatalization gradience is increasing linearly from the ends of the oral cavity towards its middle as depicted in Figure 3 and interlanguage phonology seems equally gradient with variable L1/L2 phonetic instances, depicted in the figure in bold.

Figure 3: Intra-Language, Inter-Language & Interlanguage Gradience in Greek and English

Front alveolars /t, d, s, z/ → **palatoalveolars [ʧ, ʤ, ʃ, ʒ]** → **palatals [c, ɟ, ç]** ← ([∅] ←)

palatalized velars [k^j, g^j] ← fronted velars [k^j, g^j] ← **velars /k, g, x/** ← **glottal /h/** ***Back***

This study supported the notion that L2 speakers are overwhelmingly *inarticulate* and *deaf* concerning L2 sounds (e.g. Piske and Young-Scholetn, 2009). L1 articulatory habit and perceptual bias impedes phonological differentiation of L1/L2 sounds despite education level or long RE. Phonological awareness, unless acquired naturalistically in early childhood, appears to be hardly an automatic but an arduous, intentional process building up *gradually* and is the chief facilitator of nN articulatory success. L2 acquisition of phonology gradates from high levels of L1 perceptual and articulatory constraints (the core of transfer) to increasingly lower levels to near-native realizations. This is schematically illustrated in Figure 4:

Figure 4: Gradience of Transfer in SLA phonology

High Transfer → **Lower Transfer** → **Lower Transfer** → **Low Transfer** → **Near Native**

High perceptual/articulatory constraints → Decreasing perceptual/articulatory constraints → Near Native

5. REFERENCES

Archibald, J., Young-Scholten, M. 2003. The second language segment revisited. *Second Language Research*, 20(3). 163-167.

Best, C.T. 1995. A direct realist view of cross-language speech perception. In: Strange, W. (ed.) *Speech perception and linguistic experience: Issues in cross-language research*). Timonium, MD: York Press. 171–204.

Best, C., Strange, W. 1992. Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics* 20(3). 305–330.

Coutsougera, P. 2007. The impact of orthography on acquisition of L2 phonology: inferring the wrong phonology from print. In: Ashby, M., Maidment, J. (eds.) *Proceedings of the Phonetics Teaching and Learning Conference*. University College London.

Edwards, J. G. H., Zampini, M. L. (eds.). 2008. *Phonology & Second Language Acquisition*. John Benjamins Publishers.

Flege, J. E. 1991. Perception and production: the relevance of phonetic input to L2 phonological learning. In: Huemner, T. Ferguson, C.A. (eds.) *Crosscurrents in Second Language Acquisition and Linguisitc Theories.* John Benjamins Pub. Co. 249-269.

Jakobson, Roman. 1968. Child Language Aphasia and Phonological Universals. Mouton Publ. The Hague.

Keating, P., Lahiri, A. 1993. 'Fronted velars, palatalized velars, and palatals'. *Phonetica*. 50 (2). 73-101.

Kuhl, P. K. 2000. A new view of language acquisition. *Proceedings of the National Academy of Science*. 97. 11850–11857.

Major, R. C. 2001. *Foreign Accent: the Ontogeny and Phylogeny of Second Language Phonology*. Lawrence Erlbaum Associates.

Moulton, W. 1962. Toward a classification of pronunciation errors. *Modern Language Journal*, 46. 101-109.

Nicolaidis, K. 2001. An electropalatographic study of Greek spontaneous speech. *Journal of the International Phonetic Association*. 31(1). 67-85.

Piske, T., Young-Scholten, M. (eds.) *Input Matters in SLA.* Second Language Acquisition Series. Bristol: Multilingual Matters.

Paradis, C. 2006. The unnatural /C^ju/ (< foreign /Cy/) sequence in Russian loanwords: A problem for the perceptual view. *Lingua* 116. 976–995.

Tarone, E. E. (1979). Interlanguage as chameleon. *Language Learning* 29: 181–191.

Wilson, C. 2006. Learning phonology with substantive bias: an experimental and computational study of velar palatalization'. *Cognitive Science*. 30 (5). 945-982.

Testing PAM and SLM:
Perception of American English approximants by native German listeners[1]

Ocke-Schwen Bohn and Catherine T. Best***

[*]English Department and Center on Autobiographical Memory Research, Department of Psychology, Arhus University, Denmark
engosb@hum.au.dk
[**]MARCS Auditory Laboratories, University of Western Sydney, Australia
Haskins Laboratories, New Haven, CT, USA
c.best@uws.edu.au

ABSTRACT

This study examined the impact of phonetic and phonological properties of L1 German (GE) on the perception of the American English (AE) approximant contrasts /r/-/l/, /w/-/r/, and /w/-/j/. GE does not have /w/, it realizes /r/ and /l/ differently from AE, and GE and AE /j/ are realized nearly identically. Thus, German lacks /w/-/j/ and /w/-/r/, but employs /r/-/l/ with a "light" [l] (as opposed to AE "dark" [ɫ]) and a uvular fricative [ʁ] (as opposed to AE "retroflex" [ɻ]). Forced-choice identification and AXB discrimination of /Cɑk/ syllables revealed both phonological and phonetic influences on the perception of AE approximants. GE listeners' identification of all contrasts was highly categorical, but discrimination was poorer than AE listeners' for /w/-/r/ and /r/-/l/ and better than AE listeners' for /w/-/j/. Phonologically-based predictions were correct only for one contrast, /r/-/l/. Neither Best's Perceptual Assimilation Model nor Flege's Speech Learning Model were fully successful in predicting how L1 GE listeners perceived AE approximants.

Keywords: Cross-language, PAM, SLM, categorical perception, L1 phonetics and phonology.

1. INTRODUCTION

Previous research has shown both phonological and phonetic influences from the L1 on perception of nonnative speech contrasts. For example, speakers of languages which do not contrast /r/ and /l/ (Korean, Japanese) have difficulty categorizing and discriminating AE /r/-/l/ (e.g., MacKain et al. 1981). However, perception is also influenced by nonnative phonetic realizations of phonological contrasts that do occur in the L1, as when L1 French listeners differ from L1 AE listeners in labeling stimuli from an AE /r/-/l/ continuum even though /r/ and /l/ are phonemic in French (Hallé et al. 1999).

The present study examined phonetic and phonological influences on the perception of AE approximants by L1 GE listeners. GE has /r/, /l/, and /j/, like AE, but it lacks /w/. Of the three "shared" approximants, only /j/ is realized nearly identically in AE and GE, although GE /j/ can occur as a voiced or devoiced palatal fricative syllable-initially. The typical phonetic realization of GE /r/ as an unrounded voiced uvular fricative or approximant [ʁ] is quite dissimilar from AE /r/, which is realized as a "bunched" central dorsal approximant [ɹ] or retroflex [ɻ] with additional labial and pharyngeal constrictions. GE /l/ is non-velarized [l], making it phonetically similar though not identical to AE /l/, which is typically velarized [ɫ].

On purely phonological grounds, GE listeners should perceive the AE /r/-/l/ contrast much like AE listeners because both languages have the /r/-/l/ contrast. Because GE lacks /w/, GE listeners may have some difficulty identifying the /w/ end of both the AE /w/-/r/ and /w/-/j/ continua unless they perceive AE /w/ phonologically as GE /v/, in which case their perception should be much like that of AE listeners.

These predictions differ somewhat from those that consider the phonetic realizations of AE approximants and their counterparts in GE. Best's Perceptual Assimilation Model (PAM, Best 1995; Best and Tyler 2007) predicts that GE listeners should categorize and discriminate AE /r/-/l/ categorically (PAM's "Two-Category assimilation"). However, they should differ within each category from AE listeners because GE /r/ is phonetically different from AE /r/, and because GE /l/ is similar but not identical to AE /l/. GE listeners' categorization boundary locations and slopes, and/or their within- and between-category discrimination levels should reflect phonetic differences in goodness of fit to their native /r/ and /l/ realizations.

GE listeners should also identify and discriminate the AE /w/-/r/ contrast categorically because it is either a Two Category (TC) assimilation type or an Uncategorized vs. Categorized contrast (UC) for L1 Germans. However, the perception of GE listeners should not be as accurate as the AE listeners' because GE /r/ is realized differently from AE /r/, and because AE /w/ is either assimilated to GE /v/, which differs from the /w/ endpoint, or heard as an uncategorized consonant (i.e., heard as speech, but in-between GE phonetic categories).

For the AE /w/-/j/ contrast, PAM predicts that GE listeners should perform very well because AE /w/-/j/ is also either a TC or UC contrast for L1 Germans with one endpoint, /j/, which is nearly identical in GE and AE. However, the difference in the nativeness of the /w/ endpoint leads to the prediction that the boundary location and the slope of the identification function will differ between GE and AE listeners.

The predictions generated by Flege's Speech Learning Model (SLM, Flege 1995) for inexperienced learners are much the same as those generated by PAM, in that they are also based on perceived phonetic similarities.[2] However, SLM differs from PAM in two important ways: 1) SLM focuses on individual phonetic categories whereas PAM focuses on pairwise phonological contrasts; 2) SLM's primary focus is on L2 production, which perception is posited to guide and constrain, whereas PAM's is directly on cross-language perception. Still, we can extrapolate basic predictions from the core principles of SLM, which classifies the relationship of sounds of the nonnative language to native sounds along a continuum ranging from "new" over "similar" to "identical".

The SLM prediction for the perception of /r/-/l/ is that the GE listeners will perceive AE [ɹ̩] as quite different from their native [ʁ] and thus classify it (eventually) as a new phone, and that they will classify the similar AE [ɫ] as being equivalent to their L1 [l] in spite of their phonetic differences, and will therefore perceive the /r/-l/ continuum categorically. However, the GE listeners will not perform as well as AE listeners especially on the [ɹ̩] endpoint because of the large phonetic distance between AE [ɹ̩] and GE [ʁ]. SLM predicts the same perceptual pattern for /w/-/r/ because AE /w/ is either treated as equivalent to GE /v/ or perceived as a new sound. Perception of /w/-/r/ by GE listeners should be categorical, but within-category perception should not be as accurate as for AE listeners. The SLM prediction for /w/-/j/ is that GE listeners' performance will be best for this contrast, which they perceive categorically, with a well-defined /j/-endpoint (because GE and AE are nearly identical) and a less well defined endpoint for /w/, which is phonetically dissimilar but treated as equivalent to GE /v/ or as a new sound.

In conclusion, purely phonological considerations lead to the expectation that L1 GE listeners will perceive AE /r/-/l/ like L1 AE listeners, and they may perceive AE /w/-/r/ and /w/-/j/ less accurately than AE listeners, depending on how they perceive /w/. The predictions of PAM and SLM, which are both concerned with phonetic realizations, are that GE listeners will not be as accurate as AE listeners for any of the approximant contrasts. PAM and SLM predict that GE listeners' perception will be best for /w/-/j/, though not equal to AE, and less accurate for /w/-/r/ and /r/-/l/.

The experiments reported below test these predictions. We compare the results of L1 GE listeners' identification and discrimination of stimuli from AE /r/-/l/, /w/-/r/, and /w/-/j/ continua to the results obtained by Best and Strange (1992) for L1 AE listeners for the same stimuli.

2. METHODS

2.1. Participants

Eighteen native North German speakers participated as unpaid volunteers (10 females, 8 males, mean age = 21.3 years, SD = 1.9). They were students at Kiel University, and met the following selection criteria: no history of hearing loss, L1 GE speaker, and limited exposure to languages other than GE (i.e., less than a total of eight months in a foreign language environment). All participants realized GE /r/ as [ʁ] (i.e., none of the participants realized GE /r/ as the apical trill [r] or came from an area where /r/ is commonly realized as [r]).

2.2. Stimulus Materials

We used the three 10-step synthetic continua of AE approximant contrasts - /rɑk/-/lɑk/, /wɑk/-/jɑk/, and /wɑk/-/rɑk/ - from Best and Strange (1992). Briefly, the 3-formant stimuli were generated with an OVE-IIIc cascade formant synthesizer, and the final /k/ burst was excised from natural speech and appended. The resulting syllables were equated for overall duration (330 ms including the final /k/ burst), amplitude, and F0 contour (rising-falling).

2.3. Procedure

Subjects were tested in three groups of six in one session each in the language laboratory of Kiel University. The signal from the three audiotapes was routed through Tandberg Educational Media Centre IS-10MM to Tandberg Educational headphones. The sequence in which the three contrasts were presented was counterbalanced across groups. For each contrast, a two-choice identification test (ISI = 3.0s) was followed by an AXB discrimination test (ISI = 1.0s, ITI = 3.0s). The subjects were provided with two response sheets for each contrast. The first sheet contained 200 lines for the identification test, and the subjects put a circle round the orthographic representation of the first sound in the syllable (W or Y, W or R, R or L) heard on each trial. The second sheet contained 140 lines for the discrimination test, and the subjects wrote “1” or “3” according to whether the second syllable was identical to the first or the third syllable in the triad of syllables presented on each trial.

3. RESULTS

One-way ANOVAs (with Language Group as between-subjects factor) were conducted on the following dependent variables: For the identification test, we obtained boundary values (estimates of the stimulus number in the continua at which the percept switches from one category to the other, indicating the subject’s 50% crossover) and slope values (estimates of the steepness of the identification function, indicating how categorical subjects were in dividing the continua) from probit analyses. We also conducted separate two-way ANOVAs (Language Group x Stimulus Number) on the full categorization functions to assess GE and AE listeners’ differences in responses to individual stimuli. In the analyses of discrimination performance, three dependent variables were selected for one-way ANOVAs: Overall per cent correct discrimination scores, per cent correct discrimination for the stimulus pair that straddled the phoneme boundary (as established through probit analyses in the identification test), and a measure of the flatness (or “peakiness”) of the discrimination function (i.e., the mean of the unsigned difference scores for adjacent stimulus pairs. We also conducted two-way ANOVAs (Language Group x Stimulus Pair) on the full discrimination functions to assess differences in the groups’ responses to individual stimulus pairs.

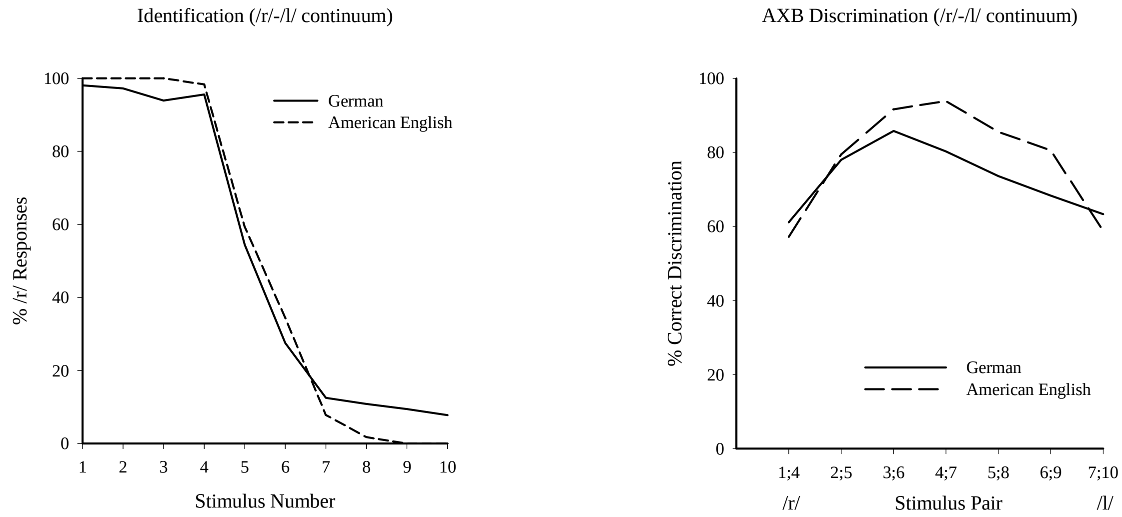
3.1. /r/-/l/

Figure 1 compares the identification and the discrimination function for the /r/-/l/ continuum as perceived by the L1 GE and the L1 AE listeners. One-way ANOVAs revealed no significant differences between the boundary values for the GE (5.4) and the AE (5.4) listeners. Likewise, the slope values for the GE (1.5) and the AE (2.2) listeners did not differ significantly. The two-way ANOVA on the full categorization function found only a significant Stimulus Number effect, $F(9,225) = 203.25$, $p < .0001$, but no Language difference or interaction, despite apparent discrepancies in labeling near both ends of the continuum, especially near the /l/ end (Figure 1, left).

ANOVAs comparing the discrimination performance of the GE and AE listeners revealed that the mean per cent correct discrimination scores for the GE listeners (73.0%) and the AE listeners (77.8%) did not differ significantly, and that the “peakiness” of the discrimination functions (i.e., the mean of difference scores for adjacent stimulus pairs) for the GE listeners (10.9%) and the AE listeners (11.9%) did not differ significantly. Also, the mean difference in cross-category discrimination accuracy (GE: 85.1% correct, AE: 92.5% correct) was nonsignificant. However, significant effects were found by the two-way ANOVA for Stimulus Pair, $F(6,150) = 26.66$, $p < .0001$, and the Language x Stimulus Pair interaction, $F(6,150) = 3.21$, p

= .005. Simple effects tests on the interaction indicate that discrimination was significantly lower for GE than AE listeners on the /l/ side of the category boundary, specifically for stimulus pair 4-7 (GE: 80% correct; AE: 94%), $F(1,25) = 6.08$, $p < .02$, pair 5-8 (GE: 74%; AE: 86%), $F(1,25) = 4.69$, $p < .04$, and pair 6-9 (GE: 68%; AE: 81%), $F(1,25) = 4.91$, $p < .03$.

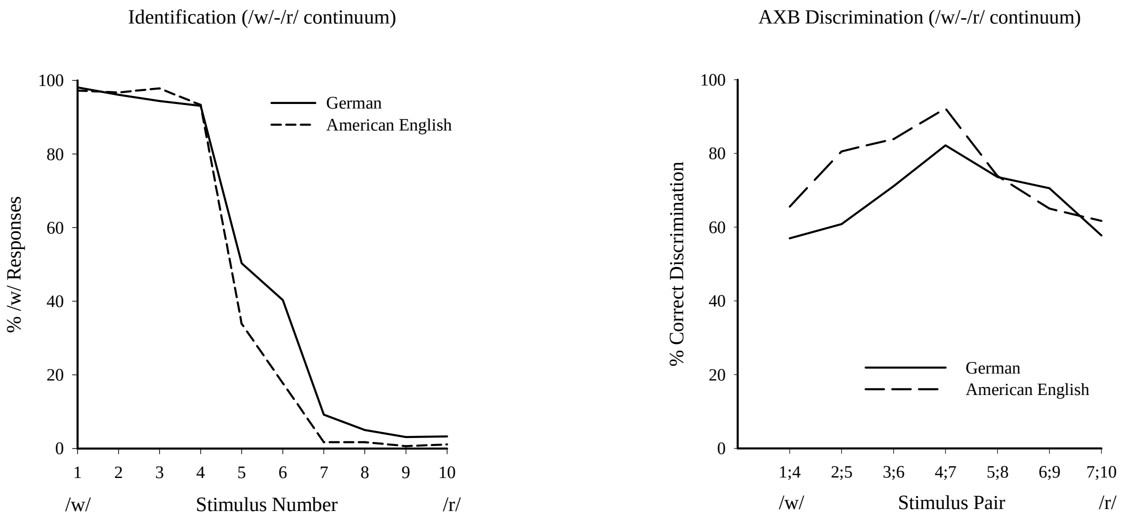
Figure 1: Identification functions (left panel) and discrimination functions (right panel) for the /r/-/l/ continuum as perceived by L1 GE and L1 AE listeners.



3.2. /w/-/r/

Figure 2 compares the identification and discrimination functions for perception of the /w/-/r/ continuum by L1 GE and L1 AE listeners. One-way ANOVAs did not reveal significant differences between the boundary values for the GE (5.4) and the AE (4.8) listeners, or between the slope values for the GE (1.5) and the AE (1.7) listeners. However, the two-way ANOVA revealed not only a significant Stimulus effect, $F(9,225) = 313.23$, $p < .0001$, but also a main effect of Language (GE mean: 48% “W” responses; AE mean: 44% “W”), $F(1,25) = 4.53$, $p < .05$, and a significant interaction, $F(9,225) = 2.69$, $p = .005$. Simple effects tests of the interaction found that the GE listeners gave significantly more “W” responses than the AE listeners for stimulus items 5 (GE: 50% correct; AE: 34%), $F(1,25) = 9.52$, $p < .002$, and 6 (GE: 40% correct; AE: 18%), $F(1,25) = 17.94$, $p < .0001$, that is, just near the /r/ side of the category boundary.

Figure 2: Identification functions (left panel) and discrimination functions (right panel) for the /w/-/r/ continuum as perceived by L1 GE and L1 AE listeners.

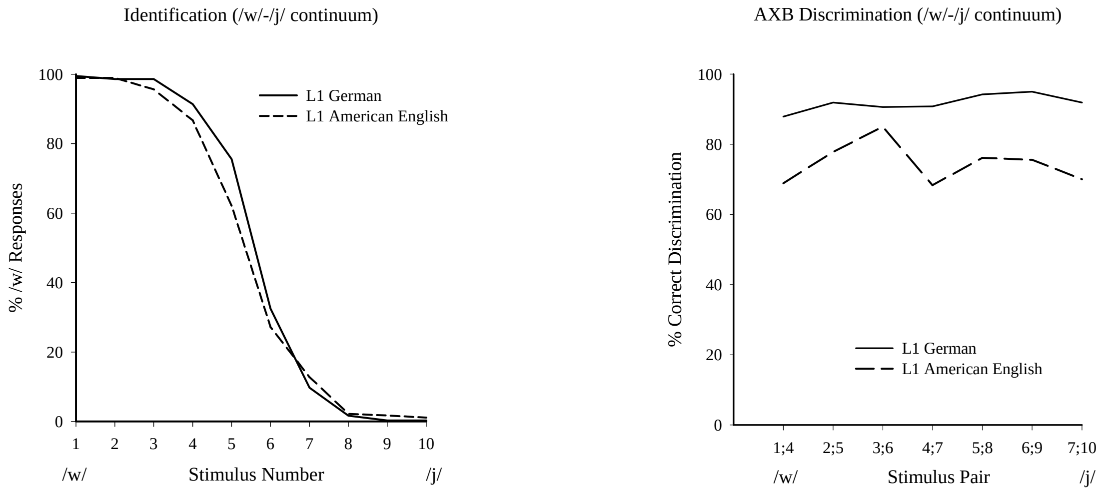


The discrimination performance of the GE and AE listeners differed significantly with respect to the mean per cent correct discrimination scores (GE: 67.6%; AE: 74.7%), F(1,25) = 5.722, p < .05. However, the “peakiness” scores of the discrimination functions for the GE listeners (11.5%) and the AE listeners (12.0%) did not differ, nor did the accuracy of the cross-category discrimination (GE: 81.5% correct, AE: 86.1% correct). This indicates that GE and AE listeners were equally categorical in their discrimination of the /w/-/r/ contrast, but that the AE listeners’ overall discrimination was more accurate than the GE listeners’. The two-way ANOVA provided further insight into this difference. The main effects of both Language, $F(1,25) = 5.72$, $p < .03$, and Stimulus Pair, $F(6,150) = 14.44$, $p < .0001$, were significant, as was their interaction, $F(6,150) = 2.77$, $p < .02$. Simple effects tests found significantly lower discrimination by GE than by AE listeners on the /w/ side of the boundary, for stimulus pairs 2-5 (GE: 61% correct; AE: 81%), $F(1,25) = 12.79$, $p < .0001$, and 3-6 (GE: 71% correct; AE: 84%), $F(1,25) = 5.40$, $p < .03$; the difference was marginally significant for stimulus pair 4-7 (GE: 82% correct; AE: 92%), $F(1,25) = 3.29$, $p = .07$.

3.3. /w/-/j/

Figure 3 compares the identification and the discrimination function for the perception of the /w/-/j/ continuum by the GE and AE listeners. One-way ANOVAs revealed no significant differences between the boundary values for the GE (5.6) and the AE (5.4) listeners, and for the slope values for the GE (1.8) and the AE (1.9) listeners. Likewise, there was no evidence of a listener group difference in the two-way ANOVA, where only the Stimulus effect was significant, $F(9,225) = 203.79$, $p < .0001$.

Figure 3: Identification functions (left panel) and discrimination functions (right panel) for the /w/-/j/ continuum as perceived by L1 GE and L1 AE listeners.



However, the GE listeners were significantly more accurate than the AE listeners at discriminating stimuli from the /w/-/j/ continuum (GE: 90.3% correct, AE: 74.5% correct, F(1,25) = 26.131, p < .001), and they were also significantly more accurate at discriminating stimuli straddling the category boundary, (GE: 92.5% correct, AE: 78.3% correct, F(1,25) = 11.36, p < .01). However, the “peakiness” scores of the discrimination functions (GE: 7.3%, AE: 9.4%) did not differ significantly. The two-way ANOVA revealed significant main effects of Stimulus Pair, $F(6,150) = 5.65$, $p < .0001$, and Language, $F(1,25) = 26.13$, $p < .0001$. Simple effects tests of the significant interaction, $F(6,150) = 2.37$, $p < .05$, showed that the GE listeners outperformed the native AE listeners on all but one stimulus pair throughout the continuum, *F*s (1,25) ranged from 6.69 to 23.72, $p < .02$ to .0001, the exception being pair 3-6, which is near the higher AE peak, and on which discrimination accuracy did not differ significantly for the listener groups (see Best and Strange 1992, for an account of their two peaks). Overall, the GE listeners show an unusual pattern of categorical identification and continuous discrimination of stimuli from the /w/-/j/ continuum.

4. DISCUSSION

The results of the experiment are not fully consistent with any of the predictions presented in the Introduction; however, they are more in line with the fine-grained phonetic predictions of PAM and/or SLM than with purely phonological predictions. For the perception of the /r/-/l/ contrast, the phonologically based prediction was supported, which states that GE listeners will not differ from AE listeners because both GE and AE have /r/-/l/ contrasts. The differences in the phonetic implementation of this contrast in AE and in GE, which motivated PAM's and SLM's prediction of less accurate perception of /r/-/l/ by the GE than the AE listeners, did not seem to affect the GE listeners' perception with respect to the between-category measures examined here. Concerning within-category perception, the GE listeners were not affected by the large phonetic difference between AE [ɹ] and GE [ʁ], whereas the difference between AE [ɫ] and GE [l] reduced discriminability on the /l/ side of the category boundary.

PAM and SLM correctly predicted that the GE listeners would differ from the AE listeners in their perception of the /w/-/r/ continuum. Their points of divergence from AE listeners are compatible with phonetically based within-category differences as addressed by these models. Comparison of the boundary and slope values for the identification functions and of the "peakiness" scores and of the accuracy of discrimination at the category boundary indicated that GE and AE listeners discriminated stimuli from the /w/-/r/ continuum equally categorically. The differences between GE and AE listeners for /w/-r/ suggest that the identification of this nonnative contrast is influenced by mismatches in phonetic detail between GE and AE, whereas discrimination of /w/-/r/ is blind to phonetic detail and guided only by phonological considerations, with AE /w/ and /r/ TC-assimilated to the GE counterparts /v/ and /r/.

None of the predictions correctly anticipated the GE listeners' identification of stimuli from the /w/-/j/ continuum, which was indistinguishable from the AE listeners', nor did the predictions anticipate that the discrimination accuracy of the nonnative listeners would be superior to them. This result poses the same problem as the results for L1 Danish (Best and Bohn 2002) and L1 French listeners (Hallé et al. 1999) on the same continua: All three nonnative groups did better than expected. By contrast, L1 Japanese listeners performed similarly to AE listeners (Best and Strange 1992). The unsolved puzzle is what phonetic principles underlie the American English and Japanese versus German, Danish, and French group patterns on /w/-/j/. A possible explanation could be that the superior and vowel-like discrimination of this contrast by German, French, and Danish listeners is related to the characteristics of the vowel systems of these languages, which, unlike English and Japanese, contrast front rounded and unrounded vowels. Native experience with contrastive lip rounding for vowels, combined with the vowel-like properties of word-initial glides, may enhance sensitivity to the /w/-/j/ contrast, in which lip rounding is also employed contrastively.

5. REFERENCES

Best, C. T. 1995. A direct realist view of cross-language speech perception. In: Strange, W., (ed.), *Speech perception and linguistic experience: Issues in cross-language research.* Timonium, MD: York Press. 171-204.

Best, C. T., Bohn, O.-S. 2002. Perception of American English glide consonants by Danish listeners. *Journal of the Acoustical Society of America* 112. 2388.

Best, C. T., Strange, W. 1992. Effects of phonological and phonetic features on cross-language perception of approximants. *Journal of Phonetics* 20. 305-330.

Best, C. T., Tyler, M. D. 2007. Nonnative and second-language speech perception. In: Bohn, O.-S., Munro, M. J. (eds.), *Language experience in second language speech learning: In honor of James Emil Flege.* Amsterdam: J. Benjamins. 13-34.

Flege, J. E. 1995. Second language speech learning: Theory, findings, and problems. In: Strange, W., (ed.) *Speech perception and linguistic experience: Issues in cross-language research.* Timonium, MD: York Press. 233-277.

Hallé, P., Best, C. T., Levitt, A. 1999. Phonetic vs. phonological influences on French listeners' perception of American English approximants. *Journal of Phonetics* 27. 281-306.

MacKain, K. S., Best, C. T., Strange, W. 1981. Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics* 2. 369-390.

NOTES

[1] Supported by NIH grant DC00403 to C. T. Best. We thank Anja Steinlen and Thorsten Piske for recruiting subjects in Germany, and Meike Bohn for entering and reducing the German data.

[2] We will consider only those predictions of SLM which concern nonnative speakers with little English language experience - "inexperienced learners" - because the focus of PAM and of the present study is on this learner group.

Dutch listeners' perception of Korean fortis, lenis, and aspirated stops: First exposure

Mirjam Broersma

Max Planck Institute for Psycholinguistics, Nijmegen, & Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, The Netherlands
mirjam@mirjambroersma.nl

ABSTRACT

This study addresses Dutch listeners' perception of Korean fortis, lenis, and aspirated stop triplets. The Korean stop triplets were expected to be extremely difficult for Dutch listeners, as all Korean stops are voiceless (in initial position), while Dutch only distinguishes voiced and voiceless stops. Dutch listeners were not familiar with Korean and received no training; they only heard six examples of each target sound before performing a phonetic categorization task.

The Korean stops were lenis, fortis, and aspirated bilabial (/p/-/p*/-/p^h/), denti-alveolar (/t/-/t*/-/t^h/), and velar (/k/-/k*/-/k^h/) stops. Stimuli were blocked by type and place of articulation; thus one block concerned either a fortis-lenis, a lenis-aspirated, or a fortis-aspirated contrast, and either bilabial, denti-alveolar, or velar stops.

Both Korean and Dutch listeners performed best on fortis-aspirated, intermediate on lenis-fortis, and poorest on aspirated-lenis contrasts. Not surprisingly, Dutch listeners performed less accurately than Korean listeners. Remarkably, however, the Dutch listeners performed significantly above chance level (if only just above it numerically) for each contrast type at each place of articulation. Thus, even at first exposure and without any training, Dutch listeners managed to identify the extremely difficult L2 sounds relatively successfully.

Keywords: Speech perception; Consonants; Identification; Non-native language; Korean.

1. INTRODUCTION

The identification of phonemes in a second language is one of the greatest challenges for non-native listeners (see e.g. the collected papers in Strange, 1995, and Bohn & Munro, 2007). The largest perceptual difficulties have been proposed to arise in cases in which the second language (L2) has two phonemes where the native language (L1) has only one, as described by the Perceptual Assimilation Model (Best, 1994; Best & Tyler, 2007). In the present study, another extremely problematic case is investigated: a case in which the L2 has *three* phonemes where the L1 has only one phoneme in the same perceptual space. This study investigates Dutch listeners' perception of Korean stop triplets. Dutch listeners were not familiar with Korean, and received no training; they only heard six examples of each target sound before performing phonetic categorization of these truly New Sounds.

Korean has a three-way stop contrast, and all of those stops are (at least in initial position) voiceless. The three stop categories are lenis, fortis, and aspirated, all occurring at bilabial (/p/-/p*/-/p^h/), denti-alveolar (/t/-/t*/-/t^h/) and velar (/k/-/k*/-/k^h/) places of articulation. Dutch, on the other hand, distinguishes prevoiced and voiceless unaspirated stops (/b/-/p/, /d/-/t/, /k/; /g/ only occurs in loanwords). Dutch listeners are likely to perceive all Korean lenis, fortis, and aspirated stops as most similar to Dutch voiceless stops. The Korean three-way distinction can therefore be expected to be very difficult to distinguish for Dutch listeners.

Some of the perceptual cues that are important for the Korean three-way distinction, i.e., the voice quality of the following vowel (with creaky voice after fortis stops, breathy voice after lenis, and breathy or modal voice after aspirated stops; Cho, Jun, & Ladefoged, 2002; Kang & Guion, 2006), do not play a role in Dutch stop voicing contrasts. Others, like VOT and F0 at onset of the following vowel (Cho et al., 2002; Kang &

Guion, 2006; Lisker & Abramson 1964) are also important cues for the stop voicing contrast in Dutch (Van Alphen & Smits, 2004). Those cues, however, have different critical values in the two languages.

This study addresses Dutch listeners' perception of Korean lenis, fortis, and aspirated stops in nonword-initial position at first exposure, and compares their performance to that of native Korean listeners. The study assesses whether Dutch listeners can identify Korean stops at a level above chance, even at first exposure and without receiving any training, and whether Dutch and Korean listeners differ in which stop categories they find most confusable.

2. METHOD

2.1. Participants

Participants were 36 native listeners of Korean and 36 native listeners of Dutch. Korean participants were students at Hanyang University (Seoul, Korea), and Dutch participants at the Radboud University Nijmegen (The Netherlands), who participated for course credits or a small financial compensation. None reported any hearing loss. None of the Dutch participants had any knowledge of or experience with Korean.

2.2. Materials

Target sounds were the nine stops /p/, /p*/, /p^h/, /t/, /t*/, /t^h/, /k/, /k*/, /k^h/. Each target sound occurred in initial position, in three phonetic contexts, followed by the vowel /i/, /u/, or /ɛ/. (Note that /ɛ/ rather than /a/ was used because the vowel /a/ was the target sound in a vowel contrast tested in the same perception experiments as the stops, the results of which are not discussed in this paper.) There were thus 27 items (9 target stops * 3 vowel contexts). Each item was recorded 20 times, yielding a total of 540 stimuli.

The materials were recorded by a 23 year old female native speaker of Korean, who had been born and raised in Seoul. She read the items, presented in Korean orthography, one by one, separated by a pause, in a clear citation style.

The recording was made in a sound proof booth with a Sennheiser microphone and stored directly onto a computer at a sample rate of 41.5 kHz. Stimuli were excised from the recording using the speech editor Praat.

Additionally, Korean vowel, fricative, and affricate materials were used, the results of which are not reported here. Those materials consisted of recordings by the same speaker of 5 Korean vowels, 3 fricatives, and 2 affricates, with 60 stimuli per phoneme.

2.3. Design

Stimuli were blocked by type and place of articulation; thus one block concerned either a fortis-lenis, or a lenis-aspirated, or a fortis-aspirated contrast, and either bilabial, or denti-alveolar, or velar stops. Participants were equally distributed over three stimulus lists. Each list contained one fortis-lenis, one lenis-aspirated, and one fortis-aspirated stop contrast, each for one place of articulation. Thus, no participant heard one type of stop contrast in more than on block, and no participant heard one place of articulation in more than block.

Additionally, participants categorized three other contrasts that are not described here, concerning Korean vowels, fricatives, and affricates.

The order of blocks was counterbalanced, and stimuli within a block were presented in a semi-random order which was different for each participant, with each phoneme occurring maximally four times in succession.

2.4. Procedure

Participants were tested one at a time in a quiet room, seated in front of a computer. They received written instructions in their native language that they would hear a series of items containing one of two sounds, numbered 1 and 2. They were instructed to decide on each trial which of the two sounds they had heard, and to indicate their response by pressing the corresponding response button, labeled 1 or 2, as fast and as accurately as possible.

Before each of the six blocks they received further instructions about the two response alternatives and the corresponding response buttons in that block. First, it was indicated on the computer screen if the block concerned vowels or consonants. Next, 6 unique examples were played of each target sound, accompanied by the number 1 or 2 on the screen. Example stimuli were similar to the experimental stimuli, and were grouped by phonetic context (i.e., examples for the /p/ - /p*/ contrast were pa, pa, p*a, p*a, pi, pi, p*i, p*i, pe, pe, p*e, p*e, in that order).

The experiment started with a short practice part in which participants categorized the Korean /i/-/u/ contrast (which was easy to distinguish for Dutch listeners), to familiarize them with the task.

The experiment was controlled with NESU (Nijmegen Experiment Set-Up) software. Stimuli were played binaurally, one at a time, over Sennheiser closed headphones at a comfortable level. Participants responded by pressing one of two buttons, labeled “1” and “2”, on a box in front of them. There was no time limit for the responses. At 600 ms after button press, the next stimulus was played.

3. RESULTS AND DISCUSSION

3.1. Korean listeners

Table 1. Korean listeners’ results; % correct, mean *d’*, mean *log β*, one-tailed One-Sample T Test for *d’* > 0, two-tailed One-Sample T Test for *log β* ≠ 0. (Higher values of *d’* indicate higher sensitivity. Negative values of *log β* indicate a bias towards the first, and positive values towards the second phoneme in the first column.)

Contrast	% Correct	*d’*	*log β*	T Test *d’*	T Test *log β*
Bilabials					
Lenis – Fortis	90.3	3.96	-1.86	t (11) = 5.8 p < .001	t (11) = -1.8 p = .09
Fortis – Aspirated	94.8	4.06	-0.74	t (11) = 9.8 p < .001	t (11) < |1| p > .1
Aspirated – Lenis	79.4	1.79	-0.44	t (11) = 10.2 p < .001	t (11) = -2.9 p < .05
Denti-alveolars					
Lenis – Fortis	89.0	3.16	-2.56	t (11) = 9.7 p < .001	t (11) = -2.4 p < .05
Fortis – Aspirated	94.0	3.89	0.74	t (11) = 8.7 p < .001	t (11) < |1| p > .1
Aspirated – Lenis	78.0	1.65	0.19	t (11) = 9.2 p < .001	t (11) = 1.5 p > .1
Velars					
Lenis – Fortis	85.7	2.68	-1.55	t (11) = 11.5 p < .001	t (11) = -1.8 p > .1
Fortis – Aspirated	91.3	3.85	-0.64	t (11) = 7.3 p < .001	t (11) < |1| p > .1
Aspirated – Lenis	85.2	2.83	1.66	t (11) = 5.8 p < .001	t (11) = 1.5 p > .1

Korean listeners' categorization results are summarized in Table 1. Mean percentages correct per contrast range from 78.0 to 94.8 %.

For each participant and each contrast separately, *d'* was calculated to assess listeners' sensitivity (with a correction for near-perfect sensitivity, MacMillan & Creelman, 1991). A *d'* of 0 indicates that listeners do not treat two phonemes as different; a *d'* of 1 corresponds to 69 % correct, and the effective upper limit of *d'* is 4.65. One-tailed One-Sample T Tests for each contrast showed that *d'* was always significantly larger than 0 (Table 1). Listeners were thus sensitive to each contrast.

As Table 1 shows, percentage correct (and, with one exception, *d'*) was largest for fortis-aspirated, intermediate for lenis-fortis, and smallest for aspirated-lenis contrasts. An ANOVA with *d'* as the dependent variable and Contrast Type (lenis-fortis, fortis-aspirated, aspirated-lenis), Place of Articulation, and Context (following vowel) as independent variables showed a main effect of Contrast Type (F (2, 99) = 21.4, $p <$ .001). To further investigate this effect, similar ANOVAs were done comparing the Contrast Types pairwise. They confirmed that *d'* was significantly larger for fortis-aspirated contrasts than for both lenis-fortis (F (1, 66) = 9.6, $p < .01$) and for aspirated-lenis contrasts (F (1, 66) = 34.3, $p < .001$), and that *d'* was larger for lenis-fortis than for aspirated-lenis contrasts (F (1, 66) = 20.5, $p < .001$). This pattern was not modified by place of articulation, as there were no interactions between Contrast Type and Place of Articulation.

Further, *log β* was calculated to assess possible response biases (McNicol, 1972). A *log β* of 0 indicates that there is no bias, a negative *log β* that there is a bias towards the first phoneme, and a positive *log β* that there is a bias towards the second phoneme mentioned in the first column of the table. Two-tailed One-Sample T Tests for each contrast showed that *log β* was significantly different from 0 in two cases; for the aspirated-lenis bilabials, there was a bias towards aspirated ('p^h') responses, and for the lenis-fortis denti-alveolars, there was a bias towards lenis ('t') responses.

3.2. Dutch listeners

Dutch listeners' categorization results are summarized in Table 2. Mean percentages correct per contrast range from 55.3 to 78.9 %.

As for the Korean listeners, for each participant and each contrast separately, *d'* was calculated as a measure of sensitivity, and *log β* as a measure of bias. One-tailed One-Sample T Tests for each contrast showed that *d'* was always significantly larger than 0 (Table 2). Despite the sometimes very low percentages correct, listeners thus showed sensitivity to each contrast.

Table 2 shows that, similar to the Korean listeners' results, percentage correct and *d'* were largest for fortis-aspirated, intermediate for lenis-fortis, and smallest for aspirated-lenis contrasts. Like for the Korean listeners, an ANOVA with *d'* as the dependent variable and Contrast Type, Place of Articulation, and Context as independent variables showed a main effect of Contrast Type (F (2, 98) = 19.1, $p < .001$). To further investigate this effect, similar ANOVAs were done comparing the Contrast Types pairwise. The analyses confirmed again that *d'* was significantly larger for fortis-aspirated contrasts than for both lenis-fortis (F (1, 66) = 4.3, $p < .05$) and for aspirated-lenis contrasts (F (1, 65) = 38.5, $p < .001$), and larger for lenis-fortis contrasts than for aspirated-lenis contrasts (F (1, 65) = 16.7, $p < .001$). Like for the Korean listeners' results, this pattern was not modified by place of articulation, as there were no interactions between Contrast Type and Place of Articulation.

Comparing the Dutch and Korean listeners' percentages correct and *d'* values (Tables 1 and 2), it is clear that the Dutch listeners had a lower accuracy than the Korean listeners for all contrasts. Indeed, in an ANOVA with *d'* as the dependent variable and Language Group (Dutch and Korean), Contrast Type, Place of Articulation, and Context as independent variables, there was a main effect of Language Group (F (1, 197) = 149.5, $p < .001$). Because there were also significant interactions among Language Group, Context, and Contrast Type (F (4, 394) = 2.7, $p < .05$), and among Language Group, Context, and Place of Articulation (F (4, 394) = 3.1, $p < .05$), the effect of Language Group was also calculated for each contrast separately. For each contrast, *d'* was significantly larger for the Korean listeners than for the Dutch listeners (Table 2).

Finally, *log β* was calculated as a measure of bias for each participant and each contrast separately. Two-tailed One-Sample T Tests for each contrast showed that *log β* was significantly different from 0 in one case; for the lenis-fortis bilabials, there was a bias towards lenis ('p') responses. This is different from the Korean listeners' results, and indeed, in an ANOVA with *log β* as the dependent variable, there was a significant interaction among Language Group, Contrast Type, and Place of Articulation (F (4, 197) = 2.5, $p < .05$).

Table 2. Dutch listeners' results; % correct, mean *d'*, mean *log β*, one-tailed One-Sample T Test for *d'* > 0, F test for *d'*: main effect of Language Group, two-tailed One-Sample T Test for *log β* ≠ 0. (Higher values of *d'* indicate higher sensitivity. Negative values of *log β* indicate a bias towards the first, and positive values towards the second phoneme in the first column.)

Contrast	% Correct	*d'*	*log β*	T Test *d'*	F test *d'*, main effect of Language Group	T Test *log β*
Bilabials						
Lenis – Fortis	67.9	1.12	-0.54	t (11) = 5.6 $p < .001$	F (1, 22) = 19.4 $p < .001$	t (11) = -2.9 $p < .05$
Fortis – Aspirated	78.9	2.24	0.55	t (11) = 4.3 $p < .001$	F (1, 22) = 8.5 $p < .01$	t (11) < |1| $p > .1$
Aspirated – Lenis	55.3	0.28	-0.03	t (11) = 2.2 $p < .05$	F (1, 21) = 45.8 $p < .001$	t (11) < |1| $p > .1$
Denti-alveolars						
Lenis – Fortis	66.1	0.96	0.14	t (11) = 3.9 $p < .01$	F (1, 22) = 35.0 $p < .001$	t (11) < |1| $p > .1$
Fortis – Aspirated	76.0	2.15	0.16	t (11) = 3.6 $p < .01$	F (1, 22) = 9.7 $p < .01$	t (11) < |1| $p > .1$
Aspirated – Lenis	56.1	0.40	-0.25	t (11) = 2.2 $p < .05$	F (1, 22) = 22.4 $p < .001$	t (11) = -1.4 $p > .1$
Velars						
Lenis – Fortis	63.0	0.77	-0.24	t (11) = 4.3 $p < .001$	F (1, 22) = 31.2 $p < .001$	t (11) = -1.8 $p > .1$
Fortis – Aspirated	76.7	1.89	-0.50	t (11) = 4.4 $p < .001$	F (1, 22) = 8.9 $p < .01$	t (11) < |1| $p > .1$
Aspirated – Lenis	55.6	0.29	-0.01	t (11) = 2.6 $p < .05$	F (1, 22) = 25.3 $p < .001$	t (11) < |1| $p > .1$

4. GENERAL DISCUSSION

The results showed that, as expected, Dutch listeners found it very difficult to distinguish the Korean three-way stop contrasts. For each of the contrast types at each place of articulation, the Dutch listeners were significantly less accurate than the Korean listeners. Despite percentages of correct responses that were sometimes as low as 55-56 %, the Dutch listeners nevertheless performed significantly above chance level for all contrast types at all places of articulation; their sensitivity as measured by *d'* was always significantly above 0.

Dutch and Korean listeners showed some differences in response biases. Importantly, however, they did not differ in which combinations of stop types they found most difficult. Both Dutch and Korean listeners performed best on the identification of fortis versus aspirated stops, intermediate on lenis versus fortis stops, and worst on aspirated versus lenis stops, at all places of articulation. This might be because the perceptual cues that have been shown to be most important for Korean listeners' recognition of the three-way stop contrast, i.e., VOT and F0 of the following vowel (Cho et al., 2002; Kang & Guion, 2006), also play a role for Dutch stop voicing perception (Van Alphen & Smits, 2004). Thus, even though Dutch listeners were not familiar with the Korean stops, they might have attempted to use the same perceptual cues as the Korean listeners did to some extent. Saliency of those cues might have resulted in a similar pattern from difficult to less difficult contrasts for Dutch and Korean listeners alike.

Thus, even at first exposure and without any training, Dutch listeners managed to identify the Korean fortis, lenis, and aspirated stops at a level above chance. As Korean stop triplets are arguably an example of the worst possible L2 contrasts for Dutch listeners, with three L2 phonemes in the same perceptual space where the L1 has only one phoneme, it is remarkable that the Dutch listeners managed to identify the sounds successfully, even if that success only meant that they recognized them with an accuracy that was just above chance level.

5. REFERENCES

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman & H. C. Nusbaum (Eds.), *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words* (pp. 167-224). Cambridge, MA: MIT.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege* (pp. 13-34). Amsterdam: John Benjamins.

Bohn, O.-S., & Munro, M. J. (Eds.). (2007). *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*. Amsterdam: John Benjamins.

Cho, T., Jun, S.-A., & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics, 30*, 193-228.

Kang, K.-H., & Guion, S. G. (2006). Phonological systems in bilinguals: Age of learning effects on the stop consonant systems of Korean-English bilinguals. *Journal of the Acoustical Society of America, 119*, 1672-1683.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20*, 384-422.

MacMillan, N. A., & Creelman, C. D. (1991). *Detection Theory: A User's Guide*. Cambridge: CUP.

McNicol, D. (1972). *A Primer of Signal Detection Theory*. Sydney: Australasian Publishing Company.

Strange, W. (Ed.). (1995). *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Baltimore: York Press.

Van Alphen, P. M., & Smits, R. (2004). Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: The role of prevoicing. *Journal of Phonetics, 32*, 455-491.

Native speakers' reactions to Polish-accented English - empirical evidence.

Agnieszka Bryła - Cruz

Maria Curie-Skłodowska University of Lublin
agnieszka.bryla@poczta.umcs.lublin.pl

ABSTRACT

The paper examines and compares the major phonetic features which contribute most to perceived accentedness of Polish English and its degree of annoyance to native speakers of two language varieties: English English and Irish English. This is done on the basis of an empirical study which analyses participants' responses to two samples of Polish-accented speech. Experimental results allow for establishing a hierarchy of phonetic aspects responsible for the listeners' judgments, comparing the perception of Polish English pronunciation by different groups of native speakers and drawing conclusions with regard to phonetic priorities in the teaching of English to Poles, and emigrants to the British Isles in particular.

1. INTRODUCTION

The massive exodus of Polish citizens to the British Isles in recent years has resulted in a situation in which the inhabitants of Ireland, England and Scotland have become important interlocutors to a great many Poles who have settled permanently or temporarily in those regions. The present paper reports on the results of a pilot study on the perception of Polish-accented English by different groups of English native speakers. Our main aim here is to compare the evaluative judgments on Polish English pronunciation passed by several speakers of English English (EE) and Irish English (IE) concerning its accentedness and annoyance triggered in listeners. Another goal of the study is to determine which phonetic features contribute most to each of these parameters. Having done that, we shall be able to establish priorities in teaching English phonetics to Poles, particularly to those who intend to emigrate to the British Isles or who already live there and wish to improve their English. It must be stressed that due to its very limited scope, this research is exploratory in nature and its conclusions tentative.

2. EXPERIMENTAL SET-UP

12 judges (7 female, 5 male, aged 25-55), including 7 speakers of Irish English (from Dublin) and 5 speakers of English English (from the Midlands), who live permanently in their countries, took part in the experiment. The two groups are homogenous with respect to their education (all are college or university graduates). They all report some familiarity with Polish accent and claim to have communicated with more than 20 Poles in English.

The experiment consisted of two stages. In Part 1 the listeners were asked to evaluate two samples of Polish English speech for accentedness and annoyance using a 5-point Likert scale (1 = very foreign / extremely annoying, 5 = native-like / not at all annoying). The samples came from two speakers (S1 and S2), aged 33 and 23, representing a different degree of foreign accent: S1 has a stronger accent than S2, according to an impressionistic evaluation of the author. The diagnostic passage (see Appendix 2) was read out by the speakers in order to eliminate the influence of non-phonetic factors.

In Part 2 the raters were provided with a list of 20 phonetic problems typical of Polish English (see Appendix 1) and were asked to indicate to what extent they contribute to the degree of perceived foreign-accentedness and irritation they cause. Since the participants are not phoneticians, much effort has been taken to describe pronunciation problems as plainly as possible.

3. RESULTS AND DISCUSSION

3.1. Part 1 of the experiment

Both IE and EE judges concur in regarding S1's pronunciation as more accented and more annoying than S2's. However, IE raters turned out to be stricter in their evaluations than EE listeners, particularly with respect to the perceived degree of annoyance ($p < 0.009$). Spearman's Rho test has shown a correlation between the degree of foreign accent and annoyance ($r = 0.6$).

3.2. Part 2 of the experiment

3.2.1.Irish raters's evaluation

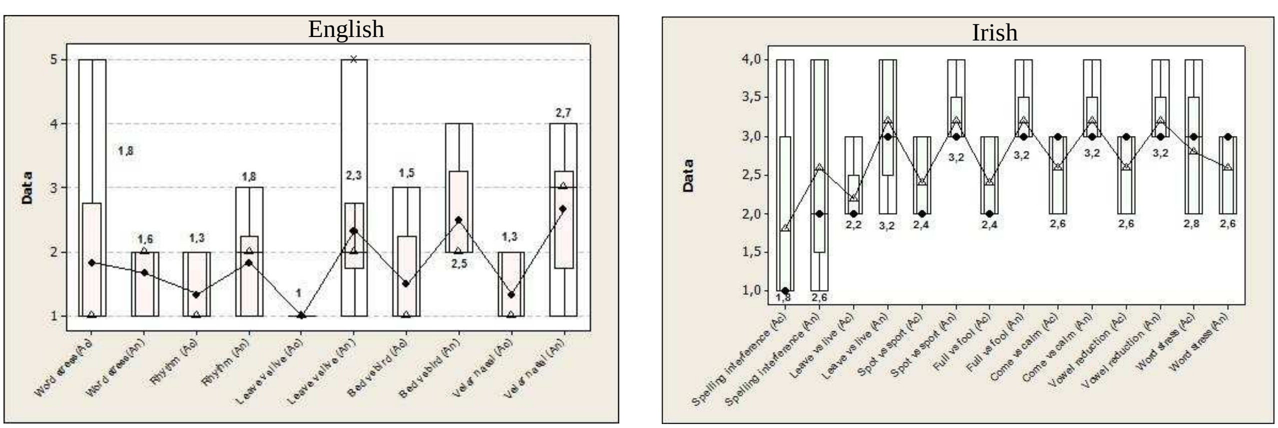
According to IE raters both segmental and suprasegmental features contribute to the impression of foreign accent and encompass: 5 vocalic contrasts, 6 consonants, 7 suprasegmental features and spelling interference. At the top of the list we can find the neutralization of the [iː] vs. [ɪ] vowel contrast, followed by word-final devoicing of obstruents, spelling-based pronunciation and violations of rhythm.

The most annoying features include word-stress errors and rhythm violations in the first place, followed by incorrect vowel realizations ([iː], [ɪ], [əː] and [e]) as well as the pronunciation of the velar nasal with an accompanying velar plosive.

The comparison of factors relevant for accentedness and annoyance judgements shows a similar high position of the role of rhythm and word-stress, which is not the case with the remaining variables. On the whole, the neutralization of the [iː] vs. [ɪ] and [əː] vs. [e] vowel contrasts as well as plosive insertion after the velar nasal contribute more to foreign accentedness than to listeners' irritation. Final devoicing, spelling interference, consonantal substitutions ([ʤ], [ʧ], [θ], [ð] and [ɦ]), failure to discriminate between vowels ([ʊː] vs. [ʊ], [aː] vs. [ʌ] and [ɔː] vs.[ɔ]), wrong intonation patters (also in interrogative sentences), lack of aspiration and vowel reduction trigger the impression of foreign accent without causing annoyance. As can be observed, a great majority of factors that contribute to Polish accent in English are not irritating for Irish raters. Of those which are, only word-stress is slightly more important for annoyance than for accentedness.

More details are presented in Figure 1. The only variables that have been taken into account are those whose mean values for the two parameters are below 4 (= pleasant to listen to / close to native).

Figure 1: Box for the most accented (AC) and the most annoying (AN) features according to EE and IE judges.



ΔMedian ● Mean Quartile- Min max and range

For EE judges foreign accent is strengthened by segmental and prosodic features as well, i.e. 4 vocalic contrasts, 2 consonants, 4 suprasegmental features and spelling interference. The highest place in the hierarchy of the most foreign elements occupies spelling interference followed by incorrect vowel articulation ([iː] vs. [ɪ], [ɔː] vs.[ɔ], [ʊː] vs. [ʊ] and [aː] vs. [ʌ]), the substitution of the dental fricatives and lack of vowel reduction.

The most irritating aspects of Polish English pronunciation include misplaced word-stress and spelling interference. The neutralization of the [iː] vs. [ɪ], [ɔː] vs.[ɔ], [ʊː] vs. [ʊ] and [aː] vs. [ʌ] vocalic contrasts and failure to reduce unstressed vowels complete the list.

In each category the highest position is occupied by spelling interference whose role is by 1.2. points greater in forming the impression of foreign accent than causing irritation. Word-stress has the same (moderate) bearing on accentedness and annoyance. Failure to preserve the aforementioned vocalic contrasts and to reduce vowels in unstressed syllables contributes more to Polish accent than to raters' irritation. While distorted rhythm and consonantal substitutions ([θ], [ð] and [ɦ]) are among the factors strengthening accentedness, they do not bother our judges. Moreover, no phonetic errors have been deemed *very annoying*.

3.2.3.Inter-group comparison

The table below presents those phonetic features that contribute most both to foreign accentedness and annoyance judgements by the two groups of native speakers.

Table 1: The mean values of features contributing to accentedness and annoyance according to IE and EE judges.

<table>
<tr><th rowspan="2">Scale</th><th colspan="4">Irish</th><th colspan="4">English</th></tr>
<tr><th>Annoyance</th><th>M</th><th>Accentedness</th><th>M</th><th>Accentedness</th><th>M</th><th>Annoyance</th><th>M</th></tr>
<tr><td rowspan="12">1. Heavily foreign
1.Very annoying</td><td colspan="2" rowspan="8"></td><td>[iː] vs. [ɪ]</td><td>1</td><td rowspan="9"></td><td rowspan="9"></td><td rowspan="12"></td><td rowspan="12"></td></tr>
<tr><td>Final devoicing</td><td rowspan="2">1.1</td></tr>
<tr><td>Spelling interference</td></tr>
<tr><td>Rhythm</td><td rowspan="2">1.3</td></tr>
<tr><td>Velar nasal</td></tr>
<tr><td>[əː] vs. [e]</td><td rowspan="4">1.5</td></tr>
<tr><td>Palato-alveolars</td></tr>
<tr><td>Dental fricatives</td></tr>
<tr><td>Word-stress</td><td>1.6</td><td>[ʊː] vs. [ʊ]</td></tr>
<tr><td rowspan="3">Rhythm</td><td rowspan="3">1.8</td><td>[aː] vs. [ʌ]</td><td rowspan="3">1.8</td><td rowspan="3">Spelling interference</td><td rowspan="3">1.8</td></tr>
<tr><td>Word-stress</td></tr>
<tr><td>[ɔː] vs. [ɔ]</td></tr>
<tr><td rowspan="7">2. Moderately foreign
2. Moderately annoying</td><td colspan="2" rowspan="3"></td><td>Glottal fricative</td><td rowspan="3">2</td><td rowspan="3">[iː] vs. [ɪ]</td><td rowspan="3">2.2</td><td colspan="2" rowspan="3"></td></tr>
<tr><td>Intonation</td></tr>
<tr><td>Vowel reduction</td></tr>
<tr><td>[iː] vs. [ɪ]</td><td>2.3</td><td>Aspiration</td><td rowspan="2">2.5</td><td>[ɔː] vs. [ɔ]</td><td rowspan="2">2.4</td><td colspan="2" rowspan="2"></td></tr>
<tr><td>[əː] vs. [e]</td><td>2.5</td><td>Intonation in interrogative sentences</td><td>[ʊː] vs. [ʊ]</td></tr>
<tr><td rowspan="2">Velar nasal</td><td rowspan="2">2.6</td><td colspan="2" rowspan="2"></td><td>Dental fricatives</td><td rowspan="2">2.6</td><td>Spelling interference</td><td rowspan="2">2.6</td></tr>
<tr><td>[aː] vs. [ʌ]</td><td>Word-stress</td></tr>
</table>

				Vowel reduction			
				Word-stress	2.8		
				Rhythm			
				Glottal fricative			
						[iː] vs. [ɪ]	3.2
3. Slightly foreign						**[aː] vs. [ʌ]**	
3. A bit annoying						**[ɔː] vs. [ɔ]**	
						[ʊː] vs. [ʊ]	
						Vowel reduction	

Two complementary tests (chi-square and ANOVA) were used to determine which properties were evaluated differently by the two groups of raters. In both tests the difference is statistically significant for $p < \alpha$, $\alpha = 0.05$. In the descriptions below only statistically significant differences are mentioned.

The participants regard spelling-based pronunciation as the greatest obstacle to sounding native-like. Lack of vowel reduction as well as the substitution of the glottal fricative have been categorized as moderately foreign. Distorted word-stress ($p = .0017$) and rhythm ($p = .000$) contribute to Polish accent more for IE than EE raters. The list of phonetic features which strengthen the impression of foreign accentedness is more extensive for IE judges and includes final devoicing ($p = .000$), replacing palato-alveolars with their Polish postalveolar equivalents ($p = .008$), failure to aspirate stressed fortis plosives and improper intonation ($p = .0048$). IE listeners are more severe in the scores they assign than their EE counterparts.

Both groups of judges consider inappropriate word-stress the most irritating aspect of Polish English pronunciation followed by the neutralization of the [iː] vs. [ɪ] contrast. It should be pointed out that IE rate these two factors much lower than EE ($p = .017$). Rhythm violation, non-target renditions of [əː] and [e] as well as velar plosive insertion after the velar nasal are annoying to IE listeners, whereas spelling interference, vowel reduction and lack of distinction between [ɔː] vs.[ɔ], [ʊː] vs. [ʊ] and [aː] vs. [ʌ] irritate EE judges. On the whole, the preservation of proper suprasegmental properties of English is more crucial for sounding pleasant than segmental aspects. Of the latter, vowels are more relevant than consonants.

Interestingly, even though the dental fricatives are listed by both groups among the features typical of Polish accent, they are not perceived as annoying. Relatively little attention paid by the judges to the proper articulation of the interdentals echoes Scheuer's study in which English listeners '*seemed relatively immune to this acoustic cue*' (2003: 98).

3.3. Pedagogical implications

Kenworthy (1987: 8) argues that learners' lack of concern about phonetics may stem from the fact that "*they are simply not aware that the way they speak is resulting in difficulty, irritation or misunderstanding for the listener*." Having realized what reaction is evoked by particular phonetic distortions '*speakers (…) can alter the way in which they speak in order to best fit into the mainstream society*" (Giles&Billings, 2004: 199).

Below we enumerate the areas[1] which should be given priority when teaching English to Poles who intend to communicate with IE and EE speakers since they either create the impression of a strong foreign accent, are annoying for the listeners, or both:

1. **Eliminating spelling-based pronunciation**.
2. **Proper word-stress**.
3. **Preserving the following vowel contrasts: [iː] vs. [ɪ], [ɔː] vs. [ɔ], [ʊː] vs. [ʊ] and [aː] vs. [ʌ]**.
4. **Vowel reduction**.

5. **The correct articulation of the dental fricatives**.

6. **The correct articulation of the glottal fricative**.

Features such as final devoicing, proper distribution of the velar nasal, [ɘː] and [e], the palato-alveolars, rhythm and intonation should also be practiced for IE target interlocutors.

4. CONCLUSION

This study was concerned with the evaluation of Polish English pronunciation by Irish and English native speakers with respect to its accentedness and resulting annoyance.

First, the examination of the participants' reactions to two samples of Polish-accented speech showed that even though the listeners agreed as to which speaker's pronunciation was more accented and more annoying, IE listeners turned out to be harsher in their judgements.

Secondly, the analysis of phonetic features which contribute to the impression of foreign accent and cause irritation in IE and EE raters uncovered that the majority of properties characteristic of Polish English pronunciation do not evoke listeners' annoyance. The comparison of features responsible for the obtained judgments revealed that spelling-based pronunciation and misplaced word-stress were regarded as most detrimental to sounding native-like and pleasant, respectively. Both IE and EE judges also agreed that the failure to employ vowel reduction as well as the substitution of the glottal fricative sound foreign and that neutralization of the [iː] vs. [ɪ] contrast is rather annoying.

Additionally, variety-specific features responsible for foreigness and irritation judgements were pointed out proving that native-speakers do not form a homogenous group with respect to accent perception. IE raters selected more features which contribute to the impression of foreign accent. Significant differences were found with regard to the evaluation of rhythm violation, non-target renditions of [ɘː] and [e], velar plosive insertion after the velar nasal (annoying to IE judges) as well as spelling interference, vowel reduction and lack of vocalic contrasts (annoying to EE judges).

Finally, we singled out those areas which should be given priority when teaching English to Poles, and emigrants to the British Isles in particular. A special emphasis should be laid on suprasegmental properties, i.e. word-stress and vowel reduction, as well as eliminating spelling-based pronunciation. Of segmental features 4 vocalic contrasts ([iː] vs. [ɪ], [ɔː] vs.[ɔ], [ʊː] vs. [ʊ] and [aː] vs. [ʌ]) and 3 consonants ([θ], [ð] and [ɦ]) should receive special focus. For Poles intending to interact with more demanding IE speakers the list of priorities is more extensive.

5. REFERENCES:

Giles, H., A.C. Billings (2004). Assessing Language Attitudes. Speaker Evaluation Studies. In: Davies, A., Elder, C. (eds.) *The handbook of Applied Linguistics*. Oxford: Blackwell Publishing Ltd.

Gonet, W., Pietroń, G. (2004) The Polish tongue in the English ear. In: Sobkowiak, W. , Waniek-Klimczak, E. (eds.) *Zeszyty Naukowe PWSZ w Koninie*, 1/ (4), 56-65.

Kenworthy, J. 1987. *Teaching English Pronunciation*. London and New York: Longman

Majer, J. 1992. Sick or seek? Pedagogical Phonology in teacher training. In: Waniek-Klimczak, E, Melia, J. (eds.) *Accents and Speech in Teaching English Phonetics and Phonology*. Berlin: Peter Lang 153-176

Nowacka, M. 2009. Native and non-native perception of Polish-accented speech: do Polish and native raters have the same ear for nativelikeness? *Issues in Accents, of English: Variability and Norm.* Cambridge: Cambridge Scholars Publishing

Prator, C. A. 1985 < http:// web. ntpu. edu. tw/~ ckliu/ pronunciation/ clifford. pdf.> retrieved August, 6th 2009

Scheuer, S.2003 What to teach and what not to teach? Some reflections on the relative salience of L2 phonetic errors. In: Sobkowiak, W., Waniek-Klimczak, E. *Zeszyt Naukowy Neofilologii (2)*. Konin: PWSZ

Szpyra-Kozłowska, J. 2005. Intelligibility versus Polish accent in English. *Studia Phonetics Posnaniensia*. *An International Journal for Linguistic Phonetics* 7. 59-73

Szpyra-Kozłowska, J. Stasiak, S. 2009. From Focus on Sounds to Focus on Words in English Pronunciation Instruction. (in press)

6. APPENDIX 1.

1. their speech sounds flat and monotonous
2. they ‘chop’ sentences making unnatural pauses between groups of words
3. they raise their voice excessively when asking questions
4. they do not make a difference between vowels in words like *c**a**lm* and *c**o**me*
5. they often accentuate a wrong syllable in a word, e.g. say *COMputer* instead of *comPUTER* or *DEVelopment* instead of *deVElopment*, *Event* instead of *eVEnt*
6. they sometimes put too much emphasis on some words in a sentence, e.g. *Should he spend all OF his time just studying?*; *He also learns TO choose the language and customs right for various situations*.
7. they do not make a difference between vowels in words like *sp**o**rt* and *sp**o**t*
8. they often pronounce words as they are spelt, e.g. *doubt* [daubt], *calm* [kalm]
9. they do not make a difference between vowels in words like *leave* and *live*
10. they pronounce [k] in *kind* or [t] in *time* too weakly, with no extra breath
11. the way they pronounce [r] sound, e.g. in *right*, *various* or *across*
12. they do not make a difference between vowels in words like *f**u**ll* and *f**oo**l*
13. they pronounce [r] in e.g. *manner*, *first* or *mark*
14. they pronounce [h] sound too harshly, e.g. in *house*, *his* or *he*
15. they pronounce ‘th’ in *clothing* as [v] in *very*, [d] in *dog* [z] in *zebra*
16. they put a vowel between two final consonants e. g. *custom* as [‘kʌstom]
17. they pronounce *is* as [i**s**] or *spend* as [spen**t**]
18. they pronounce the underlined part in ***sh**ould*, spee**ch**, *ca**s**ually*, ***j**ust* too harshly
19. they do not make a difference between vowels in words like *b**e**d* and *b**i**rd*
20. they say *sitti**ng*** as [sitti**nk**], *lo**ng*** as [lo**nk**]

7. APPENDIX 2.

When a student decides to go to study abroad, he might have many questions and more than one doubt. Where should he live? Share a flat, or look for a bed in a dormitory? Should he spend all of his time just studying in front of the computer? He’d better calm down, because marks are not the most important thing. He should live life to the full and take advantage of the many social and sport events which are offered. At first it is not easy for him to be comfortable in manner and confident in speech. He feels like a fool or comes across as a rare bird. Little by little he spots what kind of clothing is usually worn to be casually dressed for classes. He also learns to choose the language and customs right for various situations. But let me tell you, my friend, this long-awaited feeling doesn’t develop fast, does it?

(An extract adapted from Prator, 1985).

NOTES

[1] The outcome of our research alludes to previous studies on Polish-accented English, e.g. Szpyra- Kozłowska & Stasiak (2009, in press), Nowacka (2008), Szpyra-Kozlowska (2005), Gonet&Pietroń (2004), Scheuer (2003) and Majer (2002). It should be noted that most of them employed speakers of just one variety of English as judges of Polish English pronunciation and concentrated on a smaller selection of phonetic features.

Teaching foreign sC onset clusters:
Comparing the effects of three types of instruction

Walcir Cardoso

Concordia University & The Centre for the Study of Learning and Performance
walcir@education.concordia.ca

ABSTRACT

This study explores the effects of three types of instruction on the development of foreign homorganic onset clusters (/sl/, /sn/, /st/ - sC). Thirty-seven native speakers of BP participated in a four-week language course to learn Slavir, a language designed for this study, "famous for having the largest number of sC-initial words". The participants were divided into three groups, each corresponding to one of the hypotheses associated with the three types of instruction: Teachability (Pienemann 1984), Projection of Markedness (Projection; Zobl 1983), and a combination of the two (Mixed; Shirai 1997; Lightbown 1998). A mixed ANOVA analysis of the participants' production of sC-initial words in Slavir revealed that the group that focused exclusively on the more marked /st/ (Projection group) had the best overall performance in the acquisition of the three clusters, followed by the Teachability and Mixed groups, in that order. In general, the results support Zobl's (1983) Projection Model of Markedness wherein the instructional effect of a focus on the most marked /st/ projects to the acquisition of the less marked structures.

Keywords: Projection of Markedness, sonority, teachability, English, Brazilian Portuguese.

1. INTRODUCTION

The literature on phonological acquisition is replete with studies that show that the oral production of /s/ + consonant onset clusters (sC; e.g., /st/op, /sl/eep) is particularly problematic for first (L1) and second or foreign language (L2) learners (e.g. L1 acquisition: Goad and Rose 2004; Yavaş and Barlow 2006; L2 acquisition: Cardoso and Liakin 2009; Carlisle 1988, 2006). Some of these studies also reveal that the acquisition of sC follows a "natural order of acquisition" or developmental sequence in which the /s/ + sonorant sequences /sl/ and /sn/ tend to appear before /st/ (e.g., Boudaoud 2008; Cardoso and Liakin 2009; Yavaş and Barlow 2006; Carlisle 1988, 2006). From a pedagogical perspective, these studies raise an interesting question with regards to instructional intervention: Will the effects of a focus on the form that is acquired late (and assumed to be difficult) project to the forms that are usually acquired early (and assumed to be easy)? Or will the reverse lead to a more successful mastery of developmental sequences, as is often implied in the design of L2 instructional materials in which these sequences are often introduced starting from the easy end of the hierarchy (Ellis 2002, Yabuki-Soh 2007)? The L2 literature provides a limited number of studies that show the effects of either (or a combination) of these two views on the instruction of morphosyntax (e.g. Gass 1982; Ellis 1984; Shirai 1997). Surprisingly, there are no studies that investigate the acquisition of developmental sequences from a phonological perspective, except for suggestions that problematic L2 sounds be taught first in the most difficult and latest acquired environment (Eckman and Iverson 1997; see also Doughty and Williams 1999: 219 for a similar observation). It is one of the goals of this study to address the issue from a phonological perspective and thus lay groundwork for future research on the acquisition of phonological developmental sequences.

2. BACKGROUND

The production of foreign /s/ + consonant onset clusters is notoriously difficult for second or foreign language learners whose first languages disallow such sequences (e.g. Carlisle 1988; Major 1996). In the context of Brazilian Portuguese (BP) speakers learning an "sC language" such as English, French or German, for instance, learners variably syllabify the cluster via a prothetic [i] (i-epenthesis), thus triggering the resyllabification of the original onset into a nucleus-coda sequence (e.g. /st/op → [is.t]op, /sn/ow →

[is.n]ow). While i-epenthesis can be usually attributed to an L1 effect (sC onset clusters are non-existent in BP), the phenomenon is rather complex and is motivated by a variety of linguistic and extralinguistic factors (Cardoso and Liakin 2009). One of these factors include the concept of sonority, defined via a combination of features that include amplitude or intensity (Ladefoged 1993), acoustic energy (Goldsmith 1989), and propensity for voicing (Kenstowicz 1994). Together, and focusing exclusively on the set of relevant segments, these features determine a sonority hierarchy that ranges from the least sonorous stop /t/ to the more sonorous liquid /l/: /t/ + /s/ + /n/ + /l/ (where "+" indicates "more sonorous than"). In order to constitute onset clusters, the /s/ + consonant combination should follow a pattern in which sonority progressively rises towards the nucleus of the syllable (Sonority Sequencing Principle - SSP, Selkirk 1984). While both the /sl/ and /sn/ satisfy the SSP requirement because sonority rises from /s/ to the following segment, /st/ constitutes an SSP violation because sonority sequencing decreases in the second consonant. In addition, onset cluster syllabification tends to favour sequences that have a "maximal and most evenly-distributed rise in sonority" (Minimal Sonority Distance - Clements 1990). This preference favours the sequence /sl/, which has a wider sonority distance than /sn/. Appealing to the concept of markedness (e.g. de Lacy 2006), one may then assume that these three sequences constitute a hierarchy in which /sl/ is the least marked, followed by /sn/ and then the most marked /st/: /sl/ < /sn/ < /st/ (where "<" indicates less marked than"). The implication of this generalization based on markedness is that learners will have less difficulty in acquiring the least marked /sl/ than the more marked /sn/ and /st/ clusters. Not surprisingly, this is exactly what is found in a number of studies that investigate L2 sC acquisition, as will be discussed next.

The majority of the literature on L2 sC acquisition indicates that the path to sC development initially favours unmarked segments such as /s/ + liquids, which are acquired earlier and with less difficulty than their more marked counterparts (e.g., Boudaoud 2008; Carlisle 1988, 1991, 2006; Escartin 2005; Tropf 1987; but see Major 1996 and Rauber 2006 for slightly different results involving different communities of BP L1 speakers). In a study involving the same community of BP speakers examined in this investigation, Cardoso and Liakin (2009) found a relatively similar pattern in which 10 participants produced /sl/ and /sn/ (the difference between these two SSP-abiding clusters was not significant) more accurately than the more marked /st/. Interestingly, similar order of acquisition patterns have also been observed in L1 acquisition (e.g. Yavaş and Barlow 2006).

These results allow us to establish a "natural order of acquisition" or developmental sequence for sC clusters, one that reflects the markedness hierarchy just discussed: /sl/ (>>) /sn/ >> /st/ (where ">>" indicates "acquired before" and "(>>)" suggests an inconclusive pattern). The concept of a natural order of acquisition, formalized by Krashen (1981) in the form of the Natural Order Hypothesis, predicts that the acquisition of certain linguistic forms follow a predictable developmental sequence, regardless of the learners' L1 (see Kwon 2005 for a review of the literature on natural order morphemes).

From a pedagogical perspective, it remains unclear whether the instruction of developmental sequences should start at either of the two ends of the easy/less marked to hard/more marked hierarchy, or if, instead, it should consist of an equal combination of the items that encompass the developmental set. This conundrum yielded three proposals that reflect exactly these three instructional possibilities. The Teachability Hypothesis (Pienemann 1984) predicts that a novel linguistic structure can only be acquired when learners are cognitively mature - they must follow a natural order of acquisition in which stages cannot be skipped. L2 acquisition studies that reinforce this hypothesis include those of Bardovi-Harlig (1995 - English pluperfect), Ellis (1984 - English Wh-questions in children; 1989 - German word order), and Piennemann et al. (1988 - English word order). The Projection Model of Markedness (Zobl 1983, 1985), in contrast, predicts that an instructional focus on advanced/marked structures leads to the learning of earlier/unmarked structures. This hypothesis is supported by a large body of literature including the following studies: Eckman et al. (1988 - English relative clauses), Gass (1982 - English relative clauses), Mitchell (2001 - French relative clauses), Yabuki-Soh (2007 - Japanese relative clauses), and Zobl (1985 - English possessive determiners). Finally, the Mixed Approach (Shirai 1997; Ammar and Lightbown 2004) is a combination of the two previous approaches and, accordingly, it advocates an emphasis on all items that comprise a given developmental sequence. Among the studies that support this hypothesis is that of Ammar and Lightbown

(2004 - English relative clauses). See Yabuki-Soh (2007) for a comprehensive discussion of these hypotheses for the L2 acquisition of morphosyntax.

Absent from the current literature is an examination of the effects of these proposals on the outcome of phonological developmental sequences such as sC clusters. One of the goals of this study is to address this gap in the literature. In addition, the current study aims to examine sC development from a laboratory perspective, one that can easily monitor some of the factors that are known to affect L2 acquisition (e.g., frequency distribution of sC in the L2 input, word familiarity; Bybee 2007).

3. RESEARCH QUESTIONS AND PREDICTIONS

The purpose of the present study is to explore the effects of three types of instruction on the development of foreign sC onset clusters and, at the same time, to observe how these same clusters are acquired at the end of a series of teaching sessions. This study's research questions are thus formulated as follows:

1. Which type of instruction is more effective in the learning of sC cluster production?
2. How are foreign sC clusters acquired in three distinct instructional settings? More specifically, how do the patterns observed compare with those found in the second language acquisition literature?

It remains difficult to predict outcomes pertaining to question 1 because, as discussed earlier, the literature on L2 instruction and developmental sequences is inconclusive for the acquisition of structures that follow a natural sequence of development and, more importantly, non-existent for phonological L2 development. Accordingly, three outcomes are possible, each favouring one of the instructional types under consideration. With respect to question 2, it is predicted that the acquisition of sC will reflect the predictions based on sonority and its markedness effects, and to what is usually observed in the development of these clusters in second (and first) language acquisition: The marked /st/ structure will be produced comparatively less accurately than the less marked /sn/ and /sl/, in that order.

4. METHOD

Thirty students participated in this study. They were grade 2 and 3 students (16 females, 14 males) with ages ranging from 15 to 20 years (M = 16.3, SD: 1.06), enrolled in a standard secondary public school in the city of Belém (Brazil), a monolingual community of Portuguese speakers. Accordingly, the participants were all monolingual native speakers of Portuguese, without any previous oral experience with an sC language (including both speaking and listening). Students with formal learning experience of an sC language were excluded from the pool of potential participants. As part of the secondary curriculum in Brazil, students have compulsory weekly 1-hour English classes that focus exclusively on the acquisition of morphosyntax (grammar), receptive (written) vocabulary and reading comprehension in order to fulfill the requirements for the high-stakes tests (vestibular) for entry into a university program. Oral interactions in English between the instructor and the students and their peers are non-existent. The thirty participants were assigned to one of the three experimental groups, using gender as the sole criterion for the distribution: The Projection of Markedness Group (P Group; 5 females, 5 males), the Teachability Group (T Group; 6 females, 4 males) and the Mixed Group (M Group; 5 females, 5 males).

The study employed a quasi-experimental, within groups pretest/posttest design. To observe the developmental stages of sC acquisition, two additional tests were included after each teaching session (total of tests: 4 - Pretest, Test 1, Test 2, Test 3 (posttest)). The experiment lasted five weeks and consisted of three teaching sessions, each designed according to the three types of instruction considered in the study: While the P Group was taught exclusively /st/-initial words, the T Group was taught one sC per session, following their natural order of acquisition (/sl/ > /sn/ > /st/). Finally, the M Group was taught all three sC sequences throughout the duration of the Slavir course.

The testing materials used in this study (including teaching sessions and tests) consisted of 75 sC-initial words (/st/ = 33, /sn/ = 21, /sl/ = 21) created using WordGenerator 1.9 (http://billposer.org/Software/WordGenerator.html), a computer application that generates hypothetical words based on specifications such as segmental content, syllable structure and word size. Following a combination of English and BP phonotactics, the selection of the words observed the following criteria: (1) Word size: words were all disyllabic (e.g., [slu.ba] 'book'); (2) Foot structure: trochaic, stressed on the leftmost syllable as is the case for most English sC-initial words (e.g., ['sna.pu] 'bird'); (3) Skeletal structure: sCV.CV (where C = consonant and V = vowel; e.g., [stu.bi] 'tree'); (4) Segmental content: the vowels ([a e i o u] and consonants (e.g., [p t k g b d] consisted of segments that exist in BP; and (5) Semantic content: Words were assigned to specific meanings randomly. The rationales behind these criteria were motivated by attempts to manipulate only the relevant foreign sC onset and to minimize the influence of potential extraneous factors (e.g., an illicit coda such as [k] in [slu.bik] could jeopardize the intended focus on the sC sequence).

The Slavir teaching sessions were taught by one fluent female English teacher who had no problems in pronouncing the relevant sC clusters. She received approximately two hours of training on how to conduct a typical Slavir class. On the first day of class, the participants were told that, within the period of a month, they were going to learn some words from Slavir, "a minority language spoken in eastern Kazakhstan, famous for having the world's largest number of sC-initial words”. The weekly instructional sessions were conducted in a standard classroom in the school premises and lasted approximately 30 minutes. The teaching sessions, conducted in the participants' native language, focused mostly on vocabulary with a secondary focus on pronunciation. Briefly, the sessions consisted of the following teaching strategies (based on Thornbury's 2002 recommendations for teaching vocabulary): (1) Introduction of the word via a 8" by 11" flashcard with a picture and its associated sC word; (2) Pronunciation via the production of the word followed by choral repetitions; (3) Personalization activities (e.g., "[stova], I use it to listen to music"); (4) Dictations (for the development of sC awareness); and (5) Word retrieval activities (via translations: "what's the word in Slavir for "hand"?, and elicitations and gestures: "what's that thing over there?"). Whenever possible, an attempt was made to present and discuss the sC words in preceding pausal environments (e.g., "[slova], this is the word for hand in Slavir") to prevent them from being lost via resyllabification (e.g., [a slova] → [as.lova] 'the hand', where the original /sl/ onset sequence syllabifies as a coda-onset cluster). Finally, to reduce the effects of type/token frequency distribution in the L2 input, the amount of oral and visual exposure to each sC word was carefully monitored so that they received the same quantity and quality of treatment across the three experimental groups.

To test the participants' "proficiency in Slavir" and thus establish an initial state for sC acquisition, they were tested on their first meeting with the Slavir teacher via a word reading aloud task containing 15 sC-initial words (5 of each sC type) and 15 Portuguese-like pseudowords used as distractors (pretest). Relatively similar tests were administered after each of the three teaching sessions. Tests 1 through 3 consisted of 15 new sC-initial items mingled with some of the words taught in the classroom, used as distractors and consequently discarded from the analysis. The production tests were audio recorded via a Marantz CDR300 CD/RW Recorder and an Audio-Technica AT831b lavaliere microphone. The data collected were then transcribed (using the software Transcriber 1.4) independently by two research assistants and me and, whenever necessary, analyzed acoustically via Praat 5.1.04. All instances of sC clusters were then coded according to their accuracy in production (sC or [i]sC), as well as the following independent variables: sC cluster (/sl/, /sn/, /st/), Test (pretest, 1, 2, 3) and Instructional Group (P, T, M). A mixed between-within subjects analysis of variance was used to calculate differences between the independent variables included in the investigation.

5. RESULTS

The descriptive statistics for the results appear in Table 1. Observe that the results for the pretest are not provided because, in that test, the participants categorically transferred their knowledge of the phonotactics of Portuguese into Slavir, producing all sC sequences preceded by the epenthetic vowel [i]. An initial mixed between-within subjects analysis of variance did not show a significant interaction between the three variables under consideration, namely test (time), sC cluster and instructional group, $F(8, 108) = 1.67$, $p = .11$. Consequently, no pairwise analyses could be carried out. They results in Table 1 show that, overall, the P Group outperformed the other two groups in sC production, particularly in later stages of sC development, represented here as tests 2 and 3 (posttest).

Table 1: Descriptive statistics for sC production

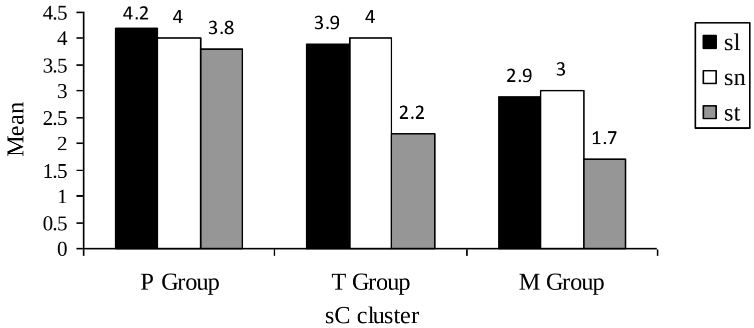
	P Group						T Group						M Group					
	sl (n = 5)		sn (n = 5)		st (n = 5)		sl (n = 5)		sn (n = 5)		st (n = 5)		sl (n = 5)		sn (n = 5)		st (n = 5)	
Test	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1	1.7	.95	1.4	.84	1.8	.63	2.6	.52	0.9	.74	0.2	.42	2.2	.42	1.7	.82	1.2	.63
2	2.1	.32	2.3	.48	2	.82	2.5	.53	2.3	.68	0.2	.42	1.5	.85	1.9	.99	1	.47
3	4.2	.63	4	.67	3.8	.63	3.9	.57	4	.47	2.2	.79	2.9	1.1	3	.94	1.7	.65

Because the research questions addressed in this study focus on the effects of different types of instruction on the outcome of sC instruction (the "final" stage), and due to the fact that the analysis of variance revealed no significant interaction between the original variables included in the investigation, another mixed between-within subjects analysis of variance was conducted including only the last stage of sC development (test 3). This analysis confirmed a significant interaction between sC cluster and group, $F(4, 54) = 3.15$ $p = .02$. This indicates that, in Test 3, the three experimental groups behaved significantly differently with respect to the production of sC clusters. Post hoc pairwise comparisons (based on a Bonferroni adjustment of alpha) were used to determine the mean differences between sC clusters and groups, as will be reported in the discussion that follows.

The first research question asked: Which type of instruction is more effective in the learning of sC cluster production? The post hoc comparisons revealed that, with respect to the /sl/ cluster, the P and T groups both outperformed the M Group ($p = .004$ and $p = .029$ respectively). A similar pattern was also detected for the set of /sn/ clusters, in which the T and M groups outperformed the M Group ($p = .013$ for both). Finally, the production of the most marked /st/ clusters was significantly improved in the P group and less likely so in both the T and M groups ($p < .001$ for both interactions; see also Figure 0 for an alternative depiction of the results). In sum, the group that was taught the more marked /st/ (P Group) had the best performance in the acquisition of sC clusters, as predicted by the Projection Model of Markedness Hypothesis.

The second question asked: How are foreign sC clusters are acquired in the three distinct instructional settings? How do the patterns observed compare with those found in L2 acquisition? For ease of exposition, the mean distribution of sC clusters by instructional group is illustrated in Figure 1. The post hoc comparisons detected no significant differences among the sC clusters in the P Group. In Group T, however, the clusters /sl/ and /sn/ were more accurately produced than /st/ ($p < .001$ for both groups). The same pattern was also observed in Group M in which both /sl/ and /sn/ were more accurately produced than /st/ ($p = .009$ and $p = .002$ respectively). To summarize, the results partially conform to the hypothesis based on markedness involving sonority, which predicts that the form that violates sonority sequencing (/st/) will be acquired with more difficulty and, accordingly, mastered later in the acquisition process. This is exactly what was observed for groups T and M, but not for Group P. In addition, the markedness relationship between /sl/ and /sn/ was not confirmed in this study, as the difference between these two clusters was not significant in any of the three instructional groups considered.

Figure 1: Means of sC production across the three experimental groups in Time 3 (Posttest)



6. DISCUSSION AND CONCLUSION

The main goals of this study were to test the effects of three types of instruction on the development of foreign homorganic sC onset clusters and, at the same time, to investigate the development of these clusters in three controlled teaching environments.

With regards to the first goal, the findings reported yield support for Zobl's (1993) Projection Model of Markedness inasmuch as the group that was taught the more marked /st/ (P Group) achieved the best overall results in the production of the three sC clusters. This is exactly what the model predicts for linguistic items that are implicationally related in acquisition: the instructional effects of mastering the most marked /st/ cluster projects to the acquisition of the less marked forms /sl/ and /sn. The second best overall performance was found among participants included in the T Group, where participants were taught sC forms progressively from the easiest and less marked /sl/ to the more difficult and marked /st/. These results suggest that a piecemeal (rather than an all at once) approach to teaching may yield interesting pedagogical benefits. A possible explanation may be due to a potential cognitive load that the all-at-once approach has on learners arising from the way information is presented (a type of extraneous cognitive load - Chandler and Sweller 1991).

The findings relating to the second goal reveal that, overall the development of sC clusters across the three experimental groups partially conforms to what is predicted by sonority and its markedness effects and some of the current literature on sC acquisition. As predicted, the most marked /st/ cluster posed a higher level of difficulty in the three groups, thus conforming to the hypothesis that this cluster should be positioned at the more marked (and late acquired) end of the sC developmental sequence. The results pertaining to a developmental sequence between /sl/ and /sn/, as predicted by Clements' (1990) Minimal Sonority Distance (MSD) and some of the previous studies (e.g., Carlisle 1988, 2006), was not borne out, since the difference in performance between the two clusters was not significant across the three instructional groups. This suggests that although these L2 learners are highly sensitive to sonority sequencing and its markedness effects on syllabifying foreign sC clusters, they remain oblivious to other principles such as the MSD and, consequently, process these clusters in a bipartite way in which /s/ + sonorants (/sl/ and /sn/) pattern together as a set in opposition to the most marked /st/ cluster. Not being an idiosyncrasy of BP L1 speakers, similar patterns have also been found for other language backgrounds in both L1 (e.g. Yavaş and Barlow 2006) and L2 acquisition (e.g. Boudaoud 2008; and Cardoso and Liakin 2009 who examined the acquisition of English sC by the same speech community investigated in this study).

This study has shown that the type of instruction used in the classroom plays an important role in the acquisition of sC clusters. It has also shown that a piecemeal introduction to novel L2 forms may lead to better performance in oral production, particularly if the introduction starts from the more difficult and marked end of the developmental sequence, as recommended by Eckman and Iverson (1997).

7. SELECTED REFERENCES

Ammar, A., Lightbown, P. 2004. Teaching marked linguistic structures: More about the acquisition of relative clauses. In Housen, A., Pierrard, M. (eds.), *Investigations in instructed second language learning. Berlin: Mouton de Gruyter.* 167–198.

Bardovi-Harlig, K. 1995. The interaction of pedagogy and natural sequences in the acquisition of tense and aspect. In Eckman, R. et al. (eds.), *Second language acquisition theory and pedagogy*. Mahwah, NJ: Lawrence Erlbaum. 151-168.

Cardoso, W., Liakin, D. 2009. When input frequency patterns fail to drive learning. In Baptista et al. (eds.), *Recent Research in Second Language Phonetics/Phonology: Perception and Production*. Newcastle Upon Tyne: Cambridge Scholars. 174-202.

Carlisle, R. 1991. The influence of environment on vowel epenthesis in Spanish/English interphonology. *Applied Linguistics 12*. 76-95.

Chandler, P. & Sweller, J. (1991). Cognitive Load Theory and the Format of Instruction. *Cognition and Instruction* 8(4). 293-332.

Doughty, C., Williams, J. 1999. Pedagogical choices in focus on form. In XXXX (eds.), *Focus on form in classroom second language acquisition*. Cambridge: Cambridge University Press. 197-262.

Eckman, F., Iverson, G. 1998. Structure Preservation in Interlanguage Phonology. In Hannahs, S., Young-Scholten, M. (eds.), *Phonological Acquisition*. Amsterdam: John Benjamins. 183 – 207.

Pienemann, M. 1984. Psychological Constraints on the Teachability of Languages. *Studies in Second Language Acquisition* 6. 186-214.

Shirai, Y. 1997. Linguistic theory and research: Implications for second language teaching. In Tucker, G., Corson, D. (eds.), *The encyclopedia of language and education: Vol. 4. Second language education*. Dordrecht: Kluwer Academic. 1-9.

Yabuki-Soh, N. 2007. Teaching relative clauses in Japanese. *Studies in Second Language Acquisition* 29. 219-252.

Yavaş, M., Barlow, J. (2006). Acquisition of #sC clusters in Spanish-English bilingual children. *Journal of Multilingual Communication Disorders 4*(3). 182-193.

Zobl, H. 1983. Markedness and the projection problem. *Language Learning* 33. 293-313.

The Production of /.sC/ Onsets in a Markedness Relationship: A Longitudinal Study

Robert S. Carlisle

California State University, Bakersfield
rcarlisle@csub.edu

Juan Antonio Cutillas Espinosa

Universidad de Murcia
jacuti@um.es

ABSTRACT

This paper reports Time 2 findings from a longitudinal study. Time 1 research, involving 17 participants, found significant differences in the correct production of three onsets in a markedness relationship based on their sonority profiles (Clements 1990): /.sl/, /.sn/, and /.st/; /.sl/ was produced correctly more frequently than /.sn/ or /.st/, and /.sn/ was produced correctly more frequently than /.st/. Time II data was gathered exactly one year after Time 1 data and involved 12 of the original 17 participants.

Findings from Time 2 reveal the same significant differences in the correct production of the onsets as were found at Time 1. The participants did not produce a significantly greater percentage of the onsets correctly at Time 2, but a significant interactive effect revealed that over time the percentage of correct production increased linearly as the markedness of the onsets increased.

Keywords: markedness, onsets, sonority, longitudinal study

1. INTRODUCTION

Over the last several decades a number of studies have found that L2 learners modify less marked onsets significantly less frequently than they do more marked onsets and that less marked onsets reach a criterion level of acquisition before more marked onsets do. One group of studies examined the modification of onsets differing in length, markedness increasing with the length of the onset.

1.1. Markedness by Length of Onset

Phonologists universally recognize that the markedness of onsets increases with length (Clements 1990; Morelli 2003; Greenberg 1978), and research in SLA has consistently shown that shorter onsets are less frequently modified or reach a criterion level of acquisition before longer onsets do. Anderson (1987) examined the frequency with which 20 speakers each of Egyptian Arabic and Amoy and Mandarin Chinese modified English onsets, finding that both groups of participants made significantly more modifications of onsets as their length increased. Arabic speakers did not modify simple onsets at all, but they modified over 7% of the biliteral onsets. In turn, the Chinese speakers modified only .8% on simple onsets but 10.4% of the biliteral onsets.

Eckman (1991) examined the production of three triliteral onsets in four tasks by 11 participants who were native speakers of Japanese, Cantonese, and Korean, none of which allow complex onsets. Instead of using a test of statistical significance to determine whether the participants modified more marked onsets more frequently than less marked onsets, Eckman used a criterion measure of 80% correct production to determine whether the participants had acquired the structure under investigation. For example, if a participant produced onsets of the form /.spr/ correctly 80% of the time, the structure was regarded as acquired. If either or both of the two subsequences --/.sp/ and /.pr/-- were correctly produced 80% of the time, then they were also considered acquired. The hypothesis was that more marked onsets would not reach the criterion level before less marked onsets, a hypothesis that could have been falsified if the triliteral onset had reached the criterion level and both of the biliteral subsequences had not. Eckman examined three triliteral onsets across 11 participants and four tasks and found one falsification; in one case, a triliteral onset was present at the criterion level, but both biliteral subsequences were absent. Given only one falsification in approximately 130 cases, this study provided evidence that less marked onsets are acquired before more marked onsets.

In a longitudinal case study, Abrahamsson (1999) tracked the production of /.sC(C)/ onsets in Swedish by a native Spanish speaker. Over a ten month period, the participant modified .77 of the triliteral onsets and .59 of the biliteral onsets, a statistically significant difference.

1.2. Markedness by Sonority Sequencing

Another set of studies held length constant, examining just biliteral onsets that were in a markedness relationship according to their sonority profile. Carlisle (1988) examined the frequency of epenthesis before OL and ON onsets. Even though both onsets abide by the Core Syllabification Principle (CSP) of continuously rising sonority through the nucleus (Clements 1990), OL onsets are still less marked because they have a lower dispersion value, (0.56), than do ON onsets (1.17) according to the Feature Dispersion Principle (Clements 1990) . To test the possible influence of this markedness relationship, Carlisle examined the frequency of prothesis before the onsets /.sl/, /.sm/, and /.sn/, the hypothesis being that prothesis would occur less frequently before the OL onset than the ON onsets. The mean proportions of prothesis before the three onsets were .29 for /.sl/, .38 for /.sm/, and .33 for /.sn/, a significant difference among the three means. Pairwise comparisons revealed that the mean frequency of prothesis before /.sl/ was significantly less than those before /.sm/ and /.sn/ as hypothesized. In addition, /.sm/ was also more frequently modified than was /.sn/, although the two onsets are not in any known implicational relationship. However, a possible explanation may be found in Clements's Sequential Markedness Principle (1990: 313) stated below:

(1) For any two segments A and B and any given context X Y, if A is simpler than B, then XAY is simpler than XBY.

Given that anterior coronals are less marked than are labials, the sequence /.sn/ is less marked than /.sm/ and should therefore be modified less frequently.

In a later study, Carlisle (1991) examined the production of /.sl/ and /.st/ onsets by 11 native Spanish-speaking adults; the two onsets differ in that /.sl/ conforms to the CSP, and /.st/ does not, making the former less marked than the latter. The frequency of prothesis was .36 before /.st/ and .25 before /.sl/, a significant difference. So again the less marked onset was modified less frequently than was the more marked onset.

Another study finding that complex onsets with a lower dispersion value are modified less frequently than those with a higher one was conducted by Eckman and Iverson (1993). Eckman and Iverson investigated the production of four onsets ranked in terms of their dispersion value and the Sequential Markedness Principle as displayed in (1). The least marked onset was a voiceless stop+liquid with a dispersion value of 0.56 and the most marked was a voiceless stop+glide with a dispersion value of 1.17. The two intermediate onsets were voiced stop+liquid and voiceless fricative+liquid. Eckman and Iverson gathered data for their study by interviewing 11 participants, three speakers of Cantonese and four speakers each of Japanese and Korean.

To test their predicted markedness ranking, the researchers measured their participants’ production against a criterion of 80% correct production, hypothesizing that more marked onsets would reach the criterion level only if the corresponding less marked onsets also reached the criterion level. The study contained 55 potential tests of the hypothesis, but five of the tests could not be conducted because some participants did not produce the minimum number of tokens for one of the target onsets. Of the 50 remaining tests 46 supported the general hypothesis that a more marked onset would not reach the criterion threshold unless a corresponding less marked onset had also reached the criterion level.

1.3. Longitudinal studies and markedness

Only one longitudinal study has examined the effect of markedness on the production of onsets (Carlisle 1997, 1998, 2002). The study examined the production to two pairs of /.sC/ and /.sCC/ onsets: /.sk/ and /.skr/ and /.sp/ and /.spr/, the first member of each pair being less marked than the second as previously discussed. All participants were native Spanish speakers and were defined as intermediate at Time 1 according to their overall rate of correct production, between 21% and 79%. At Time I the data was analyzed in two ways. First, the frequency of prothesis was compared across the biliteral and triliteral pairs. Second, acquisition was examined by examining whether the less marked member of a pair would reach the criterion level of 80% correct production before the more marked member of the pair would.

Results at Time 1 revealed that the 11 participants modified the triliteral onset of both pairs significantly more frequently than they modified the biliteral onset. In addition, Participants 3 and 6 produced /.sk/ but not /.skr/ at the criterion level, and Participants 2, 3, 6 and 10 produced /.sp/ at the criterion but not /.spr/. These results agree with those of previous studies in that more marked structures are modified more frequently than are less marked structures and that more marked structures normally do not reach a criterion level of acquisition before less marked structures do.

Time 2 data from the remaining 10 participants was gathered ten months after Time 1. At Time 2, even though nearly all participants correctly produced the less marked onsets more frequently than they produced the more marked onsets, an increase in correct production was not clearly evident from Time 1 to Time 2. At time 2, Participant 6 produced all four onsets at the criterion level, and participant 2, produced /.sp/ at the criterion level but not /.spr/. In contrast, Participants 3 and 10 who had produced onsets at the criterion level at Time 1 actually displayed decreased frequencies of correct production at Time 2, deceases that were so large the onsets no longer were at the criterion level.

Time 3 data from the four remaining participants was gathered nearly three and half years after that of Time 2. One participant was producing all four onsets with complete accuracy; two had increased frequency of correct production, but none of the onsets had reached the criterion level, and the fourth participant had a lower frequency of correct production even though data gathering was years apart. Finally, the last three participants were still producing the less marked onsets with a greater frequency of correct production than they were the more marked onsets.

These findings clearly indicate two points. First, at all three times of data gathering, the participants as a group produced less marked onsets with greater frequency of correct production than more marked onsets, revealing the very strong influence that markedness has on production. Second, L2 learners do not uniformly improve. In fact, frequencies of correct production can reach a criterion level at one time and then fall below it the next.

2. PURPOSE

The current study has several purposes. The first is to determine whether the influence of markedness remains as strong at Time 2 as at Time 1. The second is to calculate whether the frequency of correct production increases over time. The third is to investigate whether participants who produce onsets at the criterion level at Time 1 still do so at Time 2. The last purpose is to analyze possible causes for changes in correct production by examining crucial background variables gathered from each participant about their use of English. However, given limitations with space, those results will not be presented here.

3. METHODOLOGY

3.1. Participants

All the participants in the study were adult native Spanish-speakers who were enrolled in intermediate ESL courses at Bakersfield College at Time 1. The students were placed in the intermediate classes based on their scores on the Secondary Level English Proficiency Test (SLEP), which consists only of listening and reading comprehension sections; students were never placed according to their proficiency in pronunciation.

Because the students were not placed because of their pronunciation, at Time 1 the participants' overall frequency of correct production of the onsets had to fall within a certain range. All of the participants had to produce at least 21% of the onsets correctly, but no more than 79%. This range of production to identify intermediate learners can be defended because previous research has used 80% correct production of any particular structure as the criterion level for acquisition (Andersen 1978; Carlisle 1997; Eckman 1991; Eckman and Iverson, 1993). For more background information on the participants, see Carlisle (2006).

3.2. Data gathering instruments

The data gathering instrument used at both Time 1 and Time II consisted of 375 randomly ordered sentences, 125 each for /.sl/, /.sn/ and /.st/. The environments before the three onsets were strictly controlled; 25 environments appeared exactly five times before each onset. Each sheet contained 25 sentences for the participants to read. At Time II, participants read the sentences in different orders than they had at Time 1. The participants also answered a questionnaire with six background prompts about their use of English.

3.3. Procedure

The 12 participants at Time II were individually taped exactly one year after they were originally taped and in the same language laboratory at Bakersfield College. Participants were taped with a Sony TC-D5PROII recorder with Sony ECM-530 microphone.

3.4. Transcription and reliability

The two reseachers independently transcribed the tapes of the 12 participants, specifically noting the quality of the preceding environment, the presence of a prothetic vowel, and the quality of the onset. Interrater reliability coeffecients ranged between .87 to .97, with the average being .92. Another faculty member with training in phonetics and experience in transcribing independently resolved the differences between the first two transcribers. Participants in the study either skipped a sentence or misread the word containing the target onset 118 times. These items were removed from the statistical analysis.

3.5. Analysis

A two-way ANOVA with repeated measures was calculated with onset and time as the independent variables. This analysis enabled the researchers to determine whether the markedness relationships among the three onsets were still influencing the frequency of correct production at Time 2. The analysis also enabled the researchers to determine whether the participants were producing a significantly greater percentage of the onsets correctly at Time 2 than at Time 1. We also examined acquisition against a criterion level of 80% correct production.

Table 1: The percentage of correct production at Time 1 and Time 2.

	Time I	Time 2
.sl	61.3	62.9
.sn	54	58.1
.st	42.7	49.2

4. RESULTS AND DISCUSSION

Table 1 displays the frequency of correct production for the three onsets at Time 2. As revealed in the table, the frequency of correct production is in the expected order according to the markedness relationships. Participants produced 62.9% of the /.sl/, onsets correctly; 58.1% of the /.sn/ onsets correctly; and 49.2% of the /.st/, onsets correctly. These mean differences produced a significant result: $\underline{F}(2,22) = 13.070$, $\underline{p} <.0005$. These results are in accordance with those found at Time 1 (Carlisle 2006). At both times markedness had a significant influence on the correct production of the onsets. The influence of markedness did not diminish over the time of the study, which should not be surprising given that the overall difference in correct production was small from Time 1 to Time 2 as discussed next.

The main effect for time was not significant; at Time 1, the participants correctly produced 52.7% of the total onsets, and at Time 2, they correctly produced 56.7% of them, a difference of only 4% in the year between data gathering. However, even though the overall increase in correct production was 4%, individual participants differed greatly in their performance from Time 1 to Time 2 as revealed in Table 2. At Time 1, 11 of the 12 participants displayed the expected linear pattern of lower correct production as markedness increased. Only Participant 11 violated this pattern, having a higher percentage of correct production for /.sn/ than for /.sl/. At Time 2, eight of 12 participants adhered to the expected linear pattern of lower frequencies of correct production with the increase of markedness. This consistency in the pattern of individual production reaffirms the statistical finding that less marked onsets are correctly produced significantly more frequently than are more marked onsets.

Participants were not uniform in their differences in means between Time 1 and Time 2. Eight participants experienced gains; Participants 6 and 15 displayed little or no difference; and Participants 7 and 16 displayed dramatic deceases in correct production over time, 33.7% and 16.9% respectively.

Table 2: The percentage of correct production of the three onsets by all participants at Time 1 and Time 2.

Participant	.sl at T1	.sn at T1	.st at T1	.sl at T2	.sn at T2	.st at T2
1	71.3	70.7	67.2	76.3	78.0	75.4
2	35.7	25.0	14.0	63.6	62.9	42.1
4	50.9	31.9	25.0	54.9	52.0	42.3
6	40.7	25.0	19.3	38.8	30.9	26.9
7	**80.0**	79.2	63.6	45.5	44.4	32.0
8	72.0	68.1	55.2	**84.7**	78.2	76.6
10	**85.8**	**84.6**	61.5	**86.2**	**85.1**	72.1
11	76.9	**86.0**	71.8	**91.0**	**93.6**	**87.1**
12	47.3	42.6	35.0	49.1	56.7	45.4
13	37.1	26.1	8.9	42.6	28.2	12.8
15	54.7	31.9	15.3	56.9	27.2	17.9
16	**83.0**	76.6	75.5	65.0	59.3	60.2

A significant interaction effect obtained between the main effects of onset and time: $\underline{F}(2, 22) = 21.881$, $\underline{p} = .0005$, indicating that the pattern of correct production was different from Time 1 to Time 2. Table 1 indicates that even though Time 2 means are higher than those for Time 1, the differences in correct production were not uniform across the three onsets: 1.6% for /.sl/, 4.1% for /.sn/, and 6.5% for /.st/. This pattern of improvement is related to the markedness relationships, the more marked the onset the greater the increase in correct production. If this trend were continue over time, reduced differences among the frequencies of correct production would obtain, a finding that would be in accordance with Major's Ontogeny Phylogeny Model (2001), which in part postulates that the influence of markedness is weak in the early stages in L2 acquisition, increases in the intermediates stages, and declines in the later stages.

The researchers also examined the means of individual participants against an acquisition level of 80% correct production (such means are bolded in Table 2). At Time 1, Participants 7 and 16 produced /.sl/ at the criterion level but neither of the more marked onsets. Participant 10 produced both /.sl/ and /.sn/ at the criterion level but not /.st/. These findings are in accordance with previous findings that more marked onsets do not normally reach the criterion level before less marked onsets. However, at Time 1 Participant 11 violated the expectation in that /.sn/ reached the criterion level before /.sl/. At Time 2 Participant 8 produced /.sl/ at the criterion level; Participant 10 produced both /.sl/ and /.sn/, and Participant 11 produced all three onsets at the criterion level. All of these findings conform to the expectation that more marked onsets will not reach a criterion level before less marked onsets. However, an onset reaching the criterion level at Time 1 does no guarantee that it will remain at the criterion level at Time 2. Both Participants 7 and 16 produced /.sl/ at the criterion level at Time 1 but not at Time 2. In fact, both participants displayed a uniform decrease in the correct production of the onsets.

5. SUMMARY

The first purpose of the study was to examine whether the influence of markedness remained as strong at Time 2 as at Time 1. The statistical results of the study revealed that it was; at both times the results were significant, and the linear order of correct production was in accordance with the markedness relationships—the less marked the onset the greater the frequency of correct production. The second purpose was to determine whether the frequency of correct production increased over time. The statistical results revealed that although the mean was 4% higher at Time 2 than at Time 1, the results were not significant. Nevertheless, eight of the 12 participants had the expected order of correct production. The finding that the increase of correct production was greater for more marked onsets supports Major's claims that the influence of markedness declines in later stages of acquisition. Finally, the third purpose was to investigate whether participants who produced onsets at the criterion level at Time 1 still did so at Time 2. Two participants in this study produced /.sl/ onsets that reached the criterion level at Time 1 but not at Time 2. Some L2 learners actually display lower frequencies of correct production over time as has already been revealed in a previous longitudinal study (Carlisle 1998, 2002).

6. REFERENCES

Abrahamsson, N. 1999. Vowel epenthesis of /sC(C)/ onsets in Spanish/Swedish interphonology: A longitudinal study. *Language Learning 49*. 473-508.

Andersen, R. 1978. An implicational model for second language research. *Language Learning 28*. 221-281.

Anderson, J. 1987. The markedness differential hypothesis and syllable structure difficulty. In: Ioup, G., Weinberger, S. (eds.), *Interlanguage phonology: The acquisition of a second language sound system*. Cambridge, MA: Newbury House. 279-291.

Carlisle, R. 1988. The effect of markedness on epenthesis in Spanish/English interlanguage phonology. *Issues and Developments in English and Applied Linguistics 3*. 15-23.

Carlisle, R. 1991. The influence of syllable structure universals on the variability of interlanguage phonology. In: Volpe, A.D. (ed.), *The seventeenth LACUS forum 1990*. Lake Bluff, IL: Linguistic Association of Canada and the United States. 135-145.

Carlisle, R. 1997. The modification of onsets in a markedness relationship: Testing the interlanguage structural conformity hypothesis. *Language Learning 47*. 327-361.

Carlisle, R. 1998. The acquisition of onsets in a markedness relationship: A longitudinal study. *Studies in Second Language Acquisition 20*. 245-260.

Carlisle, R. 2002. The acquisition of two and three member onsets: Time III of a longitudinal study. In: James, A., Leather, J. (eds.), *New Sounds 2000*. Klagenfurt, Austria: University of Klagenfurt Press. 42-47.

Carlisle, R. 2006. The sonority cycle and the acquisition of complex onsets. In: Baptista, B., Watkins, A. (eds.), *English with a latin beat: Studies in Portuguese/Spanish-English interphonology.* Amsterdam: John Benjamins. 105-137.

Clements, G. 1990. The role of the sonority cycle in core syllabification. In: Kingston, J., Beckman, M. (eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech*. Cambridge: Cambridge University Press. 283-333.

Eckman, F. 1991. The structural conformity hypothesis and the acquisition of consonant clusters in the interlanguage of ESL learners. *Studies in Second Language Acquisition 13*. 23-41.

Eckman, F., Iverson, G. 1993. Sonority and markedness among onset clusters in the interlanguage of ESL learners. *Second Language Research 9*. 234-252.

Greenberg, J. 1978. Some generalizations concerning initial and final consonant clusters. In: Greenberg, J., Ferguson, C., Moravcsik, E. (eds.), *Universals of human language (Vol. 2)*. Stanford, CA: Stanford University Press. 243-279.

Major, R. 2001. *Foreign Accent: The ontogeny and phylogeny of second language phonology*. Mahwah, NJ: Lawrence Erlbaum Associates.

Morelli, F. 2003. The relative harmony of /s+stop/ onsets: Obstruent clusters and the sonority sequencing principle. In: Fery, C., van de Vijver, R. (eds.), *The syllable in Optimality Theory*. Cambridge: Cambridge University Press. 356-371.

Cross-linguistic perception of diphthongs by Spanish/Catalan speakers and English speakers

Juli Cebrian

Universitat Autònoma de Barcelona
juli.cebrian@uab.es

ABSTRACT

Catalan speakers were tested on the perceived similarity between L1 vowels and diphthongs and English vowels in order to assess if English diphthongs are perceived as closer to L1 monophthongs or to L1 diphthongs. Results from perceptual assimilation tasks in which listeners identified target stimuli as native categories and provided goodness of fit ratings indicated that for instance English /aʊ aɪ/ were perceived as closer to Catalan /au ai/ than to Catalan /a/. The same was true of English /eɪ/ and /oʊ/ which were more consistently assimilated to Catalan /ei/ and /ou/ than to Catalan /e/ and /o/. Results from rated dissimilarity tasks involving English and Catalan vowel pairs showed that listeners consistently rated pairs of L1-L2 diphthongs such as English /oʊ/-Catalan /ou/ as more similar than pairs of acoustically close L1 monophthongs-L2 diphthongs such as English /oʊ/-Catalan /o/. Furthermore, results on the perception of Catalan vowels and diphthongs by English native speakers indicated that Catalan diphthongs were heard as closer to English diphthongs than the corresponding Catalan monophthongs were. These results lend support to the view that sequences like Spanish and Catalan diphthongs should not be excluded from cross-language similarity and L2 perception and production studies.

Keywords: crosslinguistic similarity, perception, English, Catalan, diphthongs.

1. INTRODUCTION

Cross-language comparisons are often the starting point in L2 speech research since a good understanding of the learning process requires a detailed examination of the characteristics of the native or first language (L1) and the target or second language (L2), as well as an awareness of the differences and similarities between the two languages. Further, according to current models of L2 speech such as the Perceptual Assimilation Model (PAM, Best, 1995; Best & Tyler, 2007) or the Speech Learning Model (SLM, Flege, 1995, 2003), the likelihood of accurate second language category formation is directly related to the degree of similarity between the existing L1 categories and the target L2 sounds. Therefore, an accurate description of cross-language similarity should include all the sounds from the L1 and the L2 that are relevant to the feature under study.

In studies involving Spanish and English vowels, comparisons typically involve the whole set or a subset of the English vowel system and the five Spanish vowel monophthongs. For instance, Imai et al. (2002) investigated the phonetic differences between Spanish /i e a o u/ and the acoustically closest English counterparts /i eɪ ɑ oʊ u/, while Flege et al. (1994) and Fox et al. (1995) rated the degree of dissimilarity between Spanish /i e a/ and English /i ɪ eɪ ɛ æ ʌ ɑ/. Similarly, Iverson & Evans (2007) examined the assimilation of a large number of English vowels and diphthongs to the five Spanish vowel monophthongs /a e i o u/. While these studies provided informative depictions of the relationship between the vowel categories studied, it could be argued that they were incomplete in that they ignored the existence of other vowel sequences that may play a role in Spanish speakers' perception of the English categories examined, namely Spanish diphthongs like /ei/, /ai/ or /au/ as in *ley* 'law', *hay* 'there is' and *auto* 'car', respectively. For instance, Iverson and Evans found that English /aɪ/ and /aʊ/ obtained very low identification scores as the 'closest' Spanish vowel, namely /a/. However, inclusion of Spanish /ai/ in that study could perhaps have rendered different results. Still, it is a matter of debate whether sequences like Spanish /ei/ are comparable to

English /eɪ/ on account of their greater formant movement in comparison to English diphthongs (e.g., Hillenbrand et al., 1995).

The purpose of this study was to investigate whether diphthongs in languages like Spanish and Catalan should be included in crosslinguistic comparison studies. Catalan has a number of diphthongs, comparable in amount of formant movement to those of Spanish, made up of practically all the combinations the seven monophthongs (/i e ɛ a ɔ o u/) and the high glides (e.g., Recasens, 1993). Examples are /ai/ in *xai* ‘lamb’, /au/ in *dau* ‘dice’, /ei/ in *rei* ‘king’, /eu/ in *déu* ‘god’, /ɛi/ in *feina* ‘work’, /ɛu/ in *deu* ‘ten’, /iu/ in *viu* ‘alive’, /ɔi/ in *noi* ‘boy’, /ɔu/ in *nou* ‘new’ and /ou/ in *pou* ‘well’. The perceived similarity between English vowels and diphthongs and Catalan vowels and diphthongs was explored by examining the results of a series of perceptual tests involving both native speakers of Catalan and native speakers of English.

For the sake of clarity, throughout this paper the glides in Catalan diphthongs will be transcribed as /i/ and /u/, as in /ai/ and /au/, while English diphthongs will end in /ɪ/ and /ʊ/ as in /aɪ/ and /aʊ/, so as to distinguish the two languages better and to maintain the transcription conventions of each language.

2. METHODOLOGY

The data examined in this paper is a subset of the data obtained in previous studies (Cebrian, 2006, 2009; Cebrian et al., 2010). In those studies, Catalan native speakers and English native speakers were tested on the perceived (dis)similarity of Catalan and English vowels by means of perceptual assimilation tasks and rated dissimilarity tasks.

2.1. Perceptual assimilation tasks (PAT)

In these tasks participants were presented with native and non-native vowel stimuli and had to identify each stimulus in terms of one of the L1 choices presented on a computer screen. Subsequently, participants rated the vowel according to how closely it approximated the chosen vowel-response. The goodness of fit rating was indicated on a 7-point scale (1=poor exemplar, 7=good exemplar). Stimuli in Cebrian (2005, 2006) were four Catalan and four English high and mid vowels and diphthongs including English /eɪ/ and Catalan /ei/. The stimuli were isolated vowels edited out of /hVb/ words. Cebrian (2009) tested a larger number of vowels, presented in /bVs/ word frames, which included Catalan /ei, ɛi, ou, ɔu, ai, au/ and English /eɪ, oʊ, ai, au/.

The participants in Cebrian (2005, 2006) were 20 native speakers of Catalan with no knowledge of English, 12 Catalan learners of English and 20 monolingual speakers of Canadian English. Cebrian (2009) tested 10 Catalan native speakers, 20 Catalan learners of English and 25 Canadian English speakers. Although the results for the two Catalan groups within each study did not differ significantly, only the results for the Catalan speakers with no knowledge of English are presented in this paper so as to compare two groups of naive listeners with none or little exposure to the non-native vowels.

2.2. Rated dissimilarity task (RDT)

The rated dissimilarity task required listeners to rate the degree of dissimilarity between the two members of a pair of /bVt/ words on a continuous 7-point scale (1=same, 7=different) (see Cebrian et al. this volume, for details). 47 Catalan learners of English participated in this experiment. The crucial pairs examined in this paper are the L1-L1 and L1-L2 combinations of the vowels and diphthongs included in that study, namely Catalan /o/-/ou/, Catalan /e/-/ei/, Catalan /o/-English /oʊ/, Catalan /ou/-English /oʊ/, Catalan /e/-English /eɪ/ and Catalan /ei/-English /eɪ/.

3. RESULTS

The results of the PAT in Cebrian (2009), which included a larger number of diphthongs, are presented in Table 1 (English vowel stimuli) and Table 2 (Catalan vowel stimuli). Regarding the English vowels, English /eɪ/ was consistently identified as the Catalan diphthong /ei/ (84%) with a mean goodness rating of 4.6 out of 7 in Cebrian (2006), while it was perceived as the L1 monophthong /e/ only 13% of the time. The tendency to assimilate to the Catalan diphthong was replicated by Cebrian (2009) although the assimilation rates were lower (see Table 1). The difference between the two studies may be due to the difference in the consonantal context of the vowel stimuli, presented in isolation in one study and in /bVs/ words in the other. Although English /eɪ/ was not always judged to be a very close match for Catalan /ei/, the fact that the assimilation scores as the Catalan monophthong /e/ were much lower indicates that the high offglide is a crucial cue to the identification of this vowel for Catalan listeners. The remaining diphthongs analyzed by Cebrian (2009) obtained very high scores of assimilation to Catalan diphthongs and comparatively high goodness ratings, as shown in Table 1. Therefore, these results show that L1 diphthongs like /ei/ and /ou/, and not simply pure vowels like /e/ and /o/, should be considered when assessing crosslinguistic mapping of sounds like English /eɪ/ and /oʊ/ (cf. Imai et al., 2002).

Table 1: Perceptual assimilation rates for English diphthongs and goodness of fit ratings in parentheses (1 = bad exemplar, 7 = good exemplar).

English vowels	/eɪ/	/oʊ/	/aɪ/	/aʊ/
Catalan speakers (L2 to L1 assimilation)	61 as /ei/ (3.9) 17 as /e/ (2.3) 11 as /i/ (2.0) 11 as /ɛ/ (2.0)	97 as /ou/ (4.8)	100 as /ai/ (4.3)	100 as /au/ (5.0)
English speakers (L1 identification)	98 as /eɪ/ (6.4)	100 as /oʊ/ (6.5)	100 as /aɪ/ (6.7)	99 as /aʊ/ (6.7)

Table 2: Perceptual assimilation rates for Catalan diphthongs and goodness of fit ratings in parentheses (1 = bad exemplar, 7 = good exemplar).

Catalan vowels	/ei/	/ɛi/	/ou/	/ɔu/	/ai/	/au/
English speakers (L2 to L1 assim.)	95 as /eɪ/ (5.0) 5 as /ɛ/ (3.0)	87 as /eɪ/ (4.1) 8 as /ai/ (5.0)	99 as /oʊ/ (5.1)	49 as /aʊ/ (3.1) 47 as /oʊ/ (4.5)	99 as /aɪ/ (4.3)	99 as /aʊ/ (3.7)
Catalan speakers (L1 ident.)	99 as /ei/ (5.8)	99 as /ɛi/ (6.6)	99 as /ou/ (6.3)	98 as /ɔu/ (6.5)	99 as /ai/ (6.1)	99 as /au/ (6.4)

Regarding the assimilation of Catalan diphthongs to English vowels, the non-native sounds were consistently assimilated to English diphthongs rather than to monophthongs. Given the greater formant movement of the Catalan diphthongs, this is not surprising. Catalan /ɔu/ yielded a different result, being assimilated to two different L1 vowels, indicating that the starting point of this vowel does not have a clear match among the English diphthongs.

Table 3: Mean dissimilarity ratings for each type of vowel pair (1 = same, 7 = different).

Type of vowel pair	Vowels	Dissimilarity rating (SD)
L1 monophthong – L1 diphthong	Catalan /e/-/ei/	4.7 (0.9)
L1 monophthong – L1 diphthong	Catalan /o/-/ou/	4.7 (1.1)
L1 monophthong – L2 diphthong	Catalan /e/-English /eɪ/	5.8 (0.8)
L1 monophthong – L2 diphthong	Catalan /o/-English /oʊ/	5.0 (1.1)
L1 diphthong – L2 diphthong	Catalan /ei/-English /eɪ/	3.4 (0.9)
L1 diphthong – L2 diphthong	Catalan /ou/-English /oʊ/	3.2 (1.0)

The results of the RDT strongly supported the outcome of the PATs, as shown in Table 3. The Catalan listeners perceived the pairs made up of a Catalan monophthong and an English diphthong, such as Catalan /ei/-English /eɪ/, as more dissimilar than pairs consisting of a Catalan diphthong and an English diphthong, such as Catalan /e/ and English /eɪ/. These results lend support to the view that sequences like Spanish and Catalan diphthongs should not be excluded from cross-language similarity and L2 perception and production studies.

4. CONCLUSIONS

Despite the tendency to exclude diphthongs from cross-linguistic vowel comparison involving languages like Spanish and English, the current study provides evidence from a language close to Spanish, Catalan, that L1 diphthongs play an important role in the categorization of English vowels by Catalan learners of English. English diphthongs were found to be assimilated to Catalan diphthongs much more frequently than to Catalan monophthongs, including English diphthongs of comparatively more limited formant movement like /eɪ/ and /oʊ/. The results of a rated dissimilarity task showed that Catalan listeners perceive pairs of a Catalan monophthong like /e/ and an English diphthong like /eɪ/ as more dissimilar than pairs made up of Catalan diphthong like /ei/ and an English diphthong /eɪ/. This outcome clearly confirmed the preference observed in the perceptual assimilation task. Given the similar nature of the diphthongs in Spanish and Catalan, we can conclude that a complete examination of crosslinguistic similarity between Spanish and other languages should not exclude the diphthongs. For instance, Iverson and Evans' (2007) finding that Spanish learners of English preferred and produced /aʊ/ tokens that were closer to English /aʊ/ than to the supposedly acoustically closest Spanish vowel /a/ could be reinterpreted if the Spanish diphthong /au/ was considered in the analysis, which might show that Spanish speakers categorize English /aʊ/ in terms of Spanish /au/.

Work supported by Grant 2009PBR00008 from the Catalan Government.

5. REFERENCES

Best, C. 1995. A direct realist view of cross-language speech perception. In: Strange, W. (ed.), *Speech perception and linguistic experience: Issues in cross-language research.* Timonium, MD: York Press. 171-204.

Best, C., M. Tyler. 2007. Non-native and second language speech perception. In: Bohn, O.-S., Munro, M. J. (eds.), *Language experience in second language speech learning.* Amsterdam: John Benjamins. 15-34.

Cebrian, J. 2005. Measuring phonetic distance between Catalan and English vowels: A bidirectional cross-language mapping study. Paper presented at the 2nd Phonetics and Phonology in Iberia conference. Universitat Autonoma de Barcelona, Spain.

Cebrian, J. 2006. Experience and the use of duration in the categorization of L2 vowels. *Journal of Phonetics* 34. 372-387.

Cebrian, J. 2009. Effects of native language and amount of experience on crosslinguistic perception. *Journal of the Acoustical Society of America*, Vol. 125, No. 4, Pt. 2, April, p. 2775.

Cebrian, J., Mora, J. C., Aliaga-Garcia, C. 2010 (this volume). Assessing crosslinguistic similarity by means of rated discrimination and perceptual assimilation tasks. *Proceedings of the New Sounds 2010 Symposium on the Acquisition of L2 Speech.* Poznan: Adam Mickiewicz University.

Flege, J. E. 1995. Second language speech learning: Theory, findings and problems. In: Strange, W. (ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press. 233-277.

Flege, J. E. 2003. Assessing constraints on second-language segmental production and perception. A N.O. Schiller i A. S. Myere (eds.), *Phonetics and phonology in language comprehension and production*. New York: Mouton de Gruyter. 319-355.

Flege, J. E., Munro, M. J., Fox, R. A. 1994. Auditory and categorical effects on cross language vowel perception. *Journal of the Acoustical Society of America* 95. 3623-3641.

Fox, R., Flege, J.E., & Munro, M. (1995). The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis. *Journal of the Acoustical Society of America*, 97. 2540-2551.

Hillenbrand, J., Getty, L.A., Clark, M.J. & Wheeler, K. 1995. Acoustic characteristics of American English vowels, *Journal of the Acoustical Society of America*, **97**. 3099-3111.

Imai, S., Flege, J. E., Wayland, R. (2002). Perception of cross-language vowel differences: A longitudinal study of native Spanish learners of English. *Journal of the Acoustical Society of America,* 111. 2364.

Iverson, P., Evans, B.G. (2007). Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration. *Journal of the Acoustical Society of America* 122(5). 2842-2854.

Recasens, D. 1993. *Fonètica i fonologia*. Barcelona: Enciclopèdia Catalana.

Assessing crosslinguistic similarity by means of rated discrimination and perceptual assimilation tasks

Juli Cebrian[1], Joan C. Mora[2] and Cristina Aliaga-Garcia[2]

[1]Universitat Autònoma de Barcelona, [2]Universitat de Barcelona
juli.cebrian@uab.es, mora@ub.edu, cristinaaliaga@ub.edu

ABSTRACT

This study compares the results of a perceptual assimilation task to those obtained from a rated discrimination (or dissimilarity) task in the assessment of the degree of crosslinguistic similarity between the vowel systems of Catalan and English. Native speakers of Catalan were tested on native and non-native vowels presented in /bVt/ syllables. Stimuli in the rated discrimination task included same-category vowel pairs and adjacent-category vowel pairs involving two native (L1) vowels, two non-native (L2) vowels and one L1 and one L2 vowel. The results indicated that some non-native vowels were very consistently perceived as instances of an L1 category, with varying degrees of goodness of fit. In the discrimination task, L1 same-category pairs were rated by Catalan speakers as being more similar than L2 same-category pairs, but some L1-L1, L2-L2 and L1-L2 different-category pairs received comparable ratings. Further, some adjacent-category L1-L2 pairs obtained ratings within the range obtained for same-category pairs, such as Catalan /i/ and English /i/, suggesting that some L2 vowels are highly assimilated to L1 categories. The results of the two tasks complemented each other pointing to the advantage of combining more than one method of assessment to obtain more reliable measures of cross-linguistic perceptual similarity.

Keywords: Perceptual similarity, cross-linguistic perception, vowel discrimination, English, Catalan.

1. INTRODUCTION

The assessment of crosslinguistic perceptual similarity remains an important methodological issue in cross-language speech perception research. Current models of L2 speech learning such as the Speech Learning Model (SLM, Flege 1995) and the Perceptual Assimilation model (PAM, Best 1995; Best and Tyler 2007) make use of phonetic similarity as a key predictor of difficulties in the perception and production of L2 sounds and relate such difficulties to the development of accurate representations for L2 sound categories. For example, one recurrent finding of studies comparing L1 and L2 vowels through measures of spectral acoustic similarity in a F1-F2 space is that, with sufficient exposure, L2 vowels that are acoustically more distinct from the nearest L1 vowel are perceived (and produced) more accurately than L2 vowels that are more similar, but not identical, to an L1 vowel (Bohn and Flege 1990, 1992). This suggests that L2 sound category formation is enhanced by greater degrees of perceptual dissimilarity between L1 and L2 sounds and blocked when L2 sounds are perceived as identical or very similar exemplars of a L1 category (single-category perceptual assimilation, in PAM's terms).

Two crucial issues in this kind of research are the extent to which the degree of perceived auditory distance in a L2-L1 vowel pair reflects their acoustic distance in an F1-F2 vowel space, and whether the degree of crosslinguistic phonetic dissimilarity is best determined through acoustic or auditory (perceptual) measures. Flege et al. (1994) tested English monolinguals and native Spanish speakers of English on their perception of Spanish and English vowel categories by means of a perceptual dissimilarity rating task involving pairs of English vowels, pairs of Spanish vowels and pairs of Spanish and English vowels. Their results showed that vowel dissimilarity ratings varied as a function of the acoustic distance in an F1-F2 space, the greater the distance between the vowels the more dissimilar the vowels were perceived to be. The degree of vowel dissimilarity the listeners perceived was found to be independent from their L1 for vowel pairs that were non-adjacent in an F1-F2 vowel space, but not for adjacent vowel pairs. The native English listeners judged adjacent vowel pairs to be more dissimilar and were more successful at identifying adjacent vowel pairs in an oddity discrimination task (ODT) in terms of two categories than were the native Spanish listeners, but also among the Spanish participants greater dissimilarity ratings for a given pair of vowels went hand in hand with better performance in the corresponding ODT triad. This suggests that the differential classification of a pair of vowels increases their perceived dissimilarity. Another important finding of the study is that the degree of perceived dissimilarity did not vary significantly as a function of experience with English.

Three studies have previously explored the perceived similarity between Catalan and English vowels. Cebrian (2006) tested Catalan speakers' categorization of English vowels by means of a perceptual assimilation task (PAT) in which listeners identified isolated vowel stimuli in terms of L1 vowels and provided goodness of fit ratings. The study focused on (Canadian) English /i, ɪ, eɪ, ɛ/ and the acoustically closest Catalan counterparts /i, e, ei, ɛ/ and compared a group of Catalan speakers with no knowledge of English and a group of experienced L2 English learners. Relatively high degrees of assimilation (84-99%) were found between Catalan /i/, /ei/ and /ɛ/ and English /i/, /eɪ/ and /ɛ/, respectively, but English /ɪ/ obtained lower assimilation scores as Catalan /e/ (66%, goodness rating of 3.5/7). Experience was not found to have an effect. Cebrian (2009) extended the study to include a larger number of Catalan and Canadian English vowels and diphthongs presented in /bVs/ words. The results mostly replicated the findings in the earlier study and showed that English /æ/ and /oʊ/ were also consistently assimilated to L1 vowel categories. Experience was found to play a role in a study by Rallo (2005) that examined the assimilation of the seven Catalan vowels to American English vowels. Rallo also assessed vowel discrimination in an oddity task in which listeners were presented with AXX triads and had to select the odd item out or a "none" option. Stimuli were presented in /sVt/ words. Some learners were able to distinguish some English-English and some Catalan-English pairs, which was interpreted as evidence of L2 category formation. The relation between the two tasks was however difficult to interpret since some L2 vowels, such as English /æ/ and /ɛ/, that were poorly discriminated from Catalan /a/ and /ɛ/, respectively, did not receive consistent assimilation scores to the Catalan vowels in the PAT. The perceptual assimilation rates obtained by Rallo were notably lower than those in the studies by Cebrian. This may be due to task differences since Rallo included "non-Catalan" as a possible response alternative, and differences in the consonantal context of the stimuli.

Recent research (Strange 2007) has pointed out several methodological problems in the acoustic cross-linguistic comparison of vowels, such as the effect of consonantal context on non-native vowel production (Steinlen 2005) and speaker normalization, and advocates for direct methods of assessing crosslinguistic similarity through perceptual assimilation tasks in which listeners identify L2 sounds in terms of L1 categories and provide goodness of fit ratings. The present study explores the issue of cross-linguistic similarity further by comparing the results of a perceptual assimilation task (PAT) to those obtained from a rated discrimination (or dissimilarity) task (RDT) involving the vowel systems of Catalan and English. This research is motivated by the need to obtain reliable auditory measures of perceptual similarity between Catalan and English vowels in order to predict areas of difficulty in the perception and production of L2 vowels by Spanish-Catalan bilingual learners of English.

Auditory measures of the degree of perceptual (dis)similarity of a wide range of English and Catalan monophthongs and diphthongs, both within and between the two vowel systems of focus, were thus obtained. The main aim of the present study is to assess the reliability of two methods of assessing cross-linguistic similarity: a perceptual assimilation task in which listeners were asked to label vowel tokens in terms of native-language vowel categories and produced goodness of fit ratings, and a rated discrimination task in which listeners directly produced judgments of the degree of perceived dissimilarity between L1, L2 and L1-L2 vowel pairs.

2. METHOD

2.1. Participants

A group of adult Catalan-dominant Spanish-Catalan bilinguals (N=47) enrolled in a degree in English Studies in Barcelona participated in the study and were given course credit. Their mean age was 23 (range 20-39) and none reported any hearing problems. They were selected from a larger pool on the basis of their language background (Catalan-speaking parents and self-reported percent daily use) and their scores for Catalan vowels in the PAT test (see below). Because the perceptual asimilation task involved labelling of /bVt/ tokens according to Catalan vowel categories, language dominance was a major concern and only the data from Catalan-dominant bilinguals was analysed in the present study.

2.2. Stimuli

Monosyllabic words exemplifying 10 vowel monophthongs (/i ɪ ɛ æ ɑ ɜ ʌ ɒ ɔ u/) and 2 diphthongs (/eɪ/, /əʊ/) of British English, and 7 monophthongs (/i e ɛ a ɔ o u/) and 4 diphthongs (/ai/, /ei/, /au/, /ou/) of Eastern Catalan, were selected such that the target vowels all occurred between /b/ and /t/, in accordance with English and Catalan phonotactics and syllabification. The /bVt/ words were embedded in frame/carrier phrases of similar length, as well as similar syntactic structure and identical position of the target word across the two languages. The carrier phrases included common real words which preceded the target words in order to illustrate how target vowels should be pronounced, as the /b/-V-/t/ context created some non-words in both languages. The English and Catalan carrier phrases were "*Rima amb dit. Ara dic **bit** un cop.*" and "*It rhymes with hot. (I say bot). I say **bot** again.*", respectively.

English vowel stimuli were elicited from a group of three male native speakers of Southern British English (mean age: 35) living in London and neighbouring areas, and one male speaker of the same southern variety of English residing in Barcelona but having spent most of his life in the London area. Each English speaker recorded 6 repetitions of the phrases with the words *beat* /i/, *bit* /ɪ/, *bet* /ɛ/, *Bert* /ɜ/, *bat* /æ/, *but* /ʌ/, *Bart* /ɑ/, *bot* /ɒ/, *bought* /ɔ/, *boot* /u/, *bait* /eɪ/ and *boat* /əʊ/, with normal speed and falling intonation. A group of 10 Catalan-dominant bilinguals in Catalan and Spanish living in Barcelona and neighbouring areas were recorded (mean age: 30), selected on the grounds of a questionnaire and a short interview with two of the experimenters, native Catalan speakers. All of the Catalan speakers had spent most of their lives in greater Barcelona area and spoke the same variety of Catalan. The Catalan target words were *bat* /a/, *bet* /e/, *bèt* /ɛ/, *bit* /i/, *bót* /o/, *bòt* /ɔ/, *but* /u/, *bait* /ai/, *beit* /ei/, *baut* /au/, *bout* /ou/.

All recordings were made with a digital recorder (Marantz PMD660) and a unidirectional dynamic microphone (ShureSM58). The English recordings were made in a soundproof booth in the Phonetics Laboratory at UCL (London). The Catalan recordings were equally made in sound-attenuated conditions at EUPMT (Mataró, Barcelona). Both English and Catalan speakers read the sentence list repeated in six randomized blocks. The recordings were digitized at a 44kHz sampling rate and the stimuli from both languages were normalized for intensity (70dB) to minimize talker-related loudness differences. A selection of the best tokens based on auditory judgement and spectrographic analysis produced by three of the talkers was finally obtained in each of the languages for use as stimuli in the experiments, which included a total of 2 different (word) tokens for each of the English and Catalan /bVt/ syllables.

2.3. Perception tasks

Two tasks were used: a perceptual assimilation task (PAT) that involved labelling stimuli according to L1 vowel categories and rating them for goodness of fit, and a rated discrimination task (RDT) in which listeners rated the degree of (dis)similarity for every stimuli pair on a same-different continuum. Measures of degree of perceptual similarity were thus obtainable from both tasks through either the goodness of fit ratings given to single stimuli or the (dis)similarity ratings given to stimuli pairs. The main difference between both methods is that the PAT required subjects to compare the vowel in the /bVt/ stimuli to their phonetic representations in long term memory, whereas in the RDT subjects could compare the stimuli in each pair through auditory memory.

2.3.1. Perceptual assimilation task

The perceptual assimilation task (PAT) was based on a large randomized set of 66 English and 60 Catalan /bVt/ stimuli for which Catalan speakers/English learners gave a closed-set identification response consisting of 11 possible responses including the conventional orthography and one representative exemplar for each of the Catalan vowels: *dit (i)* (/i/) 'finger', *set (è)* (/ɛ/) 'seven', *fet (é)* (/e/) 'fact', *mut (u)* (/u/) 'mute', *pau (au)*

(/au/) ‘peace’, *rei (éi)* (/ei/) ‘king’, *pou (óu)* (/ou/) ‘well’, *xai (ai)* (/ai/) ‘lamb’, *got (ò)* (/ɔ/) ‘drinking glass’, *mot (ó)* (/o/) ‘word’, and *gat (a)* (/a/) ‘cat’. On each trial, subjects also a goodness of fit rating on a 7-point scale (1=poor exemplar; 7=very good exemplar). There were a total of 252 trials which included two tokens for each of the English and Catalan vowels as pronounced by three different male talkers in each language. Both English and Catalan stimuli were included for control purposes.

2.3.2. Rated discrimination task

The rated discrimination (or dissimilarity) task (RDT) required subjects to rate the degree of dissimilarity between the two members of 30 /bVt/ word pairs on a continuous 7-point scale (1=same, 7=different). Some pairs consisted of two different productions of the same word, some contained words with vowels that are ‘adjacent’ on a F1-F2 vowel space. Thus, the stimuli included two L1-L1 same-vowel pairs (/bet/-/bet/, /but/-/but/), five L1-L1 adjacent-vowel pairs (/bit/-/bet/, /bet/-/beit/, /bet/-/bɛt/, /bɔt/-/bot/, /bot/-/bout/), three L2-L2 same-vowel pairs (/bit/-/bit/, /bɪt/-/bɪt/, /bəʊt/-/bəʊt/), four L2-L2 adjacent-vowel pairs (*beat-bit, bat-but, bat-bart, but-Bart*), and 16 L1-L2 adjacent-vowel pairs (/bit/-*beat*, /bit/-*bit*, /bet/-*bit*, /bet/-*bait*, /bet/-*bet*, /beit/-*bait*, /bɛt/-*bet*, /bɛt/-*bat*, /bat/-*bat*, /bat/-*but*, /bat/-*Bart*, /bɔt/-*bot*, /bot/-*bought*, /bot/-*boat*, /bout/-*boat*, /but/-*boot*). There were six repetitions of the 30 different vowel pairs including talker variability (three talkers per language) and two orders (i.e., *beat-bit, bit-beat*) with an interstimulus interval of 1.2 sec. Participants were instructed to click on the same-different continuum on the screen to indicate their perceived degree of (dis)similarity for each pair.

2.4. Procedure

The subjects completed the task in a quiet computer room in groups. They listened to the stimuli individually over headphones and gave their responses with a mouse click by selecting buttons appearing on the computer screen. *Praat* (Boersma and Weenink 2007) software was used to run the experiments. Both perception tasks were preceded by a short training phase to familiarize participants with the range of possible identification and rating responses. On average it took subjects about 60 minutes to complete both tasks, with a pause between the tasks and possible short breaks within a task if necessary.

3. RESULTS

3.1. Perceptual assimilation task

The results of the PAT indicated that some English vowels were strongly assimilated to L1 categories: English /i/, /æ/ and /ɛ/ were identified as Catalan /i/, /a/ and /ɛ/ more than 90% of the time and obtained goodness of fit ratings of 4.6 out of 7 or higher. Other vowels obtained high assimilation scores (80-90%) but lower goodness ratings (3.5-3.8), like English /eɪ/, /u/, /əʊ/, /ɔ/, /ʌ/, /ɪ/ as Catalan /ei/, /u/, /ou/, /o/, /a/ and /i/, respectively. English /ɑ/ obtained comparable identification rates (78%) but lower goodness of fit scores as Catalan /a/ (2.5) and English /ɒ/ was heard as Catalan /ɔ/ 70% of the time but obtained moderate goodness ratings (4.1). The results coincide in part with the earlier studies discussed above, particularly with respect to the most consistently assimilated vowels. However, the fact that those studies examined North American English vowels while the current study focuses on Southern British English vowels, in addition to other methodological differences, renders the results less comparable. Therefore, the results of the PAT will be mostly discussed in relation to the outcome of the second task.

3.2. Rated discrimination task

Table 1 presents the mean dissimilarity ratings for each vowel pair type. Listeners found pairs of same-category L2 vowels as more difficult to discriminate than same-category L1 pairs. This difference proved significant in a paired samples t-test ($p< .001$). Among the adjacent category pairs, L1-L1 (i.e., Catalan-Catalan) pairs were heard as more dissimilar than L1-L2 (Catalan-English) pairs which in turn were less similar than L2-L2 (English-English) pairs. The results for the mean scores for the three adjacent vowel pair types were examined in a 3x1 repeated measures anova which yielded a significant main effect ($F(2,92)=41$, p<.001). Separate t-tests for each language combination yielded significant results for each of the three types (p<.001). These results illustrate the fact that the L1 categories are more robust and support the general trend in Flege et al. (1994) that adjacent vowel categories that correspond more closely to native categories are better discriminated.

Table 1: Mean dissimilarity ratings for each type of vowel pair (1 = same, 7 = different).

Type of pair	Language	Mean dissimilarity rating (SD)
same category	L1(Cat.)-L1(Cat.)	1.7 (0.5)
same category	L2(Eng.)-L2(Eng.)	2.0 (0.6)
adjacent category	L2(Eng.)-L2(Eng.)	3.4 (0.8)
adjacent category	L1(Cat.)-L2(Eng.)	3.7 (0.6)
adjacent category	L1(Cat.)-L1(Cat.)	4.2 (0.8)

Differences between individual vowel pairs were explored in separate paired samples t-tests. Despite the general results for category type, some L1-L1, L2-L2 and L1-L2 adjacent-category pairs received comparable results. The vowels that received the highest dissimilarity ratings were some of the Catalan-English pairs (/ɛ/-/ae/: 6.4, /e/-/eɪ/: 5.8, /o/-/əʊ/: 5.0, /e/-/ɛ/: 4.8), but were followed closely by some adjacent-category Catalan-Catalan pairs (/o/-/ou/: 4.7, /e/-/ei/: 4.7, /e/-/ɛ/: 4.6). The difference between the Catalan /e/-English /ɛ/ and the Catalan /o/-/ou/ pairs was not statistically significant. The least dissimilar among the adjacent vowel pairs were the Catalan-English pairs /ɛ/-/ɛ/ (2.2), /ɔ/-/ɒ/ (2.3), /i/-/i/ (2.3) and /a/-/æ/ (2.5). The Catalan /i/-English /i/ pair obtained discrimination scores that did not differ significantly from the scores for English /i/-/i/ (2.3), indicating that the L1 /i/ and L2 /i/ are perceived to be as close as instances of the same vowel category. The mean rating for Catalan /ɛ/-English /ɛ/ (2.2) was significantly higher than the one for Catalan /ɛ/-/ɛ/ (1.7) (t=(46)=4.5, p<001). Still, the discrimination ratings for Catalan /ɛ/-English /ɛ/, as well as for the Catalan-English pairs /ɔ/-/ɒ/ and /a/-/æ/ (2.3-2.5) (in addition to /i/-/i/) did not differ significantly from the rates for English-English /i/-/i/ (2.3). Therefore, listeners rated some adjacent L1-L2 pairs as being as similar as same-category vowel pairs suggesting that the non-native vowels were heard as instances of the native categories in these cases.

The results of the RDT parallel those obtained in the PAT. The English vowels that obtained the highest identification scores and goodness ratings as the closest Catalan vowels (i.e., Eng. /æ/ to Cat. /a/, Eng. /i/ to Cat. /i/, Eng. /ɛ/ to Cat. /ɛ/) correspond to the adjacent vowel pairs that obtained (dis)similarity ratings that did not differ from those for same-category vowel pairs (Cat.-Eng. /ɛ/-/ɛ/, /i/-/i/ and /a/-/æ/). L2 vowels that were fairly consistently identified as a single L1 vowel but obtained low goodness ratings were perceived to be in the mid range of the dissimilarity scale (3.2-3.6) when paired with the corresponding L1 vowels, as was the case of Catalan-English /ou/-/əʊ/, /a/-/ʌ/, /u/-/u/, /o/-/ɔ/, /ei/-/eɪ/, /i/-/ɪ/. Two English vowels obtained relatively close assimilation rates but different goodness of fit ratings as Catalan vowels: English /ɒ/ was identified as Catalan /ɔ/ 70% of the time with a mean goodness rating of 4.1, while English /ɑ/ obtained was perceived as Catalan /a/ 78% of the time with a mean rating of 2.5. As we have seen, the dissimilarity rating for Catalan-English /ɒ/-/ɔ/ (2.3) in the RDT was as low as that for English /i/-/i/. By contrast, the Catalan-English /a/-/ɑ/ pair was heard to be relatively dissimilar (4.5). These results point to the fact that both PAT measures, i.e., percentage identification as a native vowel and goodness of fit rating, contribute to the predictions of discrimination ability.

4. CONCLUSIONS

In this study we examined the perceptual similarity between English and Catalan vowels and evaluated two different methods of assessing crosslinguistic perceived similarity between L1 and L2 vowels. The two methods were a task in which individual stimuli were compared to L1 phonetic representations in long term memory (PAT) and a direct comparison of two stimuli presented consecutively and evaluated in auditory memory (RDT). Some English vowels were very consistently assimilated to their Catalan counterparts and obtained comparatively high goodness ratings. Adjacent-category L1-L2 pairs involving these highly assimilated vowels and the corresponding L1 vowels also obtained the lowest dissimilarity ratings in the RDT task. Together the results indicate that some target vowels are readily assimilated to native vowels and are perceived as instances of the native categories, consistent with some earlier studies (Cebrian 2006, 2009). Target vowels with lower assimilation rates and particularly lower goodness of fit ratings in the PAT were more readily judged to be dissimlar from native categories, and therefore it is more likely that learners, with enough experience, establish categories that are separate from the native categories. These findings lend support to proposed models of L2 speech learning whose predictions are based on the notion of crosslinguistic similarity (Best's (1995) PAM, Flege's (1995) SLM). However, a detailed evaluation of these models lies beyond the scope of the current paper. The results of both tasks complemented each other in most cases and point to the advantage of combining more than one method of assessment to obtain a more complete and reliable measure of crosslinguistic similarity. Although the current PAT results resemble those in the previous studies for some vowels, methodological differences involving task design, stimuli and target language variety make the cross-study comparisons less straightforward, as mentioned above.

The participants in the current study were Catalan learners of English with some theoretical familiarity with the sound structure of English and relatively low exposure to the target language outside the classroom. The current investigation forms part of a larger longitudinal study examining the effect of experience, understood as specialized L2 learning, on both the perceived similarity between L1 and L2 vowels and the ability to discriminate between L1 and L2 vowels. As such the current results will gain further relevance when compared to the outcome from additonal testing times as well as results from L1 speakers with no knowledge of the L2.

Work supported by Grant 2009PBR00008 from the Catalan Government.

5. REFERENCES

Best, C. 1995. A direct realist view of cross-language speech perception. In: Strange, W. (ed.), *Speech perception and linguistic experience: Issues in cross-language research.* Timonium, MD: York Press. 171-204.

Best, C., M. Tyler. 2007. Non-native and second language speech perception. In: Bohn, O.-S., Munro, M. J. (eds.), *Language experience in second language speech learning.* Amsterdam: John Benjamins. 15-34.

Boersma, P., and Weenink, D. 2007. Praat: Doing phonetics by computer (Version 4.6.05) [Computer program]. Retrieved June 2, 2007, from http://www.praat.org/.

Bohn, O.-S., Flege, J. E. 1990. Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics* 11. 303-328.

Bohn, O.-S., Flege, J. E. 1992. The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition* 14. 131-158.

Cebrian, J. 2006. Experience and the use of duration in the categorization of L2 vowels. *Journal of Phonetics* 34. 372-387.

Cebrian, J. 2009. Effects of native language and amount of experience on crosslinguistic perception. *Journal of the Acoustical Society of America*, Vol. 125, No. 4, Pt. 2, April, p. 2775.

Flege, J. E. 1995. Second language speech learning: Theory, findings and problems. In: Strange, W. (ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press. 233-277.

Flege, J. E., Munro, M. J., Fox, R. A. 1994. Auditory and categorical effects on cross language vowel perception. *Journal of the Acoustical Society of America* 95. 3623-3641.

Rallo, L. 2005. Acquisition of a second language vowel system: The case of the Catalan learners of English. Unpublished PhD thesis. Universitat de Barcelona.

Steinlen, A. K. 2005. The influence of consonantal context on native and non-native vowel production: A cross-language study. Tübingen: Gunter Narr.

Strange, W. 2007. Cross-language phonetic similarity of vowels. Theoretical and methodological issues. In: Bohn, O-S, Munro, M. J. (eds.), *Language experience in second language speech learning.* Amsterdam: John Benjamins. 35-55.

Investigating Catalan-Spanish bilingual EFL learners' over-reliance on duration: vowel cue weighting and phonological short-term memory

Eva Cerviño-Povedano, Joan C. Mora

University of Barcelona
ecervino@ub.edu, mora@ub.edu

ABSTRACT

The aim of this study is to explore the relationship between PSTM and Catalan-Spanish learners' cue weighting of English /iː/-/ɪ/. A sub-goal is to assess possible long-term memory effects on serial nonword recognition (SNWR) as a measure of PSTM. Two 8-step continua were created (*feet-fit*) in which vowel duration was manipulated and presented in a lexical decision task. The participants' (*N=85*) PSTM was measured using a SNWR task consisting of 144 monosyllabic items presented in sequences of increasing length (from 4 to 7 items). The task included three different sets of items: nonwords in the L1 (Catalan), L2 (English) and words and nonwords in a language unknown to the participants, Lx (Russian). Subjects' responses on the three sets were strongly correlated, suggesting that this task provides a language-independent measure of PSTM. Learners in the High PSTM group categorised /iː/-/ɪ/ in a more native-like manner than those in the Low PSTM group, who over-relied on duration. These results suggest that Catalan-Spanish EFL learners with higher PSTM may have an advantage over those with lower PSTM, in making use of more "difficult" or less readily accessible cues such as spectral information in the categorisation of English vowels.

Keywords: Phonological short-term memory (PTSM), serial nonword recognition (SNWR), cue weighting, vowel categorisation.

1. INTRODUCTION

The categorisation of L2 vowels might present difficulty for learners due to cross-linguistic differences in cue weighting. Native English speakers categorise vowel contrasts mainly by integrating spectral and durational information (Ylinen, Uther, Latvala, Vepsäläinen, Iverson, Akahane-Yamada and Näätänen 2009), Catalan-Spanish learners find it difficult to attend to both these cues in a native-like manner. Numerous studies suggest that whereas native English speakers rely mainly on spectral information in the categorisation of vowel contrasts, Catalan-Spanish learners rely on vowel duration. A much explored contrast in the literature is /iː/-/ɪ/ . This is because its perception and production poses a challenge to Catalan-Spanish learners since they assimilate these two categories to their single native category /i/ (Flege 1991). Extensive research on Catalan-Spanish L2 learners' categorisation of /iː/-/ɪ/ has shown that they rely on vowel duration to a greater extent than native English speakers, despite the fact that duration is not a contrastive feature in either Catalan or Spanish (Bohn 1995; Escudero and Boersma 2004; Cebrian 2006, 2007; Mora and Fullana 2007; Cerviño and Mora 2009). Bohn's (1995) Desensitization Hypothesis provides an explanation to this phenomenon in that it states that when spectral cues are insufficient to learners, duration is used regardless of whether this feature is contrastive or not in their L1.

Looking into individual differences amongst learners might bring about further insight into cue weighting of L2 vowels. Phonological short-term memory (PSTM) is one of the components within learners' cognitive individual differences that has been thoroughly examined in recent years. PSTM is one of the components of working memory. It is divided into two main storages, the phonological or articulatory loop and the visual-spatial sketchpad, which control the verbal and visual-spatial domains respectively. The phonological loop is responsible for the storage of verbal information over short periods of time (Baddeley and Hitch 1974; Baddeley 1986, 2003). It is formed by a short-term phonological store and an articulatory rehearsal component. The auditory traces that reach the short-term phonological store are subject to decay within approximately two seconds, unless maintained in the articulatoy rehearsal component. Research on language acquisition has established a strong relation between subjects' variation in PSTM capacity and children's successful L1 and L2 acquisition in areas such as vocabulary, syntax, semantics and oral fluency (Blake, Austin, Cannon, Lisus and Vaughan 1994; Adams and Gathercole 1996; Cheung 1996; Gathercole, Hitch, Service and Martin 1997; Adams, Bourke and Willis 1999; Dufva and Voeten 1999; Adams and Gathercole 2000; French 2006). More recently, it has been shown that PSTM significantly predicts children's' L2 grammar development and vocabulary knowledge (French and O'Brien 2008). However, little is known about the role that PSTM might play in the acquisition of L2 adult learners. Only a few studies have looked into this issue in formal and immersion learning contexts, concluding that PSTM makes an important contribution to adults L2 oral fluency development (O'Brien, Segalowitz, Freed and Collentine 2007).

PSTM is commonly measured by tasks such as nonword repetition (NWR), serial nonword recognition (SNWR), immediate serial recall (ISR) and the digit span test (DST). Most typically, PSTM has been measured in the literature by NWR tasks. However, the reliability of these measures has been called into question; it has been suggested that PSTM might increase as a result of language development and that, therefore, PSTM task performance could be directly influenced by language knowledge (Ottem, Liam and Karlsen 2007). However, the results of some recent studies suggest that tasks, such as nonword repetition provide similar measures of PSTM irrespective of the language of the nonwords (French and O'Brien 2008). Whereas it is true that existing knowledge about language may enhance PSTM function, it does not directly influence the quality of the temporary storage in itself (Gathercole 2006). In keeping with this view, the storage capacity of short-term memory may be considered to be language-independent, implying that PSTM abilities do not change during young adulthood (O'Brien et al.2007).

The goal of the present study was to explore the relationship between PSTM and L2 phonological acquisition by adults. Studies investigating PSTM and SLA in adults are scarce and no study, to the best of our knowledge, has examined the role that PSTM might play in L2 phonological acquisition by adults. More specifically, it was our intention to investigate the cues that Catalan-Spanish learners use in the categorisation of English /iː/-/ɪ/. We hypothesised that, as suggested by previous research, learners would make use of duration cues to a greater extent than native English speakers. In addition, we aimed at exploring possible individual differences among the learners as a function of their PSTM capacity. In other words, our objective was to establish a possible relationship between the learners' PSTM and their ability to process durational and spectral information in a native-like manner in the categorisation of /iː/-/ɪ/. It was hypothesised that learners with a higher PSTM capacity would categorise this contrast more accurately than those with a lower PSTM capacity, given the body of evidence that points to the existing relationship between PSTM and language development in other areas. Finally, this study also investigated the possible influence of language knowledge on the learners' PSTM task performance. That is, to examine whether learners obtained higher or lower scores depending on the language of the nonwords presented in the PSTM task. We hypothesised that the participants' performance would not significantly vary as a function of the nonwords language in accordance with previous research (Gathercole 2006; O'Brien et al. 2007).

2. METHOD

2.1.1. The participants

The participants in this study were 84 Catalan-Spanish bilingual students enrolled in a degree in English Studies at the University of Barcelona (mean age= 21.5), and 13 native speakers of Southern-British English (mean age= 40.2). All the participants were bilingual speakers who used both languages on a daily basis. They were asked to fill in a questionnaire examining their linguistic background, which provided information such as the number of languages spoken and the proficiency level in each of them. All reported to have normal hearing and no speech-related dysfunctions.

2.1.2. Vowel categorisation task

In order to investigate the cues that Catalan-Spanish learners used in the categorisation of /iː/-/ɪ/, a lexical decision task with words including this contrast was designed. A male native speaker of Southern-British English recorded several tokens of the words *feet-fit*. The mean duration for each vowel (/iː/ and /ɪ/) was calculated. The mean for the tense vowel was 50 ms, and for the lax vowel 190 ms. Using Praat (Boersma and Weenink 2007), two continua were created that consisted of 8 equidistant steps each; the duration of the tense vowel was shortened from 190 to 50 ms. and, conversely, the duration of the lax vowel lengthened, from 50 to 190 ms. The items that made up the two continua were repeated ten times each and randomised. A total of 160 stimuli were presented aurally to the participants in a lexical decision task using Praat. In addition to written instructions shown on the screen, the participants were given one stimulus at a time and had to choose between the two displayed options: *feet* or *fit*. They were instructed to answer as fast and accurately as possible.

2.1.3. PSMT: Serial nonword recognition (SNWR)

In order to assess PSTM, a SNWR task was created. In this kind of task, participants are asked to determine whether two strings of nonwords increasing in length appear in the same or in different order. PSTM has been typically measured in the literature using word/nonword recall and repetition. However, SNWR was chosen as a measure of PSTM in this study for two main reasons. Firstly, SNWR does not involve an articulation component that, due to articulatory demands, might hinder subjects' performance and affect the interpretation of the results obtained from the task (Snowling, Chiat and Hume 1991). Secondly, SNWR has been shown to minimise the effects of lexical influences on phonological memory (Gathercole, Pickering, Hall and Peaker 2001). In this sense, SNWR could be considered a more accurate measure of PSTM than nonword recall and repetition, having been used in recent studies on individual differences in cognitive ability and L2 oral production (O'Brien et al. 2006, 2007) and perception (Isaacs and Tromfimovich *in press*).

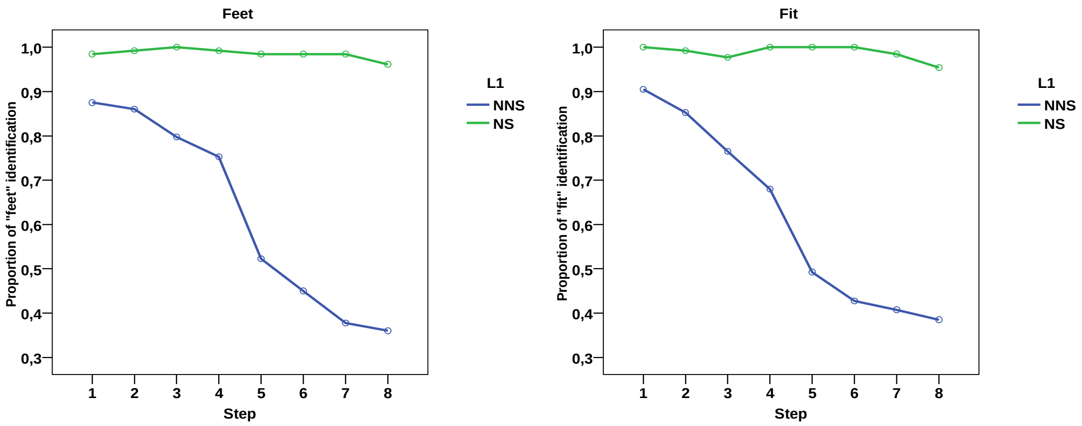
As a means to examine whether PSTM tasks might be influenced by language knowledge, nonwords in three different languages were selected for the SNWR test: Catalan, English and Russian (none of the participants knew Russian). They were recorded randomly by a native speaker at normal speed in a sound-proof booth, and three different SNWR tests were constructed, one for each of the three languages. Each test consisted of 144 nonwords, except for Russian, in which words and nonwords were mixed indistinctively. All of the nonwords followed a CVC pattern which conformed to the phonotactic regularities of each language and none of them in one language could be interpreted as a word in any of the other languages. The nonwords were organized into eight sequences and into three different sequence lengths: five, six, and seven items, all containing a different vowel sound. Each sequence length included four same and four different pair sets. In the same-pair sets, the sequence of nonwords was repeated twice. In the different-pair sets, the repetition of the second sequence was identical to the previous one, but for two adjacent elements which were transposed. In order to minimise the salience of the transposed items, the first or the last item in the sequence was never transposed. In addition, two same and two different sequence pairs of four items each was also constructed for practice before the subjects started the test.

DmDx display software (Forster and Forster 2003) was employed to present the SNWR tasks to the participants and to record their responses. The task was explained both by the experimenters and also by written instructions that were displayed on the screen. The participants started the test with the practice trials, and the task began when they reported to have fully understood the procedure. The order in which the three SNWR tasks were presented was counterbalanced across participants. In each sequence-length, the items were presented at a rate of one every 0.3 ms. and with a pause of 1ms.between same or different sequences. The duration of the nonwords sequences was measured at approximately 300 ms. with a small variation among the languages. The number of correct responses out of 24 was recorded and used as a measure of PSTM (Isaacs and Trofimovich *in press*).

3. RESULTS

The results of the lexical decision task revealed, as expected, that the Catalan-Spanish non-native speakers of English (NNSs) over-relied on duration in the lexical categorisation task, as opposed to the group of native speakers of English (NSs), who were insensitive to the duration manipulation (see Figure 1 below). The shorter the duration of the tense vowel /iː/ (*feet*), the more often it was identified as /ɪ/ (*fit*), and the longer the duration of the lax vowel /ɪ/ (*fit*), the more often it was identified as /iː/ (*feet*).

Figure 1: Duration manipulation effect on the perception of the *feet* and *fit* continua.



An overall measure of perceptual accuracy for the /iː/-/ɪ/ contrast was derived from the lexical decision task independently for every continuum (percent correct identification averaged across the 8 steps of each continuum) and this was used as a measure of perceptual accuracy for the contrast. Higher values represented less reliance on duration and more native-like performance (i.e. the use of spectral information as a primary cue). Because the NNS group obtained slightly different mean percent correct identification scores for the two continua, 63.27 (*SD*=27.07) for *feet* vs. 61.07 (*SD*=27.40) for *fit*, and this difference approached significance ($t(83)=1.87$, $p=.065$), we did not obtain an overall mean for both continua. The analysis of the relationship between PSTM and perceptual accuracy was therefore conducted independently for the *feet* and the *fit* continua.

The hypothesised language-dependent performance on the SNWR tasks was tested first by obtaining Pearson product-moment correlation coefficients among the scores obtained for the Catalan, English and Russian SNWR tasks, and then by conducting a one-way repeated-measures ANOVA on the PSTM scores (3 levels: Catalan, English and Russian). On the basis of previous research (e.g. French and O'Brien 2008) we were expecting higher PSTM scores on the Catalan SNWR task and lower scores on the English and Russian SNWR tasks. These predictions were not confirmed by the results, but significant differences in the scores were found. The scores of the three SNWR tasks were significantly correlated with one another (see Table 1), indicating that the scores obtained were equally reliable irrespective of the language of the nonwords, but the mean percent correct recognition was slightly higher in Catalan and Russian than in English, and the correlations were stronger between Catalan and Russian. The ANOVA revealed an overall significant effect for Language ($F(2, 82)=3.31$, $p=.041$) and further pairwise comparisons indicated that differences among the scores as a function of the nonword language were only significant between English and Russian ($p=0.35$)

Table 1: PSTM scores and correlations (**p= <.01)

SNWR percent correct identification				Correlations (Pearson *r*)	
Nonword language	Mean (*SD*)	Low PSTM	High PSTM	English	Russian
Catalan	63.84 (17.7)	52.87 (10.16)	76.49 (6.79)	.451**	.606**
English	60.91 (14.73)	59.07 (15.52)	69.33 (11.64)		.477**
Russian	65.38 (16.17)	56.56(12.99)	73.08 (11.22)		

In order to explore the relationship between PSTM and NNSs degree of reliance on duration as a cue in the identification of /iː/ (*feet*) and /ɪ/ (*fit*), NNSs were assigned to either Low or High PSTM capacity groups through median split (see Table 1 above for mean percent correct recognition). NNSs correctly identifying

15 (out 24) nonword sequences or below were assigned to the Low PSTM group in the Catalan and English SNWR task, 16 or below in the Russian SNWR task. Independent-samples t-tests were then conducted to test whether NNSs with higher PSTM abilities performed better on the lexical decision task. The results revealed significant differences between the Low and High (Russian) PSTM groups for the *feet* continuum ($t(82)=-2.37$, $p=.036$) and consistently present higher identification scores for the High PSTM group (see Table 2), approaching significance for the Low and High (English) PSTM groups ($t(82)=-1.92$, $p=.057$) and reaching significance for the Low and High (Catalan and Russian, averaged) PSTM groups ($t(82)=-2.01$, $p=.047$).

Table 2: % correct identification in the *feet* and *fit* continua as a function of PSTM capacity (Low vs. High).

Catalan			English			Russian		
Cont.	PSTM	Mean (*SD*)	Cont.	PSTM	Mean (*SD*)	Cont.	PSTM	Mean (*SD*)
feet	Low	62.13 (26.15)	*feet*	Low	58.05 (27.66)	*feet*	Low	57.79 (25.53)
	High	64.58 (28.37)		High	69.29 (25.37)		High	70.23 (27.69)
fit	Low	59.83 (27.69)	*fit*	Low	56.00 (26.75)	*fit*	Low	56.40 (25.46)
	High	62.50 (27.35)		High	66.92 (27.31)		High	66.99 (28.95)

4. CONCLUSIONS

The aim of this study was to investigate the possible relationship between PSTM and L2 phonological acquisition by adults. More specifically, to investigate Catalan-Spanish learners of English cue weighting of /iː/-/ɪ/, and the relationship that vowel categorisation for this contrast might have with PSTM. In order to do this, two different experiments were conducted. Firstly, a lexical-decision task was designed presenting two 8-step continua, *feet* and *fit*, in which vowel duration was shortened and lengthened, respectively. The participants were asked to categorise the /iː/-/ɪ/ contrast included in these words. The results support previous research in that the participants over-relied on duration in the categorisation of this contrast as opposed to native speakers (Bohn 1995; Escudero and Boersma 2004; Cebrian 2006, 2007; Mora and Fullana 2007; Cerviño and Mora 2009), as measured by the mean percent of correct identification averaged across the 8 steps of each continuum. An explanation to learners' over-reliance on duration might be found in Bohn's (1995) Desensitization Hypothesis that claims that when spectral differences to distinguish vowel contrasts are insufficient to learners, because previous linguistic experience did not sensitize them to spectral differences, duration will be used.

Secondly, in order to investigate possible individual differences in cognitive ability, a test of PSTM was constructed. A SNWR task was chosen as a measure of PSTM since it has been shown to have some advantages over traditional measures used in the literature, such as word/nonword recall and repetition. Some of these advantages include the fact that SNWR does not involve an articulatory component that might add difficulty to the task (Snowling et al. 1991), and that SNWR has been shown to minimise lexical influence on PSTM (Gathercole et al. 2001). To investigate possible language knowledge influence on the participants' task performance as suggested by previous research (Ottem et al. 2007), a SNWR task presenting CVC nonwords conforming to the phonotactic regularities of three different languages was constructed: L1(Catalan), L2 (English) and Lx (Russian). Statistical analysis showed that the three languages were strongly correlated with one another. However, contrary to our expectations, the participants of this study did not obtain higher PSTM scores in their L1, i.e., the Catalan SNWR task, as opposed to the English and Russian SNWR tasks. An interesting finding is that the participants obtained a slightly higher percentage of correct recognition in Catalan and Russian than they did in English, and that the correlations were stronger between Catalan and Russian. This finding is noteworthy and should be further explored in future studies, in order to clarify the reason why learners might obtain higher scores (albeit statistically non-significant) in an Lx than in their L2. Taken together, these results suggest that SNWR provides a reliable measure of PSTM irrespective of language knowledge. Finally, in order to examine the possible relationship between PSTM and native-like cue weighing of /iː/-/ɪ/, NNS were assigned to either Low or High PSTM capacity groups. The results revealed that those participants in the High PSTM capacity group in each of the three languages obtained higher identification scores in the lexical decision task than those in the Low PSTM capacity group. This finding suggests that there is a relationship between PSTM and L2 cue weighting of /iː/-/ɪ/. Additional research is needed in order to extend the results of this study to other contrasts that present difficulty for L2 learners, and that might further confirm the relationship between PSTM and L2 phonological acquisition.

5. REFERENCES

Adams, A. M., Gathercole, S. E.1996. Phonological working memory and spoken language development in young children. *The Quarterly Journal of Experimental Psychology* 41 (1). 216–233.

Adams, S., Bourke, L., Willis, C. 1999. Working memory and spoken language comprehension in young children. *International Journal of Psychology*. 34. 364-373.

Adams, A. M., Gathercole, S. E. 2000. Limitations in working memory: Implications for language development. *International Journal of Language and Communication Disorders* 35. 95–116.

Baddeley, A., Hitch, G. 1974. Working memory. In: Bower, G. A. (ed.), *The psychology of learning and motivation*. New York: Academic Press. 47-89.

Baddeley, A. 1986. *Working memory.* Oxford: Clarendon Press.

Baddeley, A. D. 2003. Working memory and language: An overview. *Journal of Communication Disorders* 36. 189-208.

Blake, J., Austin,W., Cannon, M., Lisus, A., Vaughan, A.1994. The relationship between memory span and measures of imitative and spontaneous language complexity in preschool children. *International Journal of Behavioral Development* 17. 91–107.

Bohn, O. S. 1995. Cross-language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), Speech perception and linguistic experience (pp. 279-302). York Press, Baltimore.

Boersma, P., Weenink, D. 2007. Praat: doing phonetics by computer (version 4.6.15) [Computer program], http://www.praat.org/

Cebrian, J. 2006. Experience and the use of duration in the categorisation of L2 vowels. *Journal of Phonetics* 34. 372-387.

Cebrian, J. 2007 Old sounds in new contrasts: L2 production of the English tense-lax vowel distinction. *Proceedings of the 16th International Congress of the Phonetics Sciences*, Saarbrucken, Germany.

Cerviño, E., Mora, J.C. 2009. Duration as a phonetic cue in the categorisation of /iː/-/ɪ/ by Spanish/Catalan learners of English. *Journal of the Acoustical Society of America*, Vol. 125, No. 4, Pt. 2, April, p. 2764.

Cheung, H. 1996. Nonword span as a unique predictor of second language vocabulary learning. *Developmental Psychology* 32 (5). 867-873.

Dufva, M., Voeten, J.M. 1999. Native language literacy and phonological memory as prerequisites for learning English as a foreign language. *Applied Psycholinguistics* 20. 329-348.

Escudero, P. , Boersma, P. 2004. Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition* 26. 4: 551-585.

Flege, J. E. 1991. Orthographic evidence for the perceptual identification of vowels in Spanish and English. *Quarterly Journal of Experimental Psychology,* 43. 701-731.

French, L. M. 2006. *Phonological working memory and L2 acquisition: a developmental study of Quebec francophone children learning English*. New York: Edwin Mellen Press.

French, L., O'Brien 2008. Phonological memory and children's second language grammar learning. Applied Psycholinguistics 29. 463-487.

Forster, K.I., Forster, J.C. 2003. DmDX: A window display program with millisecond accuracy. *Behavior Research Methods, Instruments & Computers* 35(1). 116-124.

Gathercole, S. E., Hitch, G. J., Service, E., Martin, A. J. 1997. Phonological short-term memory and new word learning in children. *Developmental Psychology* 33. 6: 966-979.

Gathercole, S. E., Pickering, S. J., Hall, M., Peaker, S. M. 2001. Dissociable lexical and phonological influences on serial recognition and serial recall. *Quarterly Journal of Experimental Psychology* 54(A). 1-30.

Gathercole, S.E. 2006. Nonword and word learning: the nature of the relationship. *Applied Psycholinguistics* 27. 513-543.

Isaacs, T., Trofimovich, P. (in press). Phonological memory, attention control, and musical ability: Effects of individual differences on rater judgments of L2 speech. *Applied Psycholinguistics*.

Mora, J. C., Fullana, N. 2007. Production and perception of English /iː/-/ɪ/ and /æ/ /ʌ/ in a formal setting: Investigating the effects of experience and starting age. *Proceedings of the 16th International Congress of the Phonetics Sciences*, pp. 207-221. Saarbrucken, Germany.

O'Brien, I., Segalowitz, N., Collentine, J., Freed, B. 2006. Phonological memory and lexical, narrative, and grammatical skills in second language oral production by adult learners. *Applied Psycholinguistics* 27. 377-402.

O'Brien, I., Segalowitz, N., Freed, B., Collentine, J. 2007. Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition* 29. 577-582.

Ottem, E. J., Liam, A., Karlsen, P. J. 2007. Reasons for the growth of traditional memory span across age, *The European Journal of Cognitive Psycholog,* 19.233-270.

Snowling, M., Chiat, S., Hulme, S. 1991. Words, nonwords, and phonological processes: Some comments on Gathercole, Willis, Emslie and Baddeley. *Applied Psycholinguistic* 12. 369-373.

Ylinen,S., Uther, M., Latvala, A., Vepsäläinen, S.M., Iverson, P., Akahane-Yamada, R., Näätänen, R. 2009. Training the brain to weight speech cues differently: a study of Finnish second-language users of English. *Journal of Cognitive Neuroscience* 1-14.

Learning to produce a multidimensional laryngeal contrast

Charles B. Chang[*]

University of California, Berkeley
cbchang@cal.berkeley.edu

ABSTRACT

Research on how second-language (L2) learners acquire L2 laryngeal categories has focused on languages with "voiced" and "voiceless" categories that differ in terms of one main cue: voice onset time. The present study examines how L2 learners come to produce a laryngeal contrast that requires the use of a second phonetic dimension—namely, the three-way Korean laryngeal contrast among lenis, fortis, and aspirated stops. In a five-week longitudinal study, 26 adult native English speakers learning Korean completed a reading task in which they pronounced Korean stops in a low vowel context. Results of acoustic analyses show that while the majority of learners are eventually successful at producing a full three-way contrast, there is wide variation in the way in which they produce it. This paper describes the range of variation in phonetic spaces that learners produce, shows how these differ from the findings of cross-linguistic perception studies on English speakers hearing Korean, and concludes that a perseverative kind of "equivalence classification" plays a large role in how learners link L2 laryngeal categories to L1 laryngeal categories.

Keywords: laryngeal contrast, voice onset time, fundamental frequency, Korean, equivalence classification.

1. INTRODUCTION

Research on how second-language (L2) learners acquire laryngeal categories that differ from the laryngeal categories of their first language (L1) has generally concentrated on languages with two laryngeal categories differing between L1 and L2 in terms of the same primary cue: voice onset time, or VOT (e.g. French and English: Caramazza *et al*. 1973, Flege 1987; Spanish and English: Flege and Eefting 1988; Italian and English: Flege *et al*. 1995; Portuguese and English: Major 1996). In the present study, I examine how L2 learners come to produce a laryngeal contrast that requires the use of a second phonetic dimension in addition to VOT—namely, the three-way Korean laryngeal contrast among lenis, fortis, and aspirated stops, which in initial position differ primarily in terms of VOT and fundamental frequency (f_0) onset (cf. Han and Weitzman 1970, Kim 2004, *inter alia*). How do learners use (or not use) f_0 onset in conjunction with VOT to realize this three-way contrast?

Relatively little work on L2 speech has examined Korean as L2, rather than L1. With regard to L2 perception of Korean, two studies have examined how L1 English speakers interpret Korean word-initial stop consonants. Francis and Nusbaum (2002) found that before training, L1 English speakers (naïve listeners who were not learning Korean) mostly relied on differences in VOT (and co-varying differences in rate of amplitude change) to distinguish the three laryngeal categories, but after training, seemed to use both VOT and f_0 onset differences (along with co-varying differences in the clarity of formant structure at vowel onset) to distinguish them (however, see Shin 2007 for differing results with trained learners of Korean). The perceptual data show, moreover, that after training, English speakers' perception approximates that of native Korean speakers, who break up the [VOT x f_0] phonetic space in the manner shown in Kim (2004), where tokens with short-lag VOT are consistently perceived as fortis and tokens with long-lag VOT are perceived as either lenis or aspirated depending on the VOT and on the f_0 onset.

While Francis and Nusbaum's (2002) perception study utilized identification and difference rating tasks, Schmidt's (2007) cross-linguistic perception study instead had subjects—also L1 English speakers with no knowledge of Korean—label Korean sounds as the perceptually closest English sound and rate the similarity of the English sound to the Korean sound. Her results show that subjects overwhelmingly labeled Korean lenis stops and aspirated stops as English voiceless stops and Korean fortis stops as English voiced stops. However, the Korean categories differed in terms of how similar to English categories they were perceived as being: aspirated stops were rated as more similar to English stops than lenis or fortis stops were. This suggests that for L1 English learners of Korean, the default "equivalence classifications" (Flege 1987) of Korean and English stops are aspirated-voiceless, lenis-voiceless, and fortis-voiced, but that the strength of the cross-language category identification varies across category pairings.

Whereas research on L2 perception of Korean has often focused on L2-naïve subjects, studies that have looked at L2 production of Korean have generally examined people actively learning the language. In one such study, Kim and Lotto (2002) found that intermediate Korean learners (most of whom were L1 English speakers) produced distinctions between the three stop types using VOT, but not closure duration or f_0 onset. Shin's (2007) study of elementary Korean learners resulted in similar findings with L1 English learners, who tended to rely just on VOT to produce the contrast. On the other hand, learners whose L1 was a tone language (e.g. Mandarin, Cantonese) were found to use f_0 as a cue more often than the L1 English learners.

Taken together, the results of studies of L2 perception and production of Korean suggest that L1 English speakers, and perhaps speakers of non-tone languages more generally, can be trained to use f_0 in perception, but nevertheless tend to utilize VOT rather than f_0 to distinguish the Korean laryngeal categories in production. This pattern of production contrasts with that of mature native speakers, who use both dimensions, as well as with that of children acquiring Korean as L1, who separate the lenis and aspirated categories in f_0 well before they separate them in VOT (cf. Jun 2006), although by the age of five years they use both cues reliably (Lee and Iverson 2008).

Although Schmidt's (2007) cross-linguistic perceptual findings show consistency in the way learners assimilate Korean categories to English categories, they make no predictions regarding how learners will distinguish the lenis and aspirated categories that are both assimilated to the voiceless category of English. Kim and Lotto (2002), as well as Shin (2007), suggest that learners mainly use VOT to distinguish these categories in production; however, the amount of VOT overlap between learners' lenis and aspirated stop productions—even within one place of articulation—is so large that it is unclear whether learners are actually producing a reliable three-way contrast in VOT.

A second reason to re-examine L2 learners' production of this contrast is the existence of a conflict in cues contributing to cross-language equivalence classification. If we were to pair the Korean and English laryngeal categories on the basis of phonetic similarity (specifically, in terms of similarity in VOT and f_0), aspirated stops would be paired with voiceless stops, since these categories are both long in VOT and high in f_0 onset. However, it is unclear how lenis stops and fortis stops should be classified, since each of these categories resembles voiced stops in one way and voiceless stops in another way. Lenis stops are relatively long in VOT like voiceless stops, but low in f_0 like voiced stops; fortis stops, on the other hand, are short in VOT like voiced stops, but high in f_0 like voiceless stops. Thus, linking lenis and fortis stops to English categories is not straightforward, given that most English speakers show some degree of sensitivity to the f_0 difference between voiced and voiceless stops (cf. Haggard *et al*. 1970).

In the present study, I re-examine how L1 English late learners of Korean produce the Korean laryngeal contrast, focusing on an L1 and L2 that do not share the same orthography to avoid the confound of orthographic equivalence present in the majority of studies on L2 voicing categories. The main research question is the following: given little to no explicit phonetic instruction, how successful are late learners of Korean at producing Korean laryngeal categories like native speakers? We will see if, using VOT and f_0 onset, learners manage to produce a full three-way contrast, as well as if they are consistent in their L2 phonetic spaces. Finally, we will make some generalizations about the nature of learners' deviation from the native Korean phonetic space.

2. METHODS

A production experiment was conducted weekly starting from one week into the language class that study participants were taking. Every week participants completed a reading task in which they saw a Korean stimulus (spelled in Korean orthography) and read it aloud. Stimuli were presented a total of four times, once each in four randomized blocks following a practice session of five items. Each item was presented on screen for 1.5 seconds and then replaced by a picture of a green traffic light to cue the participant to produce the item. Audio was recorded via a head-mounted condenser microphone for two seconds starting at the time point at which the green light appeared on screen, and the inter-stimulus interval from the end of this recording to the presentation of the following item was one second. All stimuli presentation and audio recording was done in DMDX 3.2.6.3 (Forster 2008) on a laptop computer.

The set of Korean stimuli consisted of 22 Korean monosyllables representing most of the phonemic contrasts in the language. The stimuli were generally of the form CV to make them as easy as possible for novice learners to read, with the vowels in the nine critical items (3 laryngeal categories x 3 stop places of articulation) being uniformly /a/. The same set of stimuli was used in every week of the study.

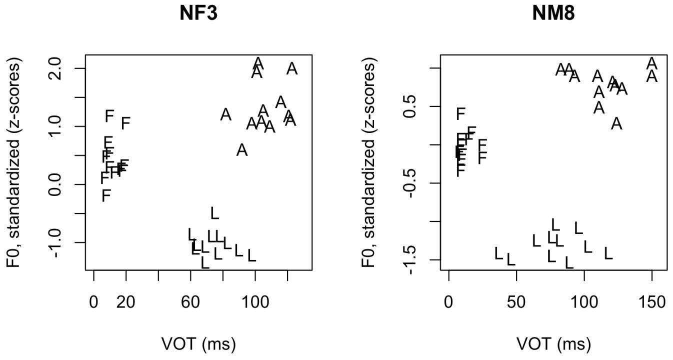
Participants were 26 late learners of Korean (4 males, 22 females; 21–26 years old), native speakers of American English with no prior exposure to Korean taking a six-week course of intensive Korean immersion instruction at the time of the study. On average these learners received four hours of instruction a day, for a total of approximately 82 hours of instruction by the end of the program (roughly equivalent to one semester of college-level Korean). In exit questionnaires, participants reported that class time constituted the majority of their experience with Korean, both in terms of listening and speaking.

Acoustic analysis of recordings was conducted using Praat 5.0.26 (Boersma and Weenink 2008). Manual measurements of VOT and f_0 onset were taken on learners' productions of critical items. VOT was measured off a wide-band Fourier spectrogram with a Gaussian window shape (window length: 5 ms; dynamic range: 50 dB; pre-emphasis: 6.0 dB/oct) as the time at voicing onset minus the time at the stop burst. To obtain stable measurements of f_0 onset, the average wavelength of the first three regular glottal periods in the vowel was calculated from the waveform and converted into a frequency value. Initial periods were skipped if they were irregular (e.g. more than 33% longer or shorter than the following period); however, tokens requiring more than five periods of the vowel onset to be skipped were discarded. In order to put male and female learners on the same f_0 scale, raw f_0 measurements were furthermore standardized to *z*-scores by learner (by subtracting the learner's mean f_0 over the duration of the study and dividing by the square root of the learner's variance in f_0 over the duration of the study).

3. RESULTS

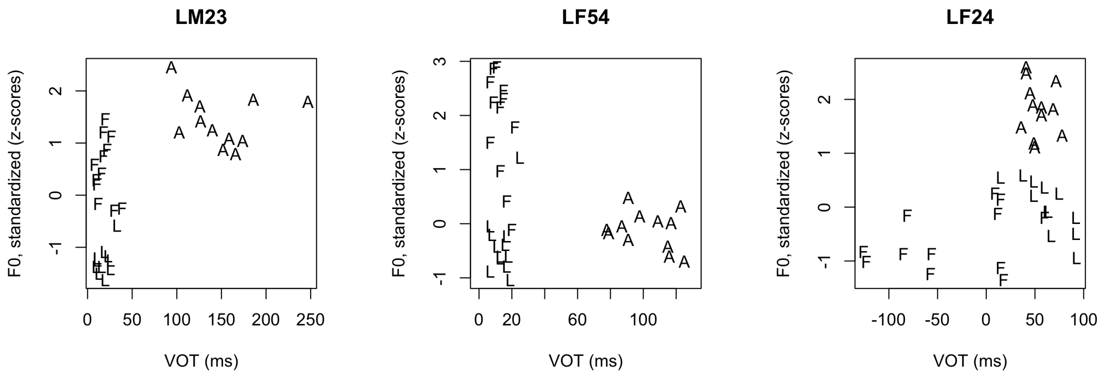
The phonetic spaces of native Korean speakers are generally consistent with Kim (2004) in terms of how the Korean laryngeal categories are realized with respect to VOT and f_0 onset. Fortis stops are produced with short VOT and an elevated f_0 onset; lenis stops are produced with longer VOT and a low f_0 onset; and aspirated stops are produced with the longest VOT and the highest f_0 onset (cf. Figure 1). For most native speakers, lenis and aspirated stops overlap considerably in VOT, and fortis and aspirated stops overlap considerably in f_0, but none of these categories overlap in both dimensions. Thus, VOT and f_0 are necessary and sufficient cues for distinguishing the three laryngeal types.

Figure 1: Representative scatter plots of native Korean speakers' productions (L = lenis, F = fortis, A = aspirated).



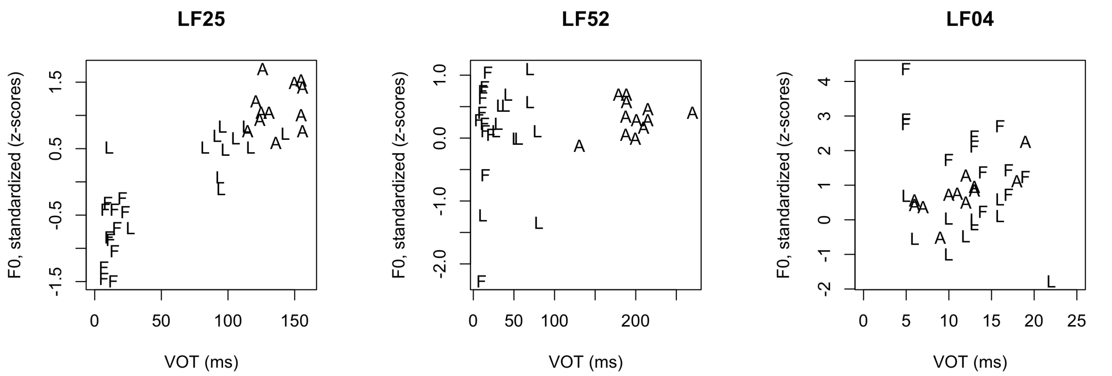
The phonetic spaces of L2 learners look markedly different. One of the most common patterns is found in Groups A (n=7) and B (n=2), where learners essentially produce two two-way contrasts, each in one dimension. In subgroup A1, lenis and fortis stops are both produced with short VOT and are contrasted on f_0, while fortis and aspirated stops tend to be produced with similar f_0 and are contrasted on VOT (cf. Figure 2, LM23). Subgroup A2 is similar, except that aspirated stops are produced with a relatively low f_0 onset in the range of the lenis stops rather than an elevated f_0 onset in the range of the fortis stops (cf. Figure 2, LF54). Subgroup A3 resembles subgroup A2, except lenis and fortis stops are reversed in the f_0 dimension: lenis stops are produced with higher f_0 than fortis stops, though lenis and aspirated stops are still produced in the same f_0 range. In Group B, fortis and lenis stops are produced in the same f_0 range and are distinguished on the basis of VOT, while lenis and aspirated stops are produced in the same VOT range and are distinguished on the basis of f_0 (cf. Figure 2, LF24).

Figure 2: Representative scatter plots of Week 5 productions in learner groups A and B (L = lenis, F = fortis, A = aspirated).



In Group C (n=7), learners produce a three-way contrast using either VOT, f_0, or both dimensions. The learners in subgroup C1 (e.g. LF25, cf. Figure 3) make use of both VOT and f_0 to make the contrast, producing fortis stops with short VOT and low f_0, lenis stops with longer VOT and higher f_0, and aspirated stops with the longest VOT and highest f_0. However, the learners in subgroup C2 (e.g. LF52, cf. Figure 3)—much like the learners described in Kim and Lotto (2002) and Shin (2007)—rely just on VOT to make a three-way contrast. In contrast, the learner in subgroup C3 (LF04, cf. Figure 3) relies almost entirely on f_0 to make the contrast, producing all three categories in the short-lag VOT range and distinguishing between them by producing lenis stops with the lowest f_0, aspirated stops with intermediate f_0, and fortis stops with the highest f_0.

Figure 3: Representative scatter plots of Week 5 productions in learner group C (L = lenis, F = fortis, A = aspirated).



Finally, a minority of learners fail to keep the three categories apart with these cues. In Group D (n=8), learners just produce a two-way contrast, showing nearly all possible types of merger, while in Group E

(n=2), learners do not keep any of these categories distinct from the others in terms of VOT and/or f_0 onset, producing all of them over the same wide phonetic space.

4. DISCUSSION AND CONCLUSIONS

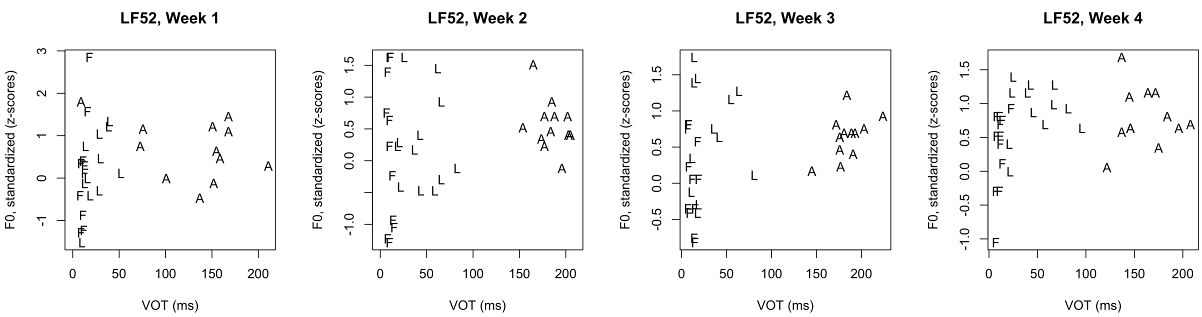
As seen above, there is wide variation in learners' success at restructuring the L1 phonetic space of two laryngeal categories into an L2 phonetic space of three laryngeal categories resembling native Korean. Some learners fail to produce a three-way contrast, merging two or more categories with different degrees of overlap, but the majority of learners do manage to produce three distinct categories. In addition, there is a dichotomy in the phonetic spaces of learners who produce a three-way contrast (the "full distinguishers") and those who only produce a two-way contrast (the "partial distinguishers"). In both groups, some learners appear to identify lenis stops as a category similar to voiced stops—full distinguishers separating fortis stops from lenis stops on the basis of f_0 onset, and partial distinguishers combining fortis and aspirated stops into a category similar to voiceless stops. However, in both groups there are other learners who identify fortis stops as the voiced-like category. Here the full distinguishers separate lenis stops from fortis stops on the basis of VOT and/or f_0, while the partial distinguishers combine lenis and aspirated stops into a voiceless-like category. These findings are consistent with the predictions of the ambiguous cross-linguistic category correspondences described above. Despite having the same L1 background, learners interpret this L2 contrast in multiple ways, resulting in disparate phonetic spaces of the L2 contrast that all depart in one or more ways from the native phonetic space.

We are left then to wonder: why is there so much variation? If we ignore the influence of affective variables, which, as suggested by background questionnaires, do not differ across the groups delineated above in any clear way, we are left with three possible explanations for the variation in learner production.

First, variation in production may be attributable to variation in input. After all, learners had different teachers, and there are some differences among the teachers in production, though the general pattern is the same (cf. Figure 1). Inspection of differences among learners along with their class affiliations does not support this hypothesis, however. For example, learners LM23 and LF54 were in the same class, yet still differ from each other: LM23 produces aspirated stops with high f_0, while LF54 produces them with low f_0 (cf. Figure 2), even though their teachers both produce them with high f_0. These facts indicate that even if some inter-learner variation is rooted in input disparities, input cannot be the whole story.

Second, there could be differences across participants with respect to how VOT and f_0 are weighted in distinguishing English voiced and voiceless stops. This variability in cue weighting could lead to variation in L2 production, in that learners would not necessarily be biased towards the same schemas of L1-L2 equivalence classifications. The fact that there is some variability among English speakers with respect to how sensitive they are to f_0 as a cue to the English voicing contrast (cf. Haggard *et al.* 1970) is consistent with this explanation—an interesting possibility that should be tested more thoroughly.

Figure 4: Scatter plots of learner LF52's productions in Weeks 1–4 (L = lenis, F = fortis, A = aspirated).



Third, learners might utilize explicit strategies to achieve L2 contrast that may or may not be based on actual L2 input patterns (such strategies being likely to differ between individuals). In fact, strategy does seem to account for what at least some learners do. For instance, learner LF52 (who produces a three-way contrast in the VOT dimension only, cf. Figure 3) expressed in study debriefings that she thought the contrast just had to do with aspiration, and so she ignored pitch. This sort of strategic bias largely accounts for why she started producing a three-way VOT contrast in Week 1 of the language program and continued to do so through Week 5, failing to make significant use of f_0 at all time points in this study (cf. Figures 3–4).

One noteworthy aspect of this production study is that the results differ substantially from those of the perception studies described above. Relatively few learners produce the L2 laryngeal categories with a phonetic space that might be predicted from cross-linguistic perception results or with one resembling that of native speakers. Moreover, there is a large amount of variation in learners' phonetic spaces, in contrast to the high degree of consistency seen in the perceptual performance of listeners in Schmidt (2007). This variation in L2 production spaces suggests that a number of factors are at work in the acquisition of L2 speech that are not necessarily seen in naïve non-native perception of an L2. Some possible sources of this variation have been discussed here, though much more work is needed to tease apart their effects.

While learners show a high degree of variation in the organization of their L2 phonetic spaces, what is consistent among them is that, with few exceptions, the production pattern they show in Week 5 is largely the same as the one they show in Week 1, suggesting that initial L1-L2 equivalence classifications tend to persevere, rather than change over the course of acquisition. The implication for L2 learning is clear: building an accurate representation of an L2 sound early in acquisition is crucial, since changing this representation significantly may become increasingly difficult later on.

5. REFERENCES

Boersma, P., Weenink, D. 2008. Praat: Doing phonetics by computer. <http://www.praat.org>. Last access 15 Jun 2008.

Caramazza, A., Yeni-Komshian, G., Zurif, E., Carbone, E. 1973. The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *Journal of the Acoustical Society of America* 54(2). 421–428.

Chang, C. In press. The implementation of laryngeal contrast in Korean as a Second Language. In: Kuno, S. et al. (eds.), *Harvard Studies in Korean Linguistics XIII*. Seoul: Hanshin Publishing Company.

Flege, J. 1987. The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15(1). 47–65.

Flege, J., Eefting, W. 1988. Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation. *Journal of the Acoustical Society of America* 83(2). 729–740.

Flege, J., Munro, M., MacKay, I. 1995. Effects of age of second-language learning on the production of English consonants. *Speech Communication* 16(1). 1–26.

Forster, J. 2008. DMDX. <http://www.u.arizona.edu/~kforster/dmdx/dmdx.htm>. Last access 15 Jun 2008.

Francis, A., Nusbaum, H. 2002. Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance* 28(2). 349–366.

Haggard, M., Ambler, S., Callow, M. 1970. Pitch as a voicing cue. *Journal of the Acoustical Society of America* 47(2B). 613–617.

Han, M., Weitzman, R. 1970. Acoustic features of Korean /P, T, K/, /p, t, k/, and /p^h, t^h, k^h/. *Phonetica* 22(2). 112–128.

Jun, S.-A. 2006. Phonological development of Korean: A case study. In: Kuno, S., Lee, I.-H., Whitman, J., Maling, J., Kang, Y.-S., Sells, P., Sohn, H.-S. (eds.), *Harvard Studies in Korean Linguistics XI*. Seoul: Hanshin Publishing Company, 17–33.

Kim, M. 2004. Correlation between VOT and F0 in the perception of Korean stops and affricates. *INTERSPEECH-2004*. 49–52.

Kim, M.-R., Lotto, A. 2002. An investigation of acoustic characteristics of Korean stops produced by non-heritage learners. In: Ree, J. (ed.), *The Korean Language in America*, Vol. 7. American Association of Teachers of Korean, 177–188.

Lee, S., Iverson, G. 2008. Development of stop consonants in Korean. *Korean Linguistics* 14. 21–39.

Major, R. 1996. L2 acquisition, L1 loss, and the critical period hypothesis. In: James, A., Leather, J. (eds.), *Second-Language Speech: Structure and Process*. Berlin: Mouton de Gruyter, 147–159.

Schmidt, A. 2007. Cross-language consonant identification: English and Korean. In: Bohn, O.-S., Munro, M. (eds.), *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*. Amsterdam: John Benjamins, 185–200.

Shin, E. 2007. How do non-heritage students learn to make the three-way contrast of Korean stops? In: Wang, H.-S. (ed.), *The Korean Language in America*, Vol. 12. American Association of Teachers of Korean, 85–105.

NOTES

* This research is reported at greater length in Chang (in press). The author gratefully acknowledges the financial support of the National Science Foundation, the UC Berkeley Center for Korean Studies, and the UC Berkeley Department of Linguistics; the helpful comments of Keith Johnson, the UC Berkeley Phonology Lab, and the audience at HISOKL 2009; the logistical assistance of the Fulbright Korean-American Educational Commission; and the dedication of all participants who completed the study.

Allophony in the L2: Can native-like phonetic output be achieved?

Will Dalton

McGill University

william.dalton@mail.mcgill.ca

ABSTRACT

Whereas the second language acquisition of non-native phonemes has been the subject of much research, relatively little attention has been paid to the acquisition of non-contrastive sounds, or allophones. Anecdotal evidence suggests that it is difficult to reach near-native levels of proficiency with the respect to the production and discrimination of non-contrastive phones. The foreign accent exhibited by many second language learners is due in part to failure to produce the correct context-dependent phone. The primary goal of this study is to determine whether learners are capable of acquiring non-native allophonic alternations and, if so, whether allophone discrimination is correlated with level of proficiency in the L2. This paper presents preliminary results from a discrimination task designed to test whether English-speaking learners of Quebec French are sensitive to the contextual information that determines which allophone of the phoneme /i/ is correct in a given context.

Keywords: Allophony, English, Quebec French, L2

1. INTRODUCTION

This research investigates the second language of acquisition of allophonic alternations to determine (a) if allophones *can* be acquired (i.e. whether learners can acquire the context-dependent distribution of non-contrastive speech sounds), and (b) if the ability to discriminate non-native allophones becomes more native-like with increasing proficiency or exposure to the L2. To do this, I investigate the acquisition of allophones in Quebec French by native speakers of English. The phones under investigation are the high, front, unrounded vowels (tense [i] and lax [ɪ]), which are contrastive (phonemic) in the L1 and non-contrastive (allophonic) in the L2.

2. ALLOPHONY IN QUEBEC FRENCH

The principal task of the learner is to learn the correct distribution of allophones. This is expected to be difficult, mainly because there is little motivation for the learner to get it right. Producing the wrong phone in a given context will not likely lead to communication difficulties on the part of the native speaker interlocutor. For example, regardless of whether the learner utters *pet*[i]*te* or *pet*[ɪ]*te* ('small' (fem.)), the native speaker will easily comprehend the meaning. This is especially true in this case, in which the L2 is Quebec French; while the latter form uses the correct allophone in this context (high vowels become lax in closed syllables), the former is consistent with other varieties of French, including standard European French.

Another factor contributing to the difficulty of acquiring this allophonic alternation is the complex distribution of high, front, unrounded vowel allophones in Quebec French. Generally, tense [i] appears in open syllables, while lax [ɪ] appears in closed syllables. However, two phonological processes in Quebec French complicate this distribution: word-final lengthening and vowel (laxing) harmony. When in word-final position, the voiced fricatives, [v, z, ʒ, ʁ] cause the preceding vowel to lengthen (Walker 1984). If the vowel is high, then the surface vowel will also be tense, in addition to long, e.g. [riːv], *[rɪv] ('shore). Therefore, a subset of closed syllables will contain a tense [i]. Additionally, in cases where the final syllable is a closed syllable with a lax [ɪ], the laxness spreads to the preceding syllable, if that syllable also contains a high front vowel, e.g. [pɪsɪn], *[pisɪn] ('pool').[1] As a result of this process, a subset of open syllables will contain a lax [ɪ] (Poliquin 2006).

Based on the complexity of the distribution of high front vowel allophones in Quebec French, as well as the surface distribution of high front vowels in English, several predictions can be made with respect to expected performance in this task. Learners may exhibit behaviour that is incorrect for Quebec French, but consistent with other dialects of French, namely the standard variety spoken in France. For example, learners may reject lax [ɪ] in closed syllables altogether and accept tense [i] across the board. This behaviour may be due to the influence of non-native or European French instructors or significant exposure to varieties of French spoken outside of Quebec. Transfer from the L1 may also affect learners' performance in this experiment. Since lax [ɪ] commonly appears in unstressed syllables in English (e.g. r[ɪ]diculous) for many speakers, French learners at lower proficiency levels (or who have had relatively little exposure to Quebec French) may incorrectly judge lax vowels in unstressed open syllables in French as appropriate. Overgeneralization may also play a role in French learners' judgments. Recall that in the majority of cases, tense [i] appears in open syllables, while lax [ɪ] appears in closed syllables. Learners may incorrectly apply this to lengthening contexts and vowel harmony contexts, leading to the acceptance of lax and tense vowels, respectively, in these contexts. Since performance is likely tied to the amount of exposure to Quebec French, which I assume to be correlated with level of proficiency, I expect that performance on the appropriateness judgment task will increase proportionally with increasing proficiency level, with the results from near-native French learners converging on those of the native speaker controls. While individual learners at the lowest level of proficiency (intermediate, in this experiment) are expected to perform either at chance (indicating that they have not yet acquired the alternation in a given context) or approaching zero (behaviour consistent with European French), I predict that, as a group, intermediate French learners will perform at or near chance level in their judgments in the discrimination task.

3. EXPERIMENT

3.1. Methods

3.1.1. Stimuli

The stimuli presented in this task consisted of pairs of words that differed only in the quality (tense or lax) of the target vowel (e.g. [anime] ~ [anɪme]), with the two variants separated by 0.5 seconds of silence. Each item consisted of two instances of the same pair, separated by 1.5 seconds of silence, so that each pair was heard twice. Participants heard a total of 120 test items, divided evenly into six conditions by target vowel context: Stressed closed syllables (Condition1), unstressed closed syllables (Condition 2), stressed open syllables (Condition 3), unstressed open syllables (Condition 4), word-final lengthening contexts (Condition 5) and vowel harmony contexts (Condition 6). All items in Condition 5 have a stressed target vowel, due to word-final stress in French. On the other hand, the target vowel is unstressed in all Condition 6 items, since vowel harmony targets the penult. Half of the items within each condition were pairs of nonce words that were designed to sound French (in terms of phonotactics, segmental inventory and stress). Stimuli were presented using Microsoft PowerPoint.

Stimuli were recorded by a female native speaker of Quebec French in a soundproof booth, using a head-mounted microphone. Recordings were made digitally using the Audacity software program at a sampling rate of 44 kHz. The native speaker was a trained linguist and attempted to control for pitch and stress as best as possible in order to ensure that both properties did not vary within a given item.

3.1.2. Participants

There were three groups of French learners: intermediate (N= 5), advanced (N=17) and near-native (N=4); as well as a group of native speakers of Quebec French (N=3) to serve as controls. Participants were recruited in Montreal; most were students or recent graduates of McGill University. Participants were grouped into proficiency levels based on self-rated scores on a four point scale (beginner, intermediate, advanced, near-native) for the following factors: reading, writing, fluency, spoken accent and listening. Scores for the last three factors were given a greater weighting, as these aspects of language ability are more relevant for this experiment. French learners received \$15 as compensation for their time, while native speaker controls received \$10.

3.1.3. Procedure

Participants were tested individually. Each was seated in a soundproof booth and listened to the stimuli through headphones. For each item, participants indicated on an answer sheet whether they perceived the first or second word in each pair to be more appropriate for Quebec French. There was no limit placed on the amount of time to judge each item; participants controlled the pace of the experiment by clicking an icon on the computer screen to advance to the next item.

4. RESULTS

Within-subjects results are demonstrably non-random, with participants exhibiting three types of responses for a given condition: (1) correct judgments at or near 100%, indicating successful acquisition of the Quebec French allophone in this context; (2) correct judgments at or near 0%, indicating that the participants' judgments are incorrect for Quebec French, but consistent with other varieties of French; and (3) correct judgments at or near 50% (chance level), indicating that the correct allophone for this context has not been acquired, but neither has it been rejected in favour of the incorrect allophone.

Table 1 below summarizes the response patterns exhibited by individual subjects. A checkmark (√) indicates that the subject's correct responses for this condition (i.e. acceptance of the correct Quebec French allophone) are greater than or equal to 75%.

The first pattern, which is exhibited by seven subjects, is across-the-board rejection of lax [ɪ]. Participants showing this behaviour correctly accept tense [i] in Conditions 3, 4 and 5 (open syllables and lengthening contexts), but they also incorrectly judge tense [i] to be acceptable in Conditions 1, 2 and 6 (closed syllables and vowel harmony contexts).

The inverse pattern, which can be seen in the judgments of four subjects, is also common. In this pattern, participants correctly accept lax [ɪ] in open syllables and vowel harmony contexts, which is appropriate for Quebec French. However, they also accept this vowel in open syllables and lengthening contexts in a clear case of overgeneralization.

A third pattern, exhibited by three subjects, suggests that transfer from the L1 plays a role in their judgments. These subjects correctly accept lax vowel [ɪ] in stressed closed syllables, but judge the same vowel in unstressed closed syllables as inappropriate. Recall that English allows [ɪ] in unstressed, reduced syllables. This behaviour suggests that the L1 distribution of high, front vowels is interfering with the acquisition of the correct vowel in closed syllable contexts.

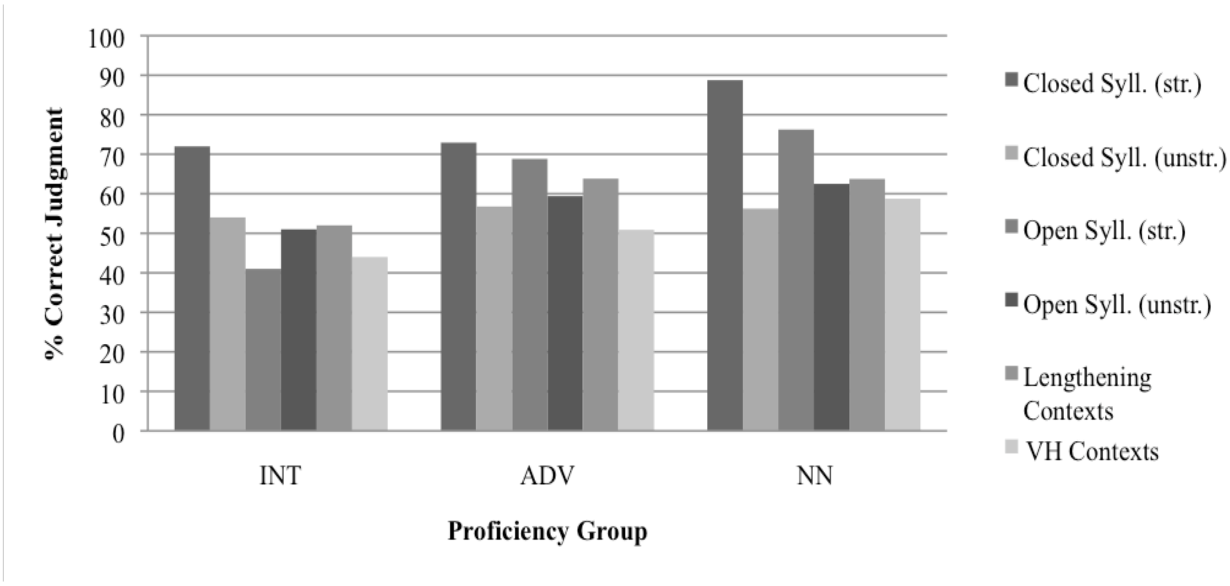
Table 1: Summary of individual results by condition

Subject	C1: Closed syll. (str.)	C2: Closed syll. (unstr.)	C3: Open syll. (str.)	C4: Open syll. (unstr.)	C5: Length. contexts	C6: VH contexts
E1, E11, E17, E18, E20, E23, E25			√	√	√	
E10, E16, E19, E26	√	√				√
E4, E8, E24	√					
E3, E7			√		√	
E14	√		√	√	√	
E21		√	√		√	
E2		√	√		√	

E5	√	√		
E6	√			√
E9	√	√	√	
E12	√	√	√	√
E13, E15				

Overall, clear patterns can be seen in the results of individual participants. When grouped by proficiency, however, no significant trend can be found, suggesting that the acquisition of Quebec French allophones may not be correlated with level of proficiency. Figure 1 shows the results of the intermediate, advanced and near native groups on all six conditions[2]. The near-native group clearly outperforms both the intermediate and advanced groups on Conditions 1 and 3 in correct judgments, suggesting that the general pattern of tense [i] in open syllables and lax [ɪ] in closed syllables is reinforced with increasing exposure to Quebec French. However, poor performance on Conditions 2 and 4 across all groups suggest that this pattern is harder to recognize in unstressed syllables. Inconsistent judgments on Condition 5 indicate that learners of all proficiency groups have not yet acquired the correct allophone for lengthening contexts. Participants also yield inconsistent judgements in vowel harmony contexts, although this is less surprising, given that laxing harmony is only semi-productive in Quebec French.

Figure 1: % correct judgments by proficiency level



5. DISCUSSION

The patterns that emerge from the results of individual participants confirm the predictions made above. A substantial portion of the participants display behaviour that is inconsistent with Quebec French, but consistent with most other dialects of French, including standard European French. These participants, independent of proficiency level, judge tense [i] to be acceptable in all contexts, including in closed syllables and vowel harmony contexts which, in Quebec French, require tense [ɪ]. There are a number of possible explanations for this behaviour: they may simply have not yet acquired the context-dependent distribution of allophones appropriate for the L2, or they may have decided, consciously or unconsciously, to conform to the phonology of European French. Even though participants were explicitly instructed to make judgements that correspond to Quebec French, they had ample time to make metalinguistic judgments. Sociolinguistic factors may be at work here; the use of lax [ɪ] in closed syllables is one of the more obvious markers of the

Quebec dialect of French, which some learners (and even some native speakers) view as substandard. This opinion may influence their judgments, leading to the avoidance of lax [ɪ] altogether. On the other hand, the participants' linguistic history may also be an important factor. Non-native instructors, or instructors who speak a non-Quebec dialect of French, may influence the behaviour of learners at the early stage of acquisition, especially if classroom learning constituted the majority of the learners' early exposure to French. It may also be the case that some learners have had substantial exposure to other dialects of French, although it should be noted that pre-screening during the recruitment stage specifically sought to eliminate these as potential subjects.

The opposite pattern as exhibited by four of the participants clearly indicates that correct usage of the lax [ɪ] allophone is indeed acquirable, although it is also subject to overgeneralization, as these participants judge the lax vowel to be appropriate in all contexts.

The fact that none of the subjects, not even the most advanced, or those with the most exposure to Quebec French, has fully acquired the distribution of allophones is worthy of note. Appropriate judgments reach significantly high levels for only two subjects (E12 and E14) in four of six conditions. Failure to fully acquire the distribution of allophones can be taken as an indication of the difficulty of the task that learners face in the acquisition of non-contrastive sounds.

6. CONCLUSION

This paper set out to answer the question of whether allophonic alternations can, indeed, be acquired; a question which has not yet been answered satisfactorily. Whereas extensive research has shown that non-native contrastive sounds can be acquired, there have been few principled investigations into the acquisition of non-contrastive sounds. The results of this experiment suggest that second language learners are capable of learning allophonic distributions. However, the fact that no subject is able to perform at native-like levels in all contexts confirms the expectation that the task of learning a complex allophone distribution is particularly challenging. Even the most proficient learners were only able to partially exploit this information in their judgments. Furthermore, sensitivity to contextual information does not seem to correlate with level of proficiency. Although participants in the near-native group performed slightly better than those of lower proficiency, no clear trend emerged. It should be noted, however, that more data is needed from the intermediate and near-native groups before a definitive conclusion is reached.

7. REFERENCES

Walker, D. 1984. *The pronunciation of Canadian French*. Ottawa: University of Ottawa Press.

Poliquin, G. 2006. *Canadian French Vowel Harmony*. Ph.D. Dissertation, Harvard University.

NOTES

[1] It should be noted that vowel harmony is not fully productive in Quebec French. Rather, it is highly variable and may be subject to frequency effects, with only higher frequency words undergoing harmony.

[2] Statistical tests of significance have not yet been performed. However, due to the clear lack of any trend, differences may be treated as non-significant.

The COREIL corpus: a learner corpus designed for studying phrasal phonology and intonation

Elisabeth DELAIS-ROUSSARIE & Hi-Yon YOO°*

*CNRS – UMR 7110/Laboratoire de Linguistique Formelle – Université Paris-Diderot
° Université Paris-Diderot – UMR 7110/ Laboratoire de Linguistique Formelle
elisabeth.roussarie@wanadoo.fr & yoo@linguist.jussieu.fr

ABSTRACT

Studies in L2 intonation and phrasal phonology are interesting not only to understand L2 acquisition, but also to get better insights on the phonology of the target language itself. Indeed, clear descriptions are still missing for many intonational and phrasal phenomena; and studying learner's speech may give perspectives to the analysis of phenomena that have remained unnoticed up to know (*e.g.* grammatical and prosodic constraints occurring in case of self-repairs, phonological status of some prosodic events, etc.). The examination of well-built data is needed to work in such directions.

The aim of the contribution is to present the COREIL corpus, an electronic oral learner corpus that has been designed to study the acquisition of phrasal phonology and intonation in French and English as a foreign language. The data collection protocol has been thought in order i) to carry research on the acquisition of suprasegmental phenomena, ii) to compare the acquisition processes along several dimensions (L1 vs. L2, differences among the learners L1, etc.). Here focus will be mostly given on the data collection protocol and the annotation schema.

Keywords: language resources, prosody, corpus annotation.

1. INTRODUCTION

Since several years, numerous researches on second language acquisition are based on the study of large corpora (cf. among others, Granger 2003, Hawkins & Buttery 2009). As a matter of fact, the use of large corpora has allowed a better evaluation of possible correlations between the learner's L1, his grammatical competence and his proficiency level in L2 (which is established among the Common European Framework of Reference for Languages). For instance, in a project such as English Profile (Hawkins & Buttery 2009), corpus-based studies are used to determine how several morpho-syntactic phenomena are acquired in English as a foreign language. However, most studies focus on the morphological and syntactic competence, and very few deal with phonological phenomena. One of the reasons of that is the difficulty to find oral corpora (see however Milde & Gut 2002 and Tortel 2008). In order to answer to this lack of data, we have elaborated a protocol and developed the COREIL corpus, a learner corpus designed to study the acquisition of suprasegmental phenomena in a foreign language.

The first section of the paper focuses on the theoretical assumptions used to build the COREIL corpus. In the second part, we describe the experimental protocol, paying a particular attention to the selection of the speakers and the construction of the different recorded tasks (reading, monologue oral production, guided conversation, etc.). The third part deals with the conventions that were used for annotating the data.

2. THEORETICAL BACKGROUNDS

2.1. Theoretical assumptions and the acquisition of L2 phonology

Many studies on the acquisition of L2 phonology rely essentially on the idea that language transfer (positive or negative) from the learners' L1 to the target language is crucial in L2 acquisition process (see, for instance, notions such as phonological deafness, but also pedagogical practices using contrastive analysis between L1 and L2). Corpus-based studies make it possible to test the validity of such hypotheses. Moreover, oral data could allow evaluating in which sense the acquisition order could be similar for children in L1 and learners in L2, and that, independently to the learner L1.

Moreover, most studies on the acquisition of L2 phonology have focused on segmental phenomena (Rasier, L. & Hiligsmann, P. 2007). Even though such studies are of great interest, working on the L2 acquisition of suprasegmental phenomena is important. Many questions need to be answered:

- Are there any differences in the acquisition of L1 and L2 prosody? If so, what are they?
- Does transfer play a fundamental role in the acquisition of prosodic phenomena?
- Are the given answers to the previous questions valid for all domains and levels of prosody (accentuation, intonation, phrasing, rhythm, etc.)?
- Does the acquisition of segmental phenomena differ from the one of suprasegmentals?
- What comes first in the acquisition of prosodic phenomena such as rhythm that imply both phonetic and phonological competences? Phonetics or phonology?

The COREIL corpus was conceived to answer these various questions. Any presupposition on the role of transfer in the acquisition of the phonology of L2 was avoided. Moreover, the recording protocol presented in section 3 was thought in order to allow comparisons along different dimensions: between learners of different L1, between L1 and L2, and so on.

2.2. Assumptions and data collection

Beyond the theoretical questions related to L2 acquisition, some assumptions on corpus design have been formulated for constructing the COREIL corpus. For instance, the corpus has been built following the AGILE methodology, which is inspired by software development (Voormann & Gut 2008). It is well known that collecting data is a long and fastidious activity, for it does not consist only in collecting data, but also in working on their format and annotation schema. Linguistic studies can start only once a “correct” corpus is achieved. But, when researchers start to analyze the data, they have to face new questions:

- Are the collected data representative enough for the planned investigations?
- Will complementary data give a better answer to the questions under investigation?
- Is the annotation schema satisfactory enough from a theoretical point of view (few presuppositions, facility to make a query, good adequacy between the annotators, and so on)?

When a researcher elaborates a corpus and carries out research on it, it is not rare to hear her saying that she would proceed differently, is she had to do it again! In order to avoid this drawback, we adopted the AGILE approach for designing the COREIL corpus. It offers many advantages: i) it allows working on a part of the collected data, even though the data collection process has not been finished yet; ii) it is possible to integrate supplementary data or tasks to the existing protocol and annotation schema without loosing the homogeneity of the existing corpus.

3. DATA COLLECTION PROTOCOL

3.1. Speakers selection

The COREIL corpus consists of texts produced by different categories of speakers: L2 adult learners, children up to 7 years of age in their L1, and adults in their L1 (as a matter of fact, the target language for L2 learners). We thought it was important to have L1 data produced by adults and children in order to have a point of comparison and to be able to evaluate the weight of transfer in the acquisition process, as well as the differences in the acquisition of phonology (and more specifically, prosody) in L1 and in L2. Moreover, for L2 learners, two parameters are strictly controlled: the choice of the learner L1 and the linguistic proficiency level.

3.1.1. Choice of the learner L1

For the French data, four distinct L1 were chosen: English, Spanish, Mandarin Chinese and Arabic. For the moment, no constraint on the learner L1 has been formulated for English. Since the project is done in France, the majority of the learners of English have French as their first language.

Even though we are aware of differences among the existing varieties of a language, we did not formulate any restriction so far. For instance, we accepted American English speakers, as well as British English one as samples of English speakers. If the analysis of the data shows that dialectal differences have a huge influence in the acquisition of L2 prosody, each variety will be treated separately, as a different language.

3.1.2. Age the speakers

No constraint was retained on the age of the L2 learner. But, as far as the children population is concerned, we consider that it is important to have both productions from children acquiring an L2 while the L1 acquisition is not yet achieved and productions from older children whose L1 acquisition is achieved. Indeed, it may be important to have access to data with a large panel of age representativity. It will allow working on questions such as the critical period, and evaluating the weight of the parameters that are at stake in L2 acquisition. Moreover, note that children and youngsters represent an interesting population to do longitudinal studies since the learning process of the pupils of one same class varies, and it is possible to follow their acquisition during the whole year (or even through several years).

3.1.3. Proficiency level in L2

Another parameter taken into account in the corpus construction was the L2 proficiency level. In second language acquisition, the proficiency level is one of the most important parameter if one wants to investigate on acquisition order, and to make comparison with L1 acquisition process. However, this parameter is one of the most difficult to evaluate, especially for adult speakers. For this corpus, we focused on A2 and B1 learners according to the CEFR evaluation grid (the A1 learners were put aside since the different tasks would be to difficult to realise for them, mostly because of the lack of grammatical and lexical knowledge).

For school children, we refer to the school program given by the ministry of Education to evaluate the level. In fact, it is easy to find in the programs what are the expectations. Evaluating the adult population was more delicate for at least two reasons. First, each of them had a different language learning background. It is thus not reliable enough to consider years (or hours) spent to study the target language (some persons may have interrupted their studies, or learned in different countries). Second, auto-evaluation is difficult because under or over-evaluation obtained, depending on the subjects. Two different methods were used in parallel:

- For university students, we used the scores they obtained at the admission test they had to perform for their foreign language class enrolment. We consider these tests reliable enough and their scores easy to use because they are generally formulated according to the CEFR.
- An auto-evaluation questionnaire was proposed to each adult before the first recording session. The questions attempt to determine what the subject can or cannot achieve in the target language. We tested both oral and written competence, in production and comprehension. The questions were developed in compliance with the language proficiency grid established for each level.

3.1.4. Learner profile

A linguistic profile was established for each subject. Different information was gathered concerning the subject L1 (monolingual or bilingual L1 acquisition, L1 spoken variety etc.), his L2 proficiency (level, result for the auto-evaluation etc.), the other spoken foreign languages, and the time he spent in countries where the target language is spoken (date, length of stay, etc.). Complete linguistic profiles can then be used to study the acquisition process and to establish comparisons.

3.2. The production tasks

As communicative and grammatical competences may differ according to the communicative tasks the speaker has to perform, the corpus is composed of productions elicited in various ways. The different tasks are grouped according to three main criteria: task typology, level of proficiency and age required to perform the task. Having a modular protocol and using tasks present several advantages such as the ability to add or modify tasks in order to get more data or to adapt according to the age and proficiency level of the speaker. At last, in longitudinal studies, comparable but different tasks can be proposed to the learners during each recording session.

For each subject, five distinct tasks are usually used. They may be classified in three main categories: reading task, monologue oral production task and conversation production task. For the reading task, the texts are either excerpts of the EUROM 1 corpus (Chan et ali. 1995) or small dialogues. The latter texts, because of their form, do request the use of different intonational patterns – because of the presence of exclamatives, questions, irony, etc.–, which is of great interest for the study of intonation.

(1) Sample of a dialogue to read

Unfortunately, I am early; too bad for me. So I get myself a newspaper and I take a seat inside the café. The barmaid comes up to me :
- « Good morning, Madam, what would you like to have ?
- One tea please.
- With milk ?
-Yes please.»

For the monologue oral production task, four different types of activities are usually given. The first one consists of a simple narration, which can take different forms depending on the age and the proficiency level of the subject. He can be asked i) to repeat a story that was just read to him or ii) to narrate a story presented on several pictures (cf. (2)). Once again, the pictures can be more or less complicated in order to adapt to the speaker age or level.

(2) Story in pictures presented to the subject

The second type of activities consists of describing a picture that belongs to one of two distinct categories. In the first case, the subject is asked to describe the content of a picture that represents something static such as a room, a painting etc. In the second one, he has to describe a picture where people do practice an activity (sports, shopping in a market etc.). The two distinct tasks require the use of specific linguistic forms (different verbs and grammatical constructions, etc.).

For the guided conversation tasks, the subject has to achieve two distinct activities. In the first one, he has to answer questions that are asked by the investigator. The productions obtained are interesting since they are produced in a very natural way.

(3) Oral production obtained by asking questions to the subject

a. in French

euh je suis né à Phoenix (.) Phoenix Arizona aux Etats Unis. C' est une grande ville il y a beaucoup de gens qui habitent là+bas. C' est peut être euh (.) je ne sais pas exactement le numero de gens qui habitent là mais peut être maintenant plus d' une million de personnes habitent à Phoenix.

b. in English

I go to primary school er in Evreux er in xx [/?] in the department [x 2] of Eure. um it was er in the countryside er my college and [x 2] after it was in [x 2] the city but my primary school was in [x 2] the countryside. So it was great because it was a little school with er everywhere er known [/?] knew [/?] everyone knew everyone sorry.

As interesting as these data are, they display some linguistic limitations. In most of the cases, the subject does produce assertions. In order to get some questions, the subject was asked to reverse the roles. He had to ask a series of questions in order to acknowledge the identity of the interlocutor (age, hobbies etc.).

3.3. Experimental settings and data recording tools

The data collection protocol was designed in order to put together different comparable texts that were elicited in taking into account parameters such as the age or the proficiency level of the subject. The modularity also allows distinguishing two experimental procedures. The subjects can be recorded in a single session, the data being used to study the weight of L1 in the L2 acquisition, for instance. But, the protocol can also be used for longitudinal studies where the same subject is recorded several times during a year in order to evaluate how the acquisition of specific prosodic phenomena (in L1 as well as in L2) occurs.

4. TRANSCRIPTION AND ANNOTATION OF THE DATA

4.1. Theoretical backgrounds

The annotation schema used for the COREIL corpus differs greatly from the one generally retained for learners' corpora. In most cases, morpho-syntactic annotation consists in coding the errors by comparing the productions obtained to correct forms in the target language, and this without taking into account the production context (see among others Lüdeling 2008). Such an approach is questionable for at least two reasons: i) by encoding only errors, the fact that the learner's language works as a system with its own rules is forgotten; ii) it is difficult to evaluate the acceptability of an utterance when treating it out of its context.

The data of this corpus have been annotated in such a way as to minimise interpretation. We actually believe that the interpreting task has to be left to the people that undergo the analysis of specific linguistic phenomena.

4.2. Orthographic transcription

The whole corpus has been orthographically transcribed, the text being aligned with the signal at the level of the utterance. The transcription conventions were elaborated in compliance with the recommendations of the TEI and EAGLES. As a matter of fact, the latter were slightly modified in order to take into account the specificity of learner productions: when the learner produces illicit forms in the target language, they are transcribed using the orthographic rules of the target language (depending on the realization obtained, *I bought a new car* could be transcribed *I buyed a new car*). When a learner produces an incorrect form from the phonetic or phonological point of view, two cases are distinguished: i) when the produced form results from a phonetic or phonological error, it is transcribed in compliance with standard orthography (for instance, when the learner pronounces for *conscious* [kɔn'ʃjus] instead of ['kɒnʃəs], it was transcribed *conscious*); ii) when the mispronunciation may be due to morpho-syntactic difficulties, multi-transcription is considered as obligatory.

4.3. Linguistic annotation

Each linguistic annotation is done on a separate tier, which allows adding further annotations when needed. At the moment, two distinct types of annotation are under development: part of speech tagging and prosodic annotation.

4.3.1. POS Tagging

POS-tagging is done with the applications Mor and Post integrated to the CLAN program (see MacWhinney 2000). Note however that specific dictionaries and tags may be developed in order to improve the accuracy of the annotation. At present, the POS-tagging is comparable of what is shown in (4) et (5).

(4) Tagging and English data

***APO:** er what kind of courses do you have?
%mor: fil|er pro:wh|what n|kind prep|of n|course-PL aux|dopro|you v|have

(5) Tagging and French data

*APO: euh je fais musique avec mon ordinateur et aussi euh je [x 2] j'écris euh j' écris des articles .
%mor: fil|euh pro:subj|je v:mdllex|faire-PRES- n|musique prep|avec det:poss|mon n|ordinateur&_ conj|et conj|aussi fil|euh pro:subj|je v|écrire-PRES fil|euh det|des&PL n|article.

4.3.2. Prosodic annotation

Annotating the prosodic features of learner productions is difficult. At the moment, there is no widely accepted transcription system that allows transcribing prosodic events without making assumptions on their phonetic or phonological status. Work is currently in progress in order to adapt the existing systems and to get a more acceptable one. To achieve this task, we pay great attention to three of the existing systems: the IVIE system (Grabe *et al* 2001), the momel-intsint model (Hirst 2005) and the system developed by Piet Mertens (Mertens 1990). As a matter of fact, these systems offer some advantages that are worth considering:

- IVIE does not make any strong assumptions on a phonetic or phonological status of the transcribed events. The grammatical analysis is thus achieved once the transcription is done;
- By using automatic procedures to encode tonal information, Momel-Intsint may be considered more robust;
- By taking the syllable as basic unit to encode the prosodic information present in the signal, Mertens' model is able to represent metrical as well as tonal phenomena.

5. CONCLUSION

The COREIL corpus, presented here, has been developed to work on the acquisition of L2 prosody. The protocol has designed to allow progressive data gathering. Moreover, tasks can be modulated to adapt to the age and the level of the subjects. The main idea behind this corpus is to build experimentations on acquisition by crossing different parameters: age, level, learner L1, etc. It enables to evaluate the weight and role of the learner L1 as well as the differences and/or similarities between L1 and L2 acquisition. The annotation schema has been thought in such a way as to facilitate the reuse of the data. Moreover any type of interpretation is avoided, limiting any linking between learner productions and the target language.

6. REFERENCES

Granger, S. 2003. The International Corpus of Learner English: A New Resource for Foreign Language Learning and Teaching and Second Language Acquisition Research. *TESOL 37 (3)*. 538-546.

Grabe, E. & B. Post. 2004. Intonational variation in the British Isles. In G. Sampson and D. McCarthy (eds.), *Corpus Linguistics: Readings in a Widening Discipline*. London and New York: Continuum International, 474-481.

Hawkins, J.A. & P. Buttery. 2009. Using learner language from corpora to profile levels of proficiency. *Studies in Language Testing*. Cambridge University Press.

Hirst, D.J. 2005. Forms and function in the representation of Speech Prosody. *Speech Communication* 46. 334-347.

Lüdeling. A. 2008. Mehrdeutigkeiten und Kategorisierung. Probleme bei der Annotation von Lernerkorpora. In: Walter et al. (eds) *Fortgeschrittene Lernervarietäten.* Niemeyer. 119-140.

MacWhinney, B. 2000. *The CHILDES Project: Tools for Analyzing Talk*. Volume 1: Transcription format and programs. Volume 2: The Database. Mahwah, NJ: Lawrence Erlbaum Associates. (These are links to recent electronic revisions).

Mertens, P. 1990. Intonation. In *Le français parlé : études grammaticales*. C. Blanche-Benveniste, M. Bilger, C. Rouget & K. Van de Eynde (eds). Paris: Editions du CNRS. 159-176.

Milde, J.-T. & U. Gut. 2002. A prosodic corpus of non-native speech. *Proceedings of the Speech Prosody 2002*, pp. 503-506.

Tortel, A. 2008. ANGLISH : base de données comparative L1 & L2 de l'anglais lu, répété, parlé. *TIPA* 27 : 111-122.

Rasier, L. & P. Hiligsmann. 2007. Prosodic transfer from L1 to L2. Theoretical and methodological issues. *Nouveaux cahiers de linguistique française 28*.

Voormann, H. & U. Gut. 2008. Agile Corpus Creation. *Corpus Linguistics and Linguistic Theory 4 (2)* : 235-251.

Speaking like the locals - the acquisition of two local variants in the spoken English of native Polish speakers living in Manchester

Rob Drummond

The University of Manchester
rob.drummond@manchester.ac.uk

ABSTRACT

This paper examines the acquisition of local variants of two linguistic features (the STRUT vowel and t-glottaling) in the speech of native Polish speakers living in Manchester, UK. The local variants of both features differ from the pedagogical model that the Polish participants will have been exposed to when learning English in Poland. In both cases there is evidence of acquisition. Multiple regression analyses using Rbrul (Johnson 2008) show length of residence to be significant in the acquisition of both features, along with current level of English and gender for t-glottaling, and attitude towards living in Manchester for STRUT. The relevance of each if these factors is discussed, along with a suggestion for the most appropriate way to view the gender effect.

Keywords: L2 dialect acquisition, L2 phonological acquisition, variation, gender.

1. INTRODUCTION

Although there is an established body of research into the acquisition of a second dialect in a person's first language (e.g. Chambers (1992); Payne (1980); Tagliamonte and Molfenter (2007); Munro et al. (1999)), research into the acquisition of a second dialect in a person's second language is less common (e.g. Baker (2008); Fox and McGory (2007); Sharma (2005)). This paper, along with the wider study of which it is a part, aims to develop further the link between the two relevant areas of linguistics (SLA and variationist sociolinguistics) by exploring the process whereby native Polish speakers living in Manchester begin to acquire phonological features of the local accent. More specifically, it reports on the findings of two strands of investigation, namely: 1) to establish whether certain features of the pronunciation of native Polish speakers do indeed change to become more like the local variants; and 2) to establish what factors might influence the degree of change. Factors under consideration include length of residence (LOR), gender, attitude towards aspects of living in Manchester, use of L1 and L2, and Level of English (LoE).

By employing variationist methods and principles in the area of SLA certain assumptions are made, primarily the idea that variation in L2 speech can indeed be viewed as being systematic, in a similar way to variation in L1 speech. There is not the space to enter the debate here, but it should be made clear that the research has been carried out having accepted this central assumption.

2. BACKGROUND

The study is concerned with the speech of native Polish speakers living in Manchester, a traditionally industrial city in the North West of England with a population of 464,200 (Manchester City Council 2010) There are a large number of Polish nationals in the city, a result of the recent wave of immigration from the eight Central and Eastern European countries (A8) acceding to the EU in May 2004. Unlike some nationals from other A8 countries, Poles have moved to most areas of the UK, but the North West of England is particularly popular (Bauere et al. 2007)

The two features under investigation are the vowel sound in STRUT words and t-glottaling. The STRUT vowel is a highly salient feature of Northern British English (NBrE), and one which generally shows little, if any, contrast with the FOOT vowel in the speech of native English speakers in the Manchester area. Those native speakers who do show some contrast tend to produce a schwa-like sound for STRUT in some contexts, but rarely produce anything close to RP STRUT. In contrast, the Polish speakers being investigated have all been exposed to a pedagogical model involving something similar to RP STRUT, and the Northern English variety represents a deviation from this model. A change in the quality of this vowel in the speech of the Polish participants from RP STRUT [ɐ] towards Northern [ʊ] is therefore seen as evidence of external influences on pronunciation, i.e. accent change.

T-glottaling, on the other hand, is not restricted to NBrE. It is a feature which is spreading across the UK, increasingly losing its connection to urban areas and non-prestige varieties of English (Milroy et al. 1994); (Williams and Kerswill 1999). Neither is it a feature that is highly salient, meaning that it tends to exist below the level of awareness, especially in the linguistic environments under investigation here. Again, for most participants the local pattern of t-glottaling represents a deviation from any pedagogical model they will have been exposed to in Poland. They are likely to have heard natural native speaker speech with a degree of t-glottaling, and some may even have been encouraged to replicate the feature, but generally, t-glottaling is not highlighted as something to reproduce. Indeed, a pilot study showed that in the speech of recently arrived native Polish speakers to Manchester there was a far lower level of t-glottaling than would be expected in the speech of a native speaker. Therefore, an increase in t-glottaling is seen as evidence of external influences on pronunciation.

3. METHOD

3.1. Selecting the participants

Pparticipants were chosen on the basis of certain criteria in order to control some of the external variables. Therefore, each participant a) grew up in Poland and arrived in England as an adult, b) was aged between 18 and 40, c) had some knowledge of English language before coming to England, and d) ideally had lived nowhere else in the UK apart from in the Manchester area. In addition to these criteria, a spread of LORs and occupations amongst the participants was sought. The final sample consisted of 40 individuals, equally balanced in terms of gender.

3.2. Gathering data

Meetings were arranged with individuals throughout 2009. Although there were other elements to the meetings, all data presented here come from an informal conversation with each participant and a written questionnaire.

The conversation was recorded using a Zoom H2 Handy Recorder placed unobtrusively on a surface near the participant. Recordings were made as .wav files using a 44.1 kHz sampling rate with 16-bit depth, saved onto an SD memory card then transferred onto a PC.

The questionnaire was divided into two main sections, with the first designed to gain socio-demographic information about the individual and the second designed to measure attitudinal and motivational factors. The second section represented the main bulk of the questionnaire, consisting of 42 questions on various aspects of attitude and motivation. The questions were all in the format of a statement followed by a seven point Likert scale, with 'strongly disagree' and 'strongly agree' at numbers 1 and 7, and numbers 2 to 6 remaining unlabelled in between. Multi-item scales, as described in Dörnyei (2002) were employed so that each main area under investigation was covered by more than one question. The internal consistency of the questions was measured using Cronbach's Alpha, and the existence of collinearity amongst the factors was checked by examining tolerance and the Variance Inflation Factor in SPSS. As a result, the following aspects were retained:

- attitude towards Manchester, its people, and living there (ATT);
- awareness of a Manchester accent (AW);
- desire to lose one's Polish accent and sound like NS (not specifically Manchester English) (CHA);
- motivation (both instrumental and integrative) to improve pronunciation. (MOT).

3.3. Identifying and coding the linguistic variables

Starting at the 5-minute point, the first 50 tokens of each variable were coded. Depending on the variation in these first 50, another 50 were gathered where possible. In all, 3375 t-glottaling and 4043 STRUT tokens were analysed. The STRUT variable was analysed both auditorily and acoustically, t-glottaling was analysed auditorily.

3.3.1. T-glottaling

Although acoustic analysis is being used more and more in the analysis of consonantal variables, it was decided that auditory analysis of t-glottaling would be sufficient for the present study. Previous research into glottal variation in /t/ shows a variety of approaches in terms of what constitutes the envelope of variation, with differences existing in both the linguistic environment of /t/ and in the nature of the variants themselves, that is whether to include both glottal replacement and glottal reinforcement. The present study follows the lead of Fabricius (2000), (2002) and Straw and Patrick (2007) by focusing on glottal replacement alone and not on any possible examples of glottal reinforcement.

The present study is concerned with word final /t/ preceded by a vowel (V/t/#), and word medial intervocalic /t/ (V/t/V). Each word final /t/ was categorised as being either pre-consonantal (PreC) *(..tha**t** **c**ountry..)*, pre-vocalic (PreV) *(..tha**t** **i**dea..),* or pre-pausal (PreP) *(..this ca**t**.)*, with the PreC category being further divided into pre-stop (PreS) /p, b, t, d, k, g, m, n/, pre-fricative and affricate (PreF) /f, v, θ, ð, s, z, ʃ, ʒ, tʃ, dʒ/, and pre-approximant (PreA) /r, l, w, j/.

Different contexts were coded for different possible variants on the basis of ease of auditory discrimination in natural speech and the focus of this particular study. The PreC context was coded simply for released [t] or 'other', and all other contexts were coded for glottal replacement or 'other'.

3.3.2. STRUT – Auditory analysis

Every instance of a word which might potentially include the STRUT vowel was identified as a token and the vowel was labelled according to the following system: 0 = RP [ɐ]; 1 = raised RP [ɐ̝]; 2 = schwa [ə]; 3 = lowered NBrE [ʊ̞]; 4 = NBrE [ʊ]; w = weak form. Realisations falling outside the usual NS variation for STRUT were coded separately.

3.3.3. STRUT - Acoustic analysis

Acoustic analysis was used to complement the findings of the auditory analysis. While it must be borne in mind that there is no direct one-to-one relationship between the auditory and acoustic analyses of vowels, particularly when restricting analysis to the first and second formants (Foulkes, Scobbie and Watt Forthcoming), it is often the case that the acoustic analysis helps clarify auditory analysis and vice versa. At the very least, using the two techniques together helps to guard against the incorrect analysis of individual tokens. Therefore, on completion of the auditory analysis for an individual speaker, all clear tokens unaffected by possible co-articulation effects were subjected to acoustic analysis using Praat.

3.4. Statistical analyis

Rbrul (Johnson 2008) was used to carry out mixed-model multiple regression analyses in which the individual speaker was included as a random effect (Johnson 2009).

4. RESULTS

4.1. T-glottaling

Table 2 shows the overall count of tokens collected. What is immediately clear is the almost categorical absence of glottal replacement in word medial position. For this reason, this environment will be excluded from further discussion. The rates of word-final t-glottaling in PreV and PreP environments are relatively low (compared to, for example, Fabricius (2000) which showed rates of 40% and 36% respectively amongst

NSs), yet are strikingly similar to each other. Ostensibly this would appear to suggest an absence of any diffusion pattern between PreV and PreP t-glottaling, with neither environment appearing more likely than the other to favour glottal replacement. However, further analysis presents an alternative.

Table 2: Total distribuiton of t-glottaling tokens for all sepakers

	PreC			**PreV**			**PreP**			
	other	**released [t]**	**total**	**glottal**	**other**	**total**	**glottal**	**other**	**total**	
word final	50.3% (787)	49.6% (779)	1000% (1566)	17.2% (151)	82.8% (726)	100% (877)	17.4% (73)	82.6% (347)	100% (420)	
word medial	×	×	×	0.4% (2)	99.6% (510)	100% (512)	×	×	×	
			1566			**1389**			**420**	**3375**

Although the mean figures for glottal replacement for all speakers are equal across the two environments, at the level of individual speakers there are differences. While there is no preference for one environment over the other amongst the 18 speakers who display glottal replacement in both contexts (PreV and PreP), of the 7 speakers who display glottal replacement in one environment only, this is always PreV. This suggests that for these speakers, glottal replacement cannot exist in PreP environment without first existing in PreV, perhaps suggesting a PreV > PreP pattern of diffusion, a pattern different from those reviewed in Straw and Patrick (2007).

4.1.1. Regression analysis

A multiple regression analysis was carried out with PreV glottal replacement as the dependent variable and with individual speaker as a random effect. The results can be seen in table 3.

Table 3: Rbrul output for glottal replacement in PreV environment for all speakers

	Factor	*Log-odds*	*Tokens*	*Factor weight*
Sex	f	0.567	465	0.638
p = 0.03	m	-0.567	412	0.362
LOR	long (49-72m)	0.979	292	0.727
p = 0.02	medium (25-48m)	0.363	359	0.59
	short (1-24m)	-1.342	226	0.207
Level of Eng	**(scale 1-9)**	+ 1 0.865		
p = 0.00				
Not significant:	*ATT, AW, CHA, MOT, Age, Use of L1/L2*			
Model	deviance 568.614	df 6	intercept -8.908	mean 0.172

Of the three significant independent variables, only LOR is directly relevant to the idea of speakers showing increased glottal replacement as a result of their being in Manchester. The log-odds and factor weights clearly show that the greater the LOR, the greater the likelihood of glottal replacement, but it is the extent of that rise that is important.

Table 4 isolates LOR, showing the very low rate of glottal replacement for those speakers who have been in the country for less than two years. 12 speakers are in this category, 6 of whom showed no glottal replacement at all. The rise between 0-2 years and 2-4 years is sharp, suggesting that two years is the point at which certain L1 norms begin to be more rapidly replaced by L2 variants. However, comparing log-odds suggests that Level of English (LoE) is a more powerful effect than LOR, resulting in the possibility that there could be low-level speakers with long LOR showing little glottal replacement. Indeed this is the case, with the four lowest level speakers showing no glottal replacement, despite LORs of 48, 40, 46 and 65 months.

Table 4: Percentage of glottal replacement in PreV for all speakers arranged by LOR.

LOR	% glottal replacement PreV
0-2 years	3.9%
2-4 years	18.3%
4-6 years	23.6%

The effect of gender is not strong, but it is interesting that females should show a greater level of replacement. This replicates findings amongst NSs in other recent studies e.g. Mathisen (1999); Mees and Collins (1999), and opens up some interesting possibilities with regard to theories on gender and variation. This will be discussed briefly below, but there is not the space on this occasion to give it the attention it deserves.

A regression analysis of the use of released [t] in the PreC environment provided the predictable results of LOR and LoE being significant ($p<0.01$) with higher LOR and LoE reflecting a greater chance of something other than released [t]. Sex was not significant in this case. A further analysis of the following linguistic environment yielded the pattern PreS, PreF, PreA ($p<0.01$), with PreS and PreF slightly favouring something other than released [t] and PreA favouring released [t].

4.2. STRUT

Table 5: Total distribution of STRUT tokens for all speakers

Variant	[ɐ]	[ɐ̝]	[ə]	[ʊ̞]	[ʊ]	other	weak	total
Tokens	1926	298	311	177	58	271	1002	**4043**

Table 5 shows that the standard RP STRUT [ɐ] is by far the most favoured, however, there is some movement towards the local variant, with 58 tokens matching the full NBrE [ʊ].

4.2.1. Regression analysis

A multiple regression analysis was carried out with the target STRUT variant as a continuous dependent variable and with individual speaker as a random effect. The reults can be seen in table 6.

Table 6: Rbrul output for target STRUT variants for all speakers

	Factor	*Units**	*Tokens*
LOR	long (49-72m)	0.662	670
p = 0.00	medium (25-48m)	-0.162	1350
	short (1-24m)	-0.501	750
ATT		+ 1 0.282	
p = 0.00			
Not significant:	*LoE, AW, CHA, MOT, Age, Sex*		
Model	deviance 6643.905	df 6	intercept -0.855

*When the dependent variable is continuous, Rbrul describes effects in units of that variable rather than log-odds

Both significant independent variables are relevant to the speakers being in Manchester. Increased LOR results in greater acquisition of the local variant, as does a positive attitude towards being in Manchester. Once again, when LOR is between 0 and 2 years, there is very little change, with only 8.9% of tokens being something other than [ɐ], compared to 27.3% for 2-4 years and 60.9% for 4-6 years.

Scores on the attitude scale were generally high, with 31 speakers scoring 5 or above on a 1-7 scale. Interestingly, two speakers who had long LORs (59m and 65m) yet minimal movement towards the local STRUT variant had low attitude scores (2 and 4 out of 7).

Two further regression analyses were carried out, one with the dependent variable as accuracy of the vowel (NS target range vs ‘other’) and the other with the dependent variable as quality of the vowel (weak vs full). In both cases the only significant factor was LoE. A higher LoE favoured target vowels over non-target vowels (log-odds = +1 0.332 $p<0.00$) as well as an increased use of weak forms (log-odds = +1 0.256 $p<0.00$).

5. DISCUSSION

Clearly there is evidence that native Polish speakers living in Manchester are, to varying degrees, acquiring local variants of the two features in question. It is unsurprising that LOR plays a significant role in all aspects of this acquisition, although it is interesting that so little acquisition has occurred in speakers with LORs of two years or less. Whether or not this two year mark represents a significant stage after which acquisition is more likely to occur will hopefully become clear as more features are examined.

The existence of a gender effect in the t-glottaling data is very interesting. However, what is unclear at this stage is precisely what aspect of gender is at work here, if indeed it is possible or desirable to separate them. On the one hand, it could be argued that the women are moving towards a supra-local variety along the lines described in Watt and Milroy (1999), especially if we view this supra-local variety not as referring to a region as is usually thought, but as referring to NS patterns of variation as opposed to NNS patterns of variation. In other words, the women are tending to acquire NS patterns, while men are tending to retain NNS patterns. This ties in with a second aspect of gender – the tendency of women to accommodate their speech more than men (Woods 1997). If women are accommodating towards the speech of NSs to a greater degree than men, it follows that they will acquire the variants more readily. The third aspect of gender is simply that women’s social activities and jobs involve contact with a wider range of people than men’s (Holmes 1997) which in this case means more contact with NSs, thus leading to a greater chance of accommodation. Having said this, self-reported use of L1/L2 was not significant in any of the multivariate analyses.

However, the most fruitful approach would appear to be a ‘gender as practice’ type approach as espoused by, for example, Eckert and McConnell-Ginet (1992). It is simply not possible or desirable either to separate the different aspects of gender that might be at work here, nor is it possible to isolate gender from other social factors, especially when we stop to consider the added dimension of potentially different Polish and English gender identities. This is not to say that gender is too complex to be investigated in this context, rather that its complexity demands a more thorough analysis than can be provided here.

The significance of attitude towards Manchester and Manchester people in the acquisition of the local STRUT variant is interesting, and is a useful quantitative illustration of what becomes apparent when one looks at the qualitative data. Variation in STRUT exists above the level of consciousness in many of the participants, and while many commented on the difference between what they had been taught and the local variant (some claiming to find it very unattractive, none claiming to find it a desirable model), it is clearly being acquired to some extent by those who feel positive about being in Manchester. A useful follow-up project would be to look at some of the reasons why certain linguistic environments are more likely to show the local variant than others.

One final interesting finding is the fact that LoE is significant in the increase of word final intervocalic glottal replacement, in the reduction of pre-consonantal released [t], in the increased use of weak forms in STRUT function words, and in the reduction of non-target STRUT vowels. Perhaps it opens a debate on what should actually be taught in terms of L2 pronunciation. Or perhaps it simply highlights some aspects of L2 English pronunciation, irrespective of how they are taught or acquired, that play a part in perceived proficiency.

REFERENCES

Baker, W. 2008. Social, Experiential and Psychological Factors Affecting L2 Dialect Acquisition. In M. Bowles, R. Foote, S. Perpinan & R. Bhatt (eds) *2007 Second Language Research Forum,* Univerity of Illinois: Cascadilla Proceedings Project.

Bauere, V., P. Densham, J. Millar, J. Salt 2007 Migrants from Central and Eastern Europe: local geographies. *POPULATION TRENDS-LONDON-,* 129**,** 7.

Chambers, J. K. 1992. Dialect Acquisition. *Language,* 68**,** 673-705.

Dörnyei, Z. 2002. *Questionnaires in second language research: Construction, administration, and processing*. Lawrence Erlbaum.

Eckert, P., McConnell-Ginet, S.(1992. Think practically and look locally: Language and gender as community-based practice. *Annual review of anthropology,* 21**,** 461-488.

Fabricius, A. 2000. T-glottalling between stigma and prestige: a sociolinguistic study of modern RP. *Unpublished PhD thesis, Copenhagen Business School*.

Fabricius, A. 2002. Ongoing change in modern RP. Evidence for the disappearing stigma of t-glottalling. *English world-wide,* 23**,** 115-136.

Foulkes, P., Scobbie, J., Watt, D. Forthcoming. Sociophonetics. In *Handbook of Phonetic Sciences,* eds. W. Hardcastle & J. Laver.

Fox, R. A. McGory, J. 2007. Second language acquisition of a regional dialect of american english by native japanese speakers. In O.-S. Bohn & M. J. Munro (eds) *Language Experience in Second Language Speech Learning: In honor of James Emil Flege.* Amsterdam: John Benjamins.

Holmes, J. 1997. Women, language and identity. *Journal of Sociolinguistics,* 1**,** 195-223.

Johnson, D. 2008. Rbrul. http://www.ling.upenn.edu/~johnson4/rbrul.html

Johnson, D. 2009. Getting off the GoldVarb Standard: Introducing Rbrul for Mixed-Effects Variable Rule Analysis. *Language and Linguistics Compass,* 3**,** 359-383.

Mathisen, A. 1999. Sandwell, West Midlands: ambiguous perspectives on gender patterns and models of change. In P. Foulkes & G. Docherty (eds) *Urban Voices: Accent Studies in the British Isles,* London: Arnold, 107-123.

Mees, I., Collins, B. 1999. Cardiff: a real-time study of glottalisation. In P. Foulkes & G. Docherty (eds) *Urban Voices: Accent Studies in the British Isles,* London: Arnold.

Milroy, J., Milroy, L., Hartley, S., Walshaw, D. 1994. Glottal Stops and Tyneside Glottalization: Competing Patterns of Variation and Change in British English. *Language Variation and Change,* 6**,** 327-57.

Munro, M. J., Derwing, T., Flege, J. 1999. Canadians in Alabama: a perceptual study of dialect acquisition in adults. *Journal of Phonetics,* 27**,** 385-403.

Payne, A. 1980. Factors controlling the acquisition of the Philadelphia dialect by out-of-state children. In *Locating Language in Time and Space,* ed. W. Labov, 143–78. New York: Academic Press.

Sharma, D. 2005. Dialect stabilization and speaker awareness in non‐native varieties of English1. *Journal of Sociolinguistics,* 9**,** 194-224.

Straw, M., Patrick, P. 2007. Dialect acquisition of glottal variation in /t/: Barbadians in Ipswich. *Language Sciences,* 29**,** 385-407.

Tagliamonte, S. A., Molfenter, S. 2007. How'd you get that accent?: Acquiring a second dialect of the same language. *Language in Society,* 36**,** 649-675.

Watt, D., Milroy L. 1999. Patterns of variation and change in three Newcastle vowels: is this dialect levelling. In P. Foulkes & G. Docherty (eds) *Urban Voices: Accent Studies in the British Isles,* London: Arnold.

Williams, A., Kerswill, P. 1999. Dialect levelling: change and continuity in Milton Keynes, Reading and Hull. In *Urban voices. Accent studies in the British Isles.,* In P. Foulkes & G. Docherty (eds) *Urban Voices: Accent Studies in the British Isles,* London: Arnold.

Woods, N. 1997. The formation and development of New Zealand English: Interaction of gender-related variation and linguistic change. *Journal of Sociolinguistics,* 1**,** 95-125.

“Age” effects on second language acquisition

James Emil Flege, Ian R. A. MacKay***

*University of Alabama at Birmingham, **University of Ottawa
flegejames@yahoo.com, imackay@uottawa.ca

ABSTRACT

Four hypotheses regarding “age” effects on ultimate L2 proficiency are considered here. All four are found to have some predictive power, but none is perfect. This suggests that age effects arise from multiple factors that co-vary with age in ways that are not yet fully understood.

Keywords: Age, AOA, foreign accent, critical period, speech.

1. INTRODUCTION

That “earlier is better” for learning a second language (L2) has been demonstrated convincingly from the 1980s on. An “age” factor seems to be especially strong for L2 speech perception and production. Hundreds of studies have examined how particular vowels and consonants (“sounds”, for short) are identified or discriminated by listeners, how L2 learners articulate the sounds making up L2 words, and the effects of inaccurate L2 articulation on native listeners’ attempts to recognize words spoken with foreign accent (FA).

Surprisingly, what the general public has gleaned about L2 speech learning is not too different from the “received wisdom” of academics not specializing in L2 speech research. It is still common to hear it said that everyone who begins to learn an L2 after puberty is destined to forever speak it with a FA, and that anyone lucky enough to begin learning an L2 before the end of a “critical” period will learn the L2 effortlessly, rapidly, and perfectly. This view not without an element of truth, but it is mistaken in important ways. The aim of this talk, then, is to briefly review some research examining age effects on overall success in learning an L2, and to underscore important questions that remain unanswered.

1.1. Age?

In L2 acquisition research, “age” refers to the *chronological* age at which L2 learning began. My own work examined immigrants whose age of L2 learning was their age of arrival (AOA) in the host country. One study, for example, examined native speakers of Italian who learned English in Canada after emigrating there from Italy (Flege et al. 1995). Another examined Koreans who learned English in the United States (Flege et al. 1999). With but few exceptions, contact with the L2 was immediate and plentiful for these research participants, for they were compelled to learn their L2 (English) due to its social and economic dominance in the surrounding community.

In studies such as these, AOA marked participants’ first substantial contact with the L2 as spoken by native speakers of the L2. A few participants who arrived as adolescents or young adults reported having studied English briefly in school before emigrating, but there is no evidence that this previous experience exerted a lasting effect. To ensure that our research focused on sequential rather than simultaneous bilinguals, we excluded individuals who arrived prior to the age of 2 years.

AOA has proven to be a useful research variable for several reasons. It can be reported accurately; participants differing in AOA are fairly easy to recruit; and constructing groups that differ in AOA ensures robust outcomes. The real importance of AOA, however, is its presumed association with “causative” variables thought likely to vary with AOA. These include variation in neural maturation (itself thought to influence degree of neural plasticity), the state of development of native language phonetic categories, and the kind and/or amount of L2 input typically received.

AOA varies in a near-continuous fashion in immigrant populations inasmuch as whole families often emigrate together. Some research studies have adopted a correlation approach, evaluating the effect of AOA in samples of participants selected randomly or on the basis of some other variable. Better controlled research has used AOA as a selection variable. A canonical research design, for example, would be one contrasting "early" learners having AOAs of, say, 2 to 10 years to a group of "late" learners having AOAs of, say, 15 to 23 years. A more ambitious design was that of Flege et al. (1995), who selected 10 groups of 24 participants each. The groups' mean AOAs ranged from 3 (*range* 1.9 to 4.1) to 21 years (*range* 20.2 to 23.2). One aim of this study, and a similar study examining Korean immigrants to the U.S. (Flege et al. 1999), was to identify the AOA of the first non-native group to differ significantly from the native English (NE) comparison group.

As far as I know, all research to date has demonstrated strong effects of AOA. Consider, for example, the seminal study of Johnson and Newport (1989). These authors examined knowledge of 12 grammatical structures said to be the most important of English. The stimuli consisted of both grammatical (*Last night the old lady <u>died</u> in her sleep*) and ungrammatical versions (**Last night the old lady* <u>die</u> *in her sleep*) of 138 English sentences. The participants were 46 Chinese and Korean speakers, all students or professors at an American university who had lived in the U.S. for at least 3 years. Their task was to label each sentence as "grammatical" or "ungrammatical". A strong correlation between AOA and the percent correct scores was obtained: the later the arrival, the lower were the scores.

A second example is a study by Flege (1991), who measured a well-known acoustic phonetic dimension in speech production, voice-onset time (VOT). This study compared early to late native Spanish learners of English. The early learners were found to produce word-initial English /t/ with mean VOT values that were very similar to English monolinguals' whereas the late learners' VOT values were intermediate to the mean values obtained for English and Spanish monolinguals.

2. EXPLAINING "AGE" EFFECTS

Although age (AOA) effects are robust, the best explanation for these effects remains uncertain, even controversial. With this in mind, I will review four general hypotheses culled from the literature that might be used to explain age effects.

2.1. The maturational constrain hypothesis (H1)

H1 states that endogenous changes that occur as humans mature render less effective mechanisms that subserve language acquisition. DeKeyser (2000: 518-519) observed strong effects of AOA on grammaticality judgment test scores. He concluded that such effects arose as the consequence of a *"severe decline* [in] *ability to induce abstract patterns implicitly"* which is an *"inevitable consequence of fairly general aspects of neurological maturation"*. In a similar vein, Scovel (2000) concluded that age effects have a *"neuromotor"* etiology, attributing the poorer performance by late than early learners to the reduction of cerebral *"plasticity"* after the closing of a *"critical period"* for L2 learning.

One strength of H1 is that it correctly predicts differences between early and late learners as well the strong AOA-L2 performance correlations observed in many studies. H1 also correctly predicts the results obtained in an analysis of data obtained by Flege et al. (1995) and Flege et al. (1999). These studies evaluated overall degree of perceived FA in sentences spoken by Italian and Korean immigrants to Canada and the U.S., respectively. Degree of FA in sentences spoken by the non-native participants were rated along with sentences produced by members of a native English (NE) comparison group. Individual non-native participants who received a mean FA rating that fell within 2.0 *SD*s of the mean rating obtained for the NE group ($n = 24$) were considered to be "accent free" (admittedly a rather lax criterion). As predicted by H1, no Koreans and just one Italian participant with an AOA greater than 12 years were found to be "accent free".

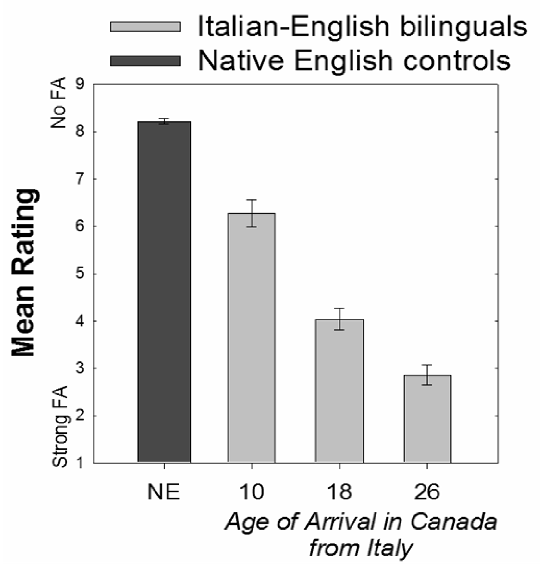
The same analysis, however, revealed a problem for H1. Less than half of the Italians and only about one-quarter of the Koreans with an AOA less than 10 years were found to be accent free. This finding is striking given that all of these participants were educated in English-medium schools, most had been immersed in English for more than 15 years, and nearly all reported using English more than their L1.

The two studies just cited used retrospective developmental designs, that is, examined adults who differed according to when in the past they began learning English. Flege et al. (2005), on the other hand, examined FA in English sentences spoken by groups of Korean children having AOAs that ranged from 5 to 15 years.

After 5 years of immersion in English, the Korean children's sentences still received significantly lower ratings than sentences produced by NE children. The same held true when the same test was performed a year later. A closer look revealed the presence of detectable FAs even in sentences spoken by Korean children who had begun learning English at the age of 6 years and had been immersed in English for more than 5 years.

Another problem for H1 is its generation of a prediction that has been falsified in several studies. If AOA effects are due to the passing of a critical period, one would expect L2 performance to decline as participants' AOAs near the end of a critical period. However, performance should remain stable thereafter because everyone *beyond* the critical period should show equally the ill effects of having passed the critical period. A re-analysis of FA ratings obtained for Italian and Korean immigrants showed, contrary to this expectation, that FAs continue to grow stronger beyond the critical period. For each study (Flege et al. 1995, 1999), AOA-FA correlations were computed for the entire sample (i.e., the 240 participants having AOAs ranging from 2 to 23 years), just for participants having AOAs ranging from 12 to 23, and finally just those having AOAs of 15 to 23 years. As the number of participants decreased, the correlations became weaker but remained significant ($p < .01$) even in the last sample which included only "post critical period" learners.

Figure 1: Ratings of sentences spoken by native English (NE) speakers and groups of Italian immigrants. Each mean value is based on 2,916 ratings (18 participants x 162). The error bars bracket +/- 1 *SE*.



Flege and MacKay (unpublished) evaluated degree of FA in English samples produced by NE speakers and native Italian groups differing in AOA. As seen in Fig. 1, the pair-wise differences between all four groups were significant ($p < .05$ by Tukey test). H1 might be used to explain the stronger FA of immigrants who arrived in Canada at the age of 18 than 10 years, but not the stronger FA of immigrants who began learning English at 26 than 18 years of age. This is because members of both of the last two groups mentioned had begun learning English after the end of the supposed critical period.

A re-consideration of the Johnson and Newport (1989) morphosyntax data also calls H1 into question. The strongest support obtained by these authors for H1 was a significant correlation between AOA and morphosyntax scores for participants having an AOA of less than 15 years, but not for those having AOAs of 16+. Crucially, the existence of a critical period at age 15 was assumed, not demonstrated. Hakuta and Bialystok (1994) showed that when the "cut point" was arbitrarily shifted to an AOA of 20 years, a moderately strong ($r = -0.49$) correlation between AOA and the morphosyntax scores was obtained for "post-critical period" learners.

Even more strikingly, the results of Flege et al. (1999) suggested that the effect of AOA on grammaticality judgment test scores may disappear when factors confounded with AOA have been controlled. These authors used a test modelled on that of Johnson and Newport (1989). As in the original study, the test scores showed a strong correlation with AOA. But, as is usually the case in L2 research, AOA was correlated with other variables that might well have influenced the test scores: years of education in English-medium schools, length of residence in the U.S., and self-reported use of English.

Flege et al. (1999) used the matched-subgroup technique in an attempt to unconfound these other potentially important variables. Two groups of 20 participants having mean AOAs of 10 and 17 years were selected from the original sample of 240 without considering any other factor. Not surprisingly, the early learners obtained much higher scores than the late learners did. Two additional groups having the same mean AOA values (viz., 10 and 17 years) were then selected, this time taking care to match the participants in each AOA group for years of education in the U.S. (*mean* 11 years for both). The matching process also eliminated significant between-group differences in years of U.S. residence and self-reported English use. Importantly, the morphosyntax scores for these *matched* groups of early and late learners did not differ (and, in fact, did not even approach significance). This finding leads one to wonder how many "age" effects reported in the literature were really due to factors confounded with AOA and not actually to the age at which L2 learning began.

2.2. The cognitive development hypothesis (H2)

H2 states that L2 learning becomes gradually less effective across the entire life span because cognitive abilities needed for speech and language learning diminish slowly across the life span. H2 is derived from a study by Hakuta et al. (2003). These authors developed estimates of English proficiency based on several responses to the U.S. Census by large numbers of Chinese- and Spanish-speaking immigrants. For both groups, the English proficiency estimates decreased continuously from 10 to 60 years, not just in the second decade of life.

H2 correctly predicts the FA difference in Fig. 1 between immigrants having AOAs of 18 and 26 years. But for H2 to really explain this finding, it will be necessary to identify a specific cognitive change that might be responsible. A more serious limitation for H2, however, is the unlikelihood of its being able to explain the difference between NE speakers and immigrants with an AOA of 10 years (Fig. 1). It is *a priori* difficult to imagine 10-year-olds having undergone a cognitive change that could reduce their ability to learn a second language in a period of life in which the L1 is still developing.

2.3. Changes in L1-L2 interactions (H3)

H3 states that as the L1 system develops, the effect of cross-language phonetic interference becomes stronger because of differences in how the L1 and L2 sound systems interact. This hypothesis is derived from Flege's Speech Learning Model (e.g., Flege 2002, 2003, 2007, 2009). H3 assumes that L1 categories develop slowly, at least into early adolescence (as shown, for example, by studies examining the recognition of L1 words in noise). According to H3, as L1 categories develop they become stronger "attractors" for sounds later encountered in an L2. This has the effect of making learners less likely to create new categories for L2 sounds that are similar but non-identical to corresponding L1 sounds, even when the cross-language phonetic differences are audible.

H3 generates predictions that can be readily tested. It predicts, for example, that as native Spanish speakers get older, their ratings of the perceived phonetic dissimilarity between tokens of English [th] and Spanish [t] (all realizations of a /t/ phoneme) will decrease. H3 also predicts that the frequency of detections of English-like VOT values inserted into Spanish speech samples will decrease as Spanish monolinguals grow older (and, possibly, that the latencies of correct detections will increase). Most importantly, H3 predicts that as an indirect consequence of the development of L1 phonetic categories, Spanish early learners will be more likely than Spanish late learners to establish a new phonetic category for English /t/, and so will be more likely than late learners to produce English /t/ correctly.

A less obvious prediction—also derived from the SLM—regards what is expected to happen when learners (presumably, mostly late learners) fail to establish a new phonetic category for L2 sounds that differ auditorily from the closest sound in L1. Here the expectation is the development of a composite (merged) category sharing properties of the perceptually "equated" L1 and L2 sounds. This leads to the prediction, for example, that late Spanish learners of English will tend to produce English /t/ with VOT values that are too short for English, and Spanish /t/ with VOT values that are too long for Spanish.

As already mentioned, Flege (1991) found that early Spanish learners of English produced English /t/ with native-like values whereas late learners produced values intermediate to those observed for Spanish and English monolingual. H3 generates the same predictions for French as for Spanish. Flege (1987) observed “compromise” VOT values for late French-English bilinguals. Also, both French learners of English and American learners of French were found to produce L1 /t/ with values that had moved in the direction of typical L2 values.

A problem for H3 is that little research has been conducted to determine if the perceived dissimilarity of pairs of L1 and L2 sounds actually does decrease as L1 categories develop, and whether such changes—if they occur—predict accuracy of L2 segmental production and/or perception (but see Baker et al. 2002).

2.4. More/better input for early than late learners (H4)

There is evidence that early learners typically get more, and perhaps more authentic, L2 input than late learners do. H4 states that L2 proficiency increases as a function of amount of L2 input. Given the assumption that input varies as a function of language use, it predicts that participants who have used the L2 often over the years will outperform those who have used L2 less often.

The prediction of H4 was supported by research examining Italian immigrants in Canada (Flege and MacKay 2004; Piske et al. 2001, 2002; MacKay et al. 2001). Groups differing in AOA (*means* 8 vs. 20 years) were subdivided accorded to self-reported frequency of Italian use (*means* of 7% and 10% vs. 43% and 53%). For both early and late learners, a frequent use of Italian (and thus an infrequent use of English) was associated with a significantly poorer identification of word-final English consonants, a poorer discrimination of English vowels, less accurate production of English vowels, and relatively stronger FAs.

Table 1: Characteristics (*means*, *SD*s) of three groups of Italian immigrants tested by Flege and MacKay (unpubl.). All differences between AOA-10 and AOA-18, but no differences between AOA-18 and AOA-26, were significant ($p < .05$).

	AOA-10	AOA-18	AOA-26
Years of residence in Canada	43(5)	37(7)	33(8)
Years of education in Canada	9(4)	<1(1)	<1(1)
Self-reported % use of English	71(14)	53(15)	47(19)
Self-reported % use of Italian	27(14)	47(15)	52(20)
Self-rated (7-pt) ability to pronounce Italian	5.3(1.1)	6.4(0.8)	6.8(0.5)
Self-rated ability to speak-understand Italian	5.5(1.0)	6.4(0.8)	6.8(0.3)

These findings discredit claims (e.g., DeKeyser 2000: 519) that input differences do not contribute to age effects on L2 speech learning. However, it is not clear at present to what *extent* input differences contribute to age effects inasmuch as L2 research has relied on self-reports of language use rather than taking actual measurements of input. Another problem for H4 is that language use differences do not predict all findings currently available. For example, H4 might explain the difference in Fig. 1 between participants having AOAs of 10 and 18 years but not the difference between those having AOAs of 18 and 26 years. As can be seen in Table 1, the former but not the later pair of groups was likely to have differed in terms of amount of L2 input.

3. CONCLUSIONS

All four hypotheses considered here seemed to have some predictive value, but none offered a convincing explanation of age effect and none was able to account for all of the evidence considered. The possibility therefore exists that all four hypotheses might help explain, to varying degrees, age-related effects seen in the literature.

Future research should be designed so that confounded factors can be eliminated *a priori* or, at the very least, be controlled statistically *a posteriori*. It will be imperative to evaluate variables thought to affect ultimate L2 proficiency directly rather than, as in the past, to manipulate "age" of L2 learning via AOA. Potentially "causative" variables worth investigating include: age-related changes in the perceived relation between sounds in the L1 and L2; variations in the kind of input received (foreign accented?), especially in early stages of L2 learning; and the amount (measured, not estimated!) of L2 input received. Until causal variables have been examined directly and variables confounded with them have been controlled, we can only speculate about the true basis (bases) for age-related differences in ultimate L2 proficiency.

4. REFERENCES

Baker, W., Trofimovich, P., Mack, M., Flege, J. 2002. The effect of perceived phonetic similarity on non-native sound learning by children and adults. In: Skarabela, B. et al. (eds.). 36-47.

Bialystok, E., Hakuta, K. 1994. *In other words: The science and psychology of second language acquisition*. New York: HarperCollins.

Birdsong, D. (ed.) 1999. *Second language acquisition and the critical period hypothesis*. Hillsdale: Lawrence Erlbaum.

Burmeister, P. et al. (eds.). 2002. A*n integrated view of language development: Papers in honor of Henning Wode*. Trier: Wissenschaftlicher.

Cole, J., Hualde, J. (eds.) 2007. *Laboratory phonology 9*. Berlin: Mouton de Gruyter.

DeKeyser, R. 2000. The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition* 22. 499-534.

Flege, J. 1987. The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15. 47-65

Flege, J. 1991. Age of learning affects the authenticity of voice onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America* 89. 395-411.

Flege, J. 1999. Age of learning and second-language speech. In: Birdsong, D. (ed.). 101-132.

Flege, J. 2002. Interactions between the native and second-language phonetic systems. In: Burmeister, P. et al. (eds.). 217-244.

Flege, J. 2003. Assessing constraints on second-language segmental production and perception. In Meyer, N. Schiller (eds.). 319-355.

Flege, J. 2007. Language contact in bilingualism: phonetic system interactions. In: Cole, J., Hualde, J. (eds.). 353-382.

Flege, J. 2009. Give input a chance! In: Piske, T., Young-Scholten,M. (eds.). 175-190.

Flege, J., Birdsong, D., Bialystok, E., Mack, M., Sung, H., Tsukada, K. 2005. Degree of foreign accent in English sentences produced by Korean children and adults. *Journal of Phonetics* 33. 263-290.

Flege, J., MacKay, I. 2004. Perceiving vowels in a second language. *Studies in second language acquisition* 26. 1-34.

Flege, J., Munro, M., MacKay, I. 1995. Factors affecting degree of perceived foreign accent in a second language. *Journal of the Acoustical Society of America* 97. 3125-3134.

Flege, J., Yeni-Komshian, G., Liu, S. 1999. Age constraints on second language learning. *Journal of Memory and Language* 41. 78-104.

Hakuta, K.; Bialystok, E., Wiley, E. 2003. Critical evidence: A test of the critical-period hypothesis for second-language acquisition. *Psychological Sciences* 14. 31-38.

Johnson, J., Newport, E. 1989. Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology* 21. 60-99.

MacKay, I., Meador, D., Flege, J. 2001. The identification of English consonants by native speakers of Italian. *Phonetica* 58, 103-125.

Meyer, A., Schiller, N. (eds.) 2003. *Phonetics and phonology in language comprehension and production*. Berlin: Mouton.

Piske, T., Flege, J., MacKay, I., Meador, D. 2002. The production of English vowels by fluent early and late Italian-English bilinguals. *Phonetica* 59. 49-71.

Piske, T., MacKay, I., Flege, J. 2001. Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29, 191-215.

Piske, T., Young-Scholten, M. (eds.) 2009. *Input matters in SLA*. Bristol: Multilingual Matters.

Scovel, T. 2000. A critical review of critical period research. *Annual Review of Applied Linguistics* 20. 213-223.

Skarabela, B., Fish, S., Do, A. (eds.). 2002. *Proceedings of the 26th annual Boston University conference on language development*, Somerville: Cascadilla Press.

Perceiving non-native vowel contrasts: ERP evidence of the effect of experience

Cheryl Frenck-Mestre, Pauline Peri, Christine Meunier & Robert Espesser
Laboratoire Parole et Langage, CNRS, Université de Provence

ABSTRACT

In two ERP experiments, we used a 3-stimulus oddball paradigm to examine the perception of American-English vowel contrasts, by native speakers, French monolinguals and late French-English bilinguals. In the first experiment, stimuli were standard /ɛ/ (75%), target /ae/ (10%) and oddball /ɪ/ (15%). In the second, the attentional demands were inversed: standard /ɛ/ (75%), target /ɪ/ (10%) and oddball /ae/ (15%). In both experiments and in all 3 groups, early acoustic discrimination of all vowels was shown by variations in the N100 response. Subsequent phonemic categorization as revealed by the P300 response differed however, across experiments and groups. Bilinguals showed a P300 response to oddball /ɪ/ akin to English speakers, though reduced in magnitude and distribution, whereas French monolinguals did not. When /ɪ/ became the target and /a/ the deviant, native English and French showed an inversion of the P300 to these vowels; bilinguals could not disengage attention and showed equal P300 responses to both. Our results indicate that while late French-English bilinguals establish new L2 vowel categories these categories are not as stable as L1 categories. Late learners' ability to discriminate and to selectively attend to a given L2 category is intermediate between that of native speakers and native speakers of their own L1.

Keywords : ERPs, P300, late bilinguals, non-native contrasts

INTRODUCTION

One of human being's irrepressible activities is that of categorisation, the act of consciously sorting the observable events in the external environment into recognizable entities that can then be further processed, stored in memory as identifiable traces and recalled from memory as needed. The perception of the sounds of the world's languages is no different. From infancy, we learn to differentiate auditory events and to categorize them as either exemplars of a common category or as members of different categories (Dehaene-Lambertz & Baillet, 1998; Kuhl, Williams, Lacerda, Stevens & Lindblom, 1992; Werker & Tees, 1984; Werker, 1994). The results of numerous studies, obtained with a wide variety of techniques, suggest moreover that one's early experience with auditory input has considerable impact on the ability later in life to create new categories, otherwise stated to reorganize the existing system to accommodate novel sounds as opposed to assimilating them to established categories (Iverson & Kuhl, 1995; Dehaene-Lambertz, 1997) although, as outlined below, various factors over and above early-life experience play a role in how easily new perceptual categories can be created. The present study used event-related cortical potentials to determine the extent to which adult learners create new perceptual categories when they learn a foreign language, how native-like their categories are and the role of experience in this ability.

There is ample evidence that non-native contrasts can be acquired after infancy; the extent of success depends nonetheless on numerous factors. Early exposure, during childhood, to a second language does not in fact guarantee the formation of native-like phonemic categories; however, the likelihood of such is increased if the L1 is not maintained (Flege & McKay, 2004; Tsukada et al., 2005). L2 contrasts that form a "single category" in the L1 may not become native-like even for early learners (Højena & Flege, 2006). Moreover, the acquisition of L2 vowel contrasts has often been found to be conditioned by the phonological features present in the native language of learners (Brown, 1998; McAllister, Flege & Piske, 2002). These results, which support the idea that the native language plays an important role and enables one to predict the relative difficulty of acquisition of a given L2 contrast, are in line with two predominant models of adult learning/ perception of non-native contrasts. Both the Speech Learning Model, or SLM (Flege 1995), which focuses on second language acquisition, and the Perception-Assimilation Model, or PAM (Best, 1994), which is aimed towards the study of perceptual capacities in monolinguals (but see Best & Tyler, 2007, for an extension to L2 learning), provide a rich theoretical framework of non-native phonemic perception. Note that neither model restricts the acquisition of non-native contrasts to the simple presence vs. absence of phonological contrasts within the L1. Indeed, a considerable body of work has shown that the capacity to perceive L2 vowel contrasts is influenced by phonetic properties of the speech signal in and of itself such that L2 contrasts can be perceived in certain phonetic contexts/positions but not others (Strange, 2007; Levy & Strange, 2008; Trofimovich, Baker & Mack, 2001). Moreover, recent psycholinguistic studies of L2 perception have shown that meta-linguistic factors can also play a role in the discrimination of L2 minimal pairs (Weber & Cutler, 2004).

The present study presents further evidence that adult L2 learners can reshape their perceptual space to accommodate L2-specific vowel contrasts but that, in line with the results of the vast majority of studies, “native-like” performance is not observed. Whereas the majority of studies of L2 (and non-native) phoneme perception have used behavioural techniques to address this question, we chose to record event-related potentials (ERPs) to examine the perceptual capacities of our participants. ERPs provide not only a millisecond precise measurement of processing but, depending upon the task, can allow one to disentangle automatic detection from attentional processing. To date, only a handful of ERP studies have examined phonemic processing specifically for L2 acquisition. These studies have used an oddball paradigm under conditions of non-attentional processing and examined the mismatch negativity response (MMN) to L2 contrasts. The results of these studies are mixed. Whereas Winkler et al. (1999) found that adult late L2 learners who had been immersed for several years in their L2 perceived non-native contrasts (in Finnish) as well as native speakers, this result was not replicated in a population of advanced adult L2 learners (of English) who were not immersed (Peltola et al., 2003). The results obtained for Finnish children in either bilingual or full immersion programs revealed very rapid onset of neural changes in response to L2 (French) vowel contrasts (Choeur et al., 2002; Peltola et al., 2005). Again however, subsequent work did not confirm these findings when the L2 was English (Pelota et al., 2007), although the pattern of results of the bilinguals was puzzling. Nonetheless, caution is warranted before concluding that young L2 learners achieved native-like capacity to perceive their L2 as no L1 control group was included in these studies.

Here, we report the results of two ERP investigations of non-native vowel processing, using an active oddball paradigm. Rather than have our participants passively attend to auditory stimuli, we required them to mentally count a target vowel (10% of trials) presented against a background “standard” (75% of trials) and a deviant (15%) vowel. This allowed us to examine the effects of attention, by varying the status of a given vowel, i.e. whether target or deviant, across experiments. To address the question of just how “native-like” L2 learners’ perception can become, we compared 3 groups: control native English speakers, control French speakers who had only had classroom learning of English, and French-English late bilinguals.

METHOD

Participants. Twenty-four right-handed adults aged 18 to 24 participated. There were 8 English speakers living in France (mean of 6 months), 8 native French “monolinguals” who had learned English as a foreign language throughout secondary school, and 8 late French-English bilinguals. All late bilinguals were native French speakers raised by native French parents, who had learned English in secondary school starting at age 11, were studying to become English instructors, had lived at least 1 year in an English speaking country and rated themselves as fluent in English for oral and written comprehension and expression.

Stimuli. The stimuli consisted of the American-English vowels /E/, /I/ and /ae/, extracted from a lexical context (/h/_/d/) produced by a trained female native speaker. Twenty-five utterances of each lexical item were recorded in a sound attenuated room and digitized at 32 kHz/ 32 bits. Stimuli were created by extracting 60 ms from either side of the steady-state of the vowel and adding 10 ms contours to the onset and offset. Seven tokens were selected for each vowel category. F0 was held constant across stimuli. F1, F2 and F3 values were automatically extracted. The mean formant values were: /E/ (F1 = 625 Hz, F2 = 2120 Hz, F3 = 2845 Hz), /I/ (F1 = 480 Hz, F2 = 2295 Hz, F3 = 2953 Hz), and /ae/ (F1 = 930 Hz, F2 = 1857 Hz, F3 = 2698 Hz). The Euclidean distance calculated between mean Bark values for F1 and F2 was larger between /ae/ and /E/ (2.32) than between /I/ and /E/ (1.32).

Design and Procedure. Stimuli were delivered binaurally via headphones. A total of 625 stimuli were presented in a fixed random order, with a probability of occurrence of .75, .10 and .15 for standard /E/, target and deviant stimuli respectively. Stimulus duration was 140 ms and SOA 1200 ms. In Experiment 1, the target was /ae/ and the deviant /I/; in Experiment 2 the target was /I/ and the deviant /ae/. Participants mentally counted the target stimuli. They were not informed of the presence of the deviant.

Electroencephalographic recording. EEG activity was recorded continuously at pre-frontal (Fp1, Fp2), frontal (F3, F4), occipital (O1, O2) and midline (Fz, Cz, Pz) electrode sites, referenced to the left mastoid. The EEG was amplified with a bandpass of 0.1 – 40 Hz (3dB cutoff) and digitized on-line at 200 Hz. EEGs were filtered offline below 15 Hz. Epochs began 100 ms prior to stimulus onset and continued 1100 ms thereafter. Average ERPs were formed off-line from trials free of muscular and/or ocular artifact.

Data Analysis. The ERP data were quantified by calculating the peak and/or mean amplitudes and latencies post-stimulus onset and relative to a 100 ms pre-stimulus baseline, for N100 (80–150 ms), for N200 (180–280 ms) and for P300 (280–500 ms) for each participant. Independent ANOVAs (Greenhouse-Geisser corrected) were performed on amplitude and latencies.

RESULTS

N100

The grand average waveforms elicited by /E/, /ae/ and /I/ are presented for N100 as a function of experiment in Fig. 1. In both experiments, the effect of vowel was significant (Exp 1: F (2, 42) = 12.54 p<.001; Exp 1: F (2, 42) = 11.54 p<.01); N100 peak amplitude was largest for /ae/, smallest for the standard /E/ and intermediate for /I/, as confirmed by pairwise comparisons (p<.01 or better, Bonferroni). No interactions obtained with group (F<1) or experiment (F<1). The differences between the three vowels are in line with the Euclidean distance between them, and are independent of the role played, whether target or deviant.

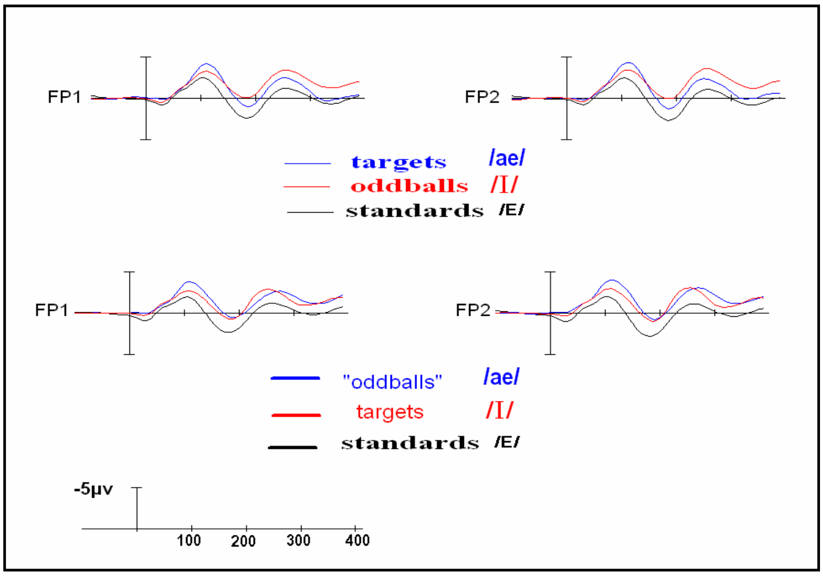


Figure 1. Grand average wave forms as a function of vowel and experiment, for all participant groups.

P300

The grand average waveforms elicited by /E/, /ae/ and /I/ are presented for each of the three participant groups in the P300 window, as a function of vowel and experiment, in Fig. 2 – 4. In Experiment 1, the target /ae/ elicited a large positive deflection compared to the standard /E/ in all groups. The deviant /I/, however, only elicited a P300 response for native English and French-English bilingual participants. In Experiment 2, where participants were now requested to mentally count the vowel /I/, both the target /I/ and deviant /ae/ produced a larger mean P300 response than standard /E/ in all groups. Importantly, the P300 response varied as a function of the role of the vowel across experiments, as confirmed by ANOVAs.

Experiment 1. Mean P300 amplitude varied significantly at midline as a function of Vowel ($F(2, 42) = 50.48$, $p<.0001$), which was modified by the higher order interaction involving Vowel, Electrode and Group ($F(8, 84) = 4.05$, $p<.001$). Post hoc comparisons revealed that in all groups, the target /ae/ produced a large P300 response than standard /E/ at all midline sites. For the deviant /I/, in both the French-English bilingual group and the native English group this vowel produced a larger P3 amplitude than standard /E/ at Cz and Pz. In the French monolingual group the deviant /I/ did not differ from the standard /E/.

Experiment 2. Mean P300 amplitude varied significantly at midline as a function of Vowel ($F(2, 42) = 15.83$, $p<.0001$), which was modified by the higher order interaction involving Vowel, Electrode and Group ($F(8, 84) = 2.44$, $p<.05$). Post hoc comparisons revealed that in the native English group, mean P300 amplitude was larger for both target /I/ and deviant /ae/ than standard /E/, but that the effect was larger for the target /I/. In the French monolingual group, target /I/ produced a larger P300 than standard /E/ whereas the deviant /ae/ tended to (p.07). In the French-English bilingual group, both target /I/ and deviant /ae/ produced larger P300 responses than the standard /E/, and did not differ from each other.

Experiment 1 vs. Experiment 2

To assess the effect of task demands, i.e. of the allocation of attentional resources to perceiving the different vowel categories, an ANOVA was performed on the effect sizes for the target and deviant vowels, i.e. the ERP response produced in response to each of these stimuli in comparison to the standard vowel /E/, in the P300 time window across experiments. For native English speakers, the effect size of /ae/ and /I/ was inversed across experiments in accordance with the task; for native French, the same inversion was found. For French-English late bilinguals however, whereas the effect size of /I/ increased Experiment 2 that of /ae/ did not decrease significantly.

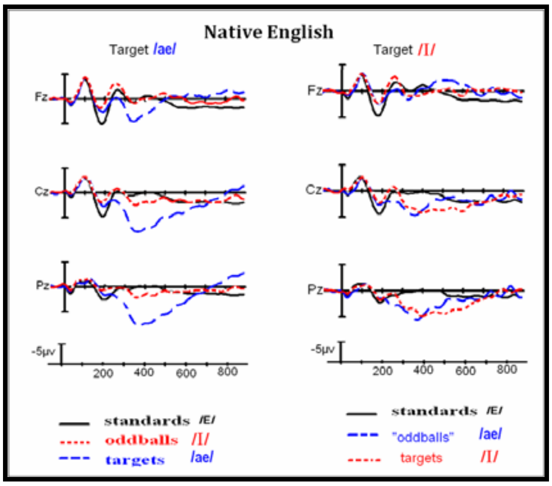


Figure 2. P300 response as a function of vowel and experiment, for native English speakers.

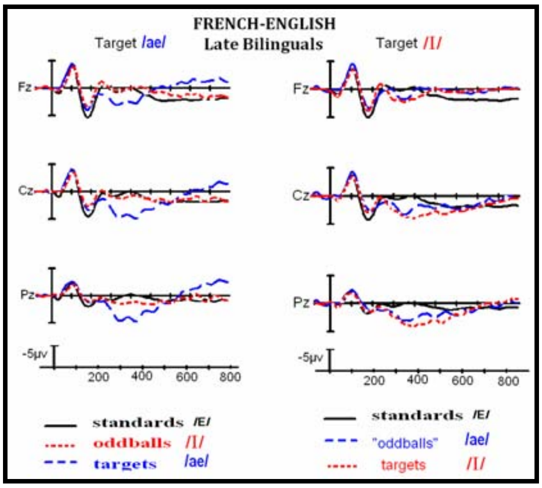


Figure 3. P300 response as a function of vowel and experiment, for French-English bilinguals.

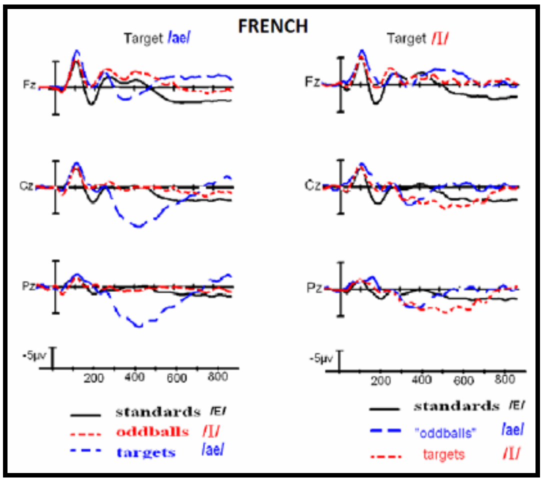


Figure 4. P300 response as a function of vowel and experiment, for French monolinguals.

General Discussion

The results from the two experiments reported here provide important insight into the perception of vowel contrasts as a function of numerous factors: the status of these contrasts--whether native or not part of the listener's native repertoire--the status of the listener--whether performing as a monolingual or a bilingual--the type of processing--whether acoustic or phonetic--and task demands. Overall, the pattern of results show that our late French-English bilingual participants were in an intermediary position between native English speakers and French native speakers who have been exposed to English through secondary education but who are not proficient in this language.

For all participants, during the first 100 ms following presentation of the three American-English vowels we found a graded ERP response, which closely followed the acoustic differences between these vowels. This graded N100 response was impervious not only to the listeners' native language repertoire but also to task demands, such that it can be considered a likely "acoustic" response rather than related to categorical perception. This result replicates previous results for the same contrasts (Frenck-Mestre et al., 2005). Following this initial response however, the three participant groups showed quite different patterns as revealed by the P300 response. Late French-English bilinguals showed a categorical response to the English vowel /I/ against the background /E/ even when attention was devoted to the target vowel /ae/. In this sense, they mimicked the pattern of native English speakers and differed from French monolinguals, who did not consciously discriminate /I/ from /E/ under these conditions. The French-English bilinguals differed from both the native English speakers and French monolinguals however, when task demands were switched. For the two groups working under "monolingual" conditions, the P300 response to the non-target vowel was significantly smaller than that to the target. For the bilingual group, even when attention was switched both American English vowels, /ae/ and /I/ the vowel /I/ elicited a reliable P300 response, with no reliable decrease in response to the non-target vowel (/ae/) across experiments, as though they were unable to reliably disengage attention. This pattern suggests that for the bilingual group, consciously sorting out the three vowel categories necessitated attention to all vowels in contrast to the "monolinguals" who were able to devote their attention exclusively to one category.

It is important to note that the comparison of the two French native speaker groups is one between "foreign language acquisition", i.e. for the group of French speakers who had predominantly been exposed to English in a classroom setting, and "second language acquisition", i.e. for the late bilingual group, who had all lived for at least 12 months immersed in the English language in addition to extensive classroom learning. Our results support the findings of ERP studies which have examined both pre-attentive and conscious processing of second-language and non-native contrasts in showing that simple exposure to these contrasts is not sufficient to produce a reliable automatic response but that immersion appears to be critical (Frenck-Mestre et al., 2005; Peltola et al., 2003). Nonetheless, even for our advanced bilingual group we did not find a pattern of vowel categorization that entirely overlapped with that of native speakers. In our first experiment, although the bilinguals consciously categorized the deviant, non-attended vowel contrast just as native speakers did, their electrophysiological response was smaller in both amplitude and distribution. In the second experiment, while the native speakers were able to selectively attend to just one of the English vowels, the group of late bilinguals did not demonstrate this capacity. These results suggest, as has been forwarded by Flege (Flege and McKay, 2004) that it may be too much to ask to expect late learners to achieve native-like performance but that they can indeed achieve very high levels of proficiency.

REFERENCES

Best, C.T. 1994. The emergence of native-language phonological influences in infants: a perceptual assimilation model. In J. Goodman & H. Nusbaum (Eds.), *The development of speech perception: the transition from speech sounds to spoken words*. Cambridge: MIT Press. 167-224.

Best, C. T., Tyler, M.D. 2007. Nonnative and second language speech perception. In O.S. Bohn and M.J. Munro (eds.) Language experience in second language speech learning. John Benjamins, Amsterdam, 13-34.

Brown, C. 1998. The role of the L1 grammar in the L2 acquisition of segmental structure. Second Language Research, 14, 136-193.

Cheour, M., Shestakova, A., Alku, P., Ceponiene, R., Näätänen, R. 2002. Mismatch negativity shows that 3–6-year-old children can learn to discriminate non-native sounds within two months. *Neuroscience Letters*, 325, 187–190.

Dehaene-Lambertz, G., Baillet, S. 1998. A phonological representation in the infant brain. *NeuroReport*, 9, 1885-1888.

Dehaene-Lambertz, G. 1997. Electrophysiological correlates of categorical phoneme perception in adults. *NeuroReport*, 8(4), 919--924.

Flege, J.E. 1995. Second-language Speech Learning: Theory, Findings, and Problems. In W. Strange (Ed) *Speech Perception and Linguistic Experience: Issues in Cross-language research*. Timonium, MD: York Press, 229-273.

Flege, J.E., McKay, I. 2004. Perceiving vowels in a second language. *Studies in Second Language Acquisition*, 26, 1-34.

Frenck-Mestre, C., Meunier, C., Espesser, R., Daffner, K., Holcomb, P. 2005. Perceiving nonnative vowels: The effect of context on perception as evidenced by event-related brain potentials. *J. of Speech, Language, and Hearing Research*, 48, 1–15.

Hojen, A., Flege, J.2006. Early learners' discrimination of second-language (L2) vowels. *J. of the Acoustical Society of America*. 119, 3072-3084.

Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *J. of the Acoustical Society of America*, 97, 553–562.

Kuhl, P. K., Williams, K.A., Lacerda, F., Stevens, K.N., Lindblom, B. 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255, 1003-1010.

Levy, E.S., Strange, W., 2008. Perception of French vowels by American English adults with and without French language experience. *Journal of Phonetics* 36, 141–157.

McAllister, R., Flege, J., Piske, T. 2002. The influence of the L1 on the acquisition of Swedish vowel quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics*, 30, 229-258.

Peltola, M.S., Tuomainen, O., Koskinen, M., Aaltonen, O. 2007. The effect of language education on the preattentive perception of native and non-native vowel contrasts. *Journal of Psycholinguistic Research*,

Peltola, M. S., Kujala, T., Tuomainen, J., Ek, M., Aaltonen, O., Näätänen, R. 2003. Native and foreign vowel discrimination as indexed by the mismatch negativity (MMN) response. *Neuroscience Letters*, 352, 25–28.

Peltola, M.S., Kuntola, M., Tamminen, H., Hämäläinen, H., Aaltonen, O. 2005. Early exposure to non-native language alters preattentive vowels discrimination. *Neuroscience Letters*, 388: 121–125.

Strange, W. 2007. Cross-language phonetic similarity of vowels: Theorertical and methodological issues. In O.S. Bohn and M.J. Munro (eds.) Language experience in second language speech learning. John Benjamins, Amsterdam, 35-56.

Trofimovich, P., Baker, W., Mack, M. 2001. Context- and experience-based effects on the learning of vowels in a second language. *Studies in the Linguistic Sciences, 31*, 167-186.

Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Flege, J.E. 2005. A developmental study of English vowel production and perception by native English adults and children. *Journal of Phonetics*, 33, 263-290.

Weber, A., Cutler, A. 2004. Lexical competition in non-native spoken- word recognition. *Journal of Memory and Language*, 50, 1-25.

Werker, J.F. 1994. Cross-language speech perception: Development does not involve loss. In J.C. Goodman & H.C. Nusbaum (eds.), *The development of speech perception*. MIT Press, Cambridge, MA, 93-120.

Werker, J.F., Tees., R.C., 1984. Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behaviour and Development*, 7, 49-63.

Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., Czigler, I., Csépe, V., Näätänen, R. 1999. Brain responses reveal the learning of foreign language phonemes. *Psychophysiology*, 26, 638–642.

L2 fluency in the production of English extemporaneous speech by Catalan/Spanish bilinguals

Natalia Fullana[1] *and Ian R. A. MacKay*[2]

[1]Universitat de Barcelona [2]University of Ottawa
nataliafullana@ub.edu imackay@uottawa.ca

ABSTRACT

This study explored oral fluency in Catalan/Spanish learners of English who had been mainly exposed to the target language in a formal learning context. Excerpts from a picture narrative task were analyzed for temporal measures of fluency. Results of low speech rate and pruned speech rate demonstrated a low level of fluency in English among the Catalan/Spanish bilinguals and evidenced a high degree of variability among learners. Moreover, exposure to English was not a determining factor in attaining a more native-like degree of fluency in the FL, though fluency was slightly better for those individuals who had stayed in an English-speaking environment for three months or more.

Keywords: fluency, L2 experience, Catalan/Spanish speakers, EFL.

1. INTRODUCTION

Second language (L2) speech acquisition research has by and large examined L2 learners in immersion settings within models of cross-linguistic speech perception and L2 learning, particularly Best's (1995) Perceptual Assimilation Model (PAM) – or the more recent version PAM-L2 (Best & Tyler 2007) – and Flege's (1995) Speech Learning Model (SLM). For the most part, studies conducted in formal learning contexts have also adopted the predictions the two models make (e.g., Cebrian 2006; Rallo-Fabra 2009), though the findings to date are far from conclusive. Specifically, the role of experience – understood as either (phonetic) instruction in the L2 or periods of short/long stays in the target language (TL) country – in L2 speech perception and production is still to be determined (e.g., Cebrian 2003; Mora 2008). Similarly, the factor of age of onset of L2 learning has often failed to provide supporting evidence for an earlier start advantage in the production and perception of L2 sounds as the SLM predicts. In fact, most investigations in formal learning contexts have reported on a late start advantage (e.g., García-Lecumberri & Gallardo 2003; Fullana 2006).

Current work has evidenced a new (or renewed) interest in the factor of input in L2 speech acquisition in both naturalistic and formal learning contexts (see, for example, Piske & Young-Scholten 2009). Additionally, recent investigations have focused on the learning of suprasegmentals (e.g., Hahn 2004; Trofimovich & Baker 2006) and on dimensions other than foreign accent, fluency having received much of the attention (e.g., Derwing, Munro & Thomson 2008; Derwing, Rossiter, Munro & Thomson 2004; Kormos & Dénes 2004; Mora 2006; Munro, Derwing & Morton 2006; O'Brien, Segalowitz, Freed & Collentine 2007; Rossiter 2009; Segalowitz 2007). As for the latter, differences in the definition of the construct of fluency have been noted at length (for a compilation of the various characterizations, see, among others, Cucchiarini, Strik & Boves 2002; Kormos 2006; Rossiter 2009). Even so, L2 speech acquisition studies often examine fluency in terms of a number of temporal and objective measures that might be interpreted from either the point of view of the listener or that of the speaker (or both). Moreover, the available results suggest that listeners' perception of L2 learners' degree of oral fluency might be predicted on the basis of speech rate – and/or pruned speech rate – and pausing phenomena (e.g., Cucchiarini et al. 2002; Derwing et al. 2004; Kormos & Dénes 2004; Rossiter 2009).

Taking all of the above into account, and in order to contribute to the fewer existing findings from formal instruction settings (e.g., Mora 2006), the present study explored oral fluency in the English of learners who studied the TL in a formal learning context. Furthermore, in line with the increasing number of investigations on the factor of input in L2 speech acquisition, this study also aimed to assess any potential effects of exposure to the foreign language (FL) and of periods of residence abroad on temporal measures of oral fluency.

2. METHOD

2.1. Participants

The sample of L2 participants in this study comprised 47 Catalan/Spanish bilinguals (41 female, 6 male). They were undergraduate students in English Teacher Education at the University of Barcelona (mean age = 23.1 years). They differed in age of onset of English learning – before, at, or after age 8 – and in exposure to English in their formal learning context – school exposure only vs. extra exposure through language courses – but all with a minimum of 7 years of formal instruction. The Catalan/Spanish bilinguals further varied in their degree of experience in English, as a small subgroup (6 participants) had lived in an English-speaking environment for a period longer than three months. As a control group, five Canadian English native speakers (NSs) (4 female, 1 male) were included in the study. They were undergraduate students in Linguistics at the University of Ottawa and reported having limited proficiency in other languages (if any), particularly French.

2.2. Materials and Procedure

For this study, participants carried out a 6-picture frame narrative task. They were given 1-2 minutes to look at the pictures and then they proceeded to tell the story (see Sánchez & Jarvis 2008, for further details on the picture narrative). Participants were recorded in a quiet room at the phonetics lab by means of a CASIO DA-7 DAT tape recorder and a YU-Brother EM-106 microphone. The resulting recordings were transferred in raw format onto a computer and later saved as 48 kHz, mono, 16-bit WAV files with sound-editing software *CoolEdit2000*. The narratives were fully transcribed in standard orthography. Filled and silent pauses, together with instances of false starts, repetitions, self-corrections, and reformulations, were coded following CHILDES conventions (MacWhinney 2000).

2.3. Analyses

Two excerpts from the narratives were selected for analysis. As in previous studies (e.g., Derwing et al. 2004), an excerpt corresponding to the beginning of the narrative was chosen so that the content was similar across participants. In this case, Excerpt 1 was 15 seconds long on average. The second excerpt included the answer given to the question '*what are the children going to do now?'* that the participants were asked once they had finished telling the story. Excerpt 2 was shorter in duration – about 8 seconds – and it was selected for analysis because it did not involve any prior planning time (as was the case of Excerpt 1). As in other L2 fluency studies (Cucchiarini, Strik & Boves 2002; Kormos 2006; Trofimovich & Baker 2006), several temporal measures of oral fluency in the two excerpts were obtained. Specifically, speech rate (syllables per second), pruned speech rate (pruned syllables per second), pause frequency, and pause duration were examined. It should be further noted that any pause longer than 100 ms was considered in the current investigation, following Trofimovich and Baker (2006). Additional measurements of extemporaneous speech, which involved acoustic analyses of vowels and consonants in Praat (Boersma & Weenink 2007), were also made, but will not be examined in detail in this paper.

Statistical analyses were performed on PASW Statistics 17. The temporal measures of oral fluency obtained were submitted to Kruskal-Wallis analyses with varying degrees of experience in English (both in the formal learning context and as a result of periods of residence abroad) and age of onset as factors. Significant differences between groups were further explored through Mann-Whitney *U* tests. The alpha level was set at .05.

3. RESULTS

Mean values for the temporal measures of oral fluency obtained in the two narrative excerpts are shown in Table 1. First, it can be observed that both groups – Canadian English and Catalan/Spanish speakers – behaved similarly when Excerpt 1 and Excerpt 2 were considered. That is, in Excerpt 2 speech rate and pruned speech rate were higher, and there were fewer and shorter pauses than in Excerpt 1. However, the two participant groups differed significantly from each other in all the fluency measures examined. Catalan/Spanish bilinguals' speaking rate in both excerpts was slower than that of Canadian English NSs. In addition, Catalan/Spanish bilinguals paused more often and their pauses were twice as long as compared to those of the English control group.

Table 1: Averaged temporal measures of oral fluency in Excerpts 1 and 2 for Canadian English native speakers (CanE NSs) and Catalan/Spanish (C/S) bilinguals. (Standard deviations are in parentheses and ranges in square brackets.)

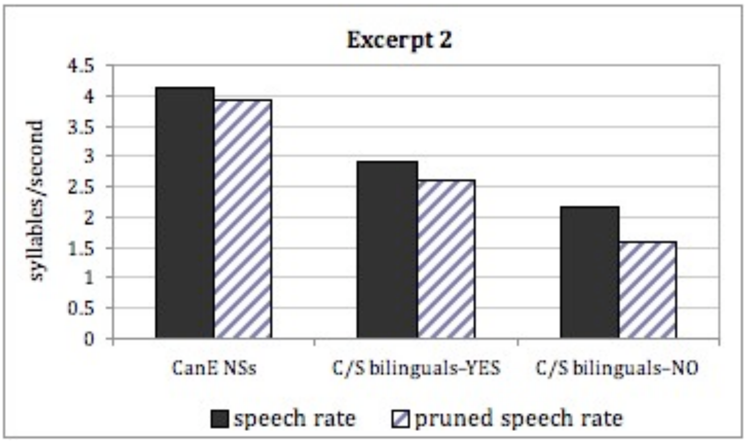
TEMPORAL MEASURES	EXCERPT 1		EXCERPT 2	
	CanE NSs	**C/S bilinguals**	**CanE NSs**	**C/S bilinguals**
Speech rate (syll/sec)	3.63 (0.5)	2.30 (0.5)	4.12 (0.9)	2.30 (0.8)
Pruned syll/sec	3.60 (0.5)	1.58 (0.5)	3.91 (0.8)	1.75 (0.9)
Pause frequency	4 [3–8]	6.7 [3–13]	3 [2–3]	2.89 [0–7]
Pause duration (ms)	218 (115.8)	550 (161)	240 (52.4)	561 (321)

As far as degree of experience in English was concerned, Kruskal-Wallis analyses showed that there were significant differences in speech rate, pruned speech rate, and frequency of pauses in both excerpts (X^2 between 6.047 and 13.399, $p < .05$). In all cases, Catalan/Spanish bilinguals exposed to English only in the school setting and those bilinguals with extra exposure had significantly lower speech rates and made more and longer pauses. When the two exposure groups were compared, Mann-Whitney *U* tests revealed no differences (*U* between 47 and 122.5, $p > .05$). In fact, both groups obtained similar results for Excerpt 1, e.g., pruned speech rate of 1.5 and 1.6 and average pause duration of 550 ms and 520 ms for the extra exposure and no extra exposure groups, respectively. In contrast, and despite the lack of significant differences, Catalan/Spanish bilinguals with extra exposure to English tended to have both higher speech rate and pruned speech rate (mean values of 2.3 and 1.8 vs. 2.1 and 1.6) and shorter pauses on average (467 ms vs. 748 ms) in Excerpt 2.

The observed advantage of a higher degree of experience in English became significant when the variable of periods of residence abroad was looked at. Therefore, participants with a period of residence abroad longer than 3 months obtained a significantly higher pruned speech rate in Excerpt 2 than Catalan/Spanish learners of English who had not lived in an English-speaking environment (mean values of 2.62 and 1.58, respectively) (*U* 43, $p < .05$). This is shown in Figure 1 below. Pruned speech rate and speech rate were also higher – along with shorter pauses – in Excerpt 1 for those learners who had had periods of residence abroad. However, the latter differences did not turn out to be significant (*U* between 47 and 122.5, $p > .05$). Furthermore, irrespective of the tendencies and significant differences noted for Catalan/Spanish bilinguals with periods of residence abroad, all the Catalan/Spanish participants differed significantly from Canadian English NSs in all the fluency measures examined.

Finally, the same significant differences between native English speakers and learners of English were reported for the factor of age of onset (X^2 between 10.647 and 13.519, $p < .05$). Additionally, Catalan/Spanish bilinguals obtained very similar values for temporal measures of oral fluency as a function of age of onset, resulting in nonsignificant differences.

Figure 1: Average speech rate and pruned speech rate for Excerpt 2 as a function of periods of residence abroad. NOTE: CanE NSs = Canadian English NSs; C/S bilinguals–YES: Catalan/Spanish bilinguals with periods of residence abroad longer than 3 months; C/S bilinguals–NO: Catalan/Spanish bilinguals without periods of residence abroad longer than 3 months.



4. DISCUSSION AND CONCLUSIONS

This study looked at oral fluency in 47 Catalan/Spanish learners of English who had been exposed to English as a foreign language in a formal learning context. Temporal measures of fluency – namely, speech rate, pruned speech rate, pause frequency, and pause duration – were obtained in two narrative excerpts. Results of low speech rate and pruned speech rate demonstrated a low level of fluency in English among the Catalan/Spanish bilinguals. These results corroborate previous findings from formal instruction contexts (e.g., Mora 2006).

The low level of oral fluency was further illustrated by the high number of pauses made by the Catalan/Spanish bilinguals and the average long duration of those pauses, in contrast to those of a control group of 5 Canadian English native speakers. However, while pausing phenomena contributed to low speech rates, an additional explanation might account for the low speech rate and pruned speech rate of Catalan/Spanish learners of English. As mentioned above, acoustic measurements of vowel and consonant segments were made in Excerpts 1 and 2. In general, Catalan/Spanish bilinguals tended to overgeneralize (or sometimes overdo) the use of lengthening both in both voiced and voiceless consonants in word-final position as well as in tense vs. lax vowels (Fullana & MacKay 2009). As a result, syllables were longer in duration, probably leading to the production of a lower number of syllables per second, unlike English NSs. This suggestion calls for further investigation. Similarly, more insight into Catalan/Spanish bilinguals' fluency in English might be gained if the values of vowel duration were correlated with listeners' fluency ratings, as recently shown by Derwing, Munro, Thomson, and Rossiter (2009). Moreover, further research on the narratives should explore other measures of fluency that have been shown to correlate with listeners' perceptions of fluency, such as phonation-time ratio and articulation rate. (e.g., Cucchiarini et al. 2002; Kormos 2006).

The results of the present investigation also showed that exposure to English was not a determining factor in attaining a more native-like degree of fluency in the TL. Nor was age of onset of English learning, which did not coincide with the findings in Mora (2006). Despite this fact, fluency was slightly better for those individuals who had stayed in an English-speaking environment for three months or more. In addition, the finding of a higher pruned speech rate in Excerpt 2 vs. Excerpt 1 – both for Canadian English native speakers and those individuals with periods of residence abroad – deserves further attention. As noted above, Excerpt 2 did not involve any planning time as opposed to Excerpt 1. It has been suggested that planning time leads to a higher level of fluency (e.g., Foster & Skehan 1996, as cited in Derwing et al. 2004). However, in the current study the opposite trend was observed. Further research into oral fluency should look into equally long stretches of speech involving planning time and lack of it.

Furthermore, these fluency results evidenced a high degree of variability, as illustrated by the high standard deviations in Table 1 and the subsequent use of nonparametric tests. This corroborates previous findings of existing individual differences and variability that might be accounted for by considering L2 speech learning as a "context-dependent, approximative, frequency-based process" (Munro 2008: 141).

Finally, in light of the long period of formal instruction these learners were exposed to and together with recent suggestions in the literature (e.g., Flege 2009), the findings of this study bring into focus the need for methods to assess both the quality and quantity of input in formal learning contexts, in addition to conducting further research on the effects of periods of residence abroad in the absence of phonetically relevant input in formal instruction learning contexts.

5. ACKNOWLEDGEMENTS

Work supported by a postdoctoral fellowship to the first author from the *Secretaría de Estado de Universidades e Investigación del Ministerio de Educación y Ciencia* and the *Fundación Española para la Ciencia y la Tecnología* (Spain). The authors would like to thank Maurice Bélanger (Linguistics Research Laboratory, University of Ottawa), the *Laboratori de Fonètica Aplicada* (*Universitat de Barcelona*, Spain) and all the participants in the study.

6. REFERENCES

Best, C. T. 1995. A direct realist view of cross-language speech perception. In: Strange, W. (ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press. 171–204.

Best, C. T., Tyler, M. D. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In: Bohn, O.-S., Munro, M. J. (eds.), *Language experience in second language speech learning: In honor of James Emil Flege*. Amsterdam/Philadelphia: John Benjamins. 13–34.

Boersma, P, Weenink, D. 2007. Praat (Version 4.6.05). Retrieved June 2, 2007, from www.praat.org.

Cebrian, J. 2003. Input and experience in the perception of an L2 temporal and spectral contrast. In: Recasens, D., Solé, M- J., Romero, J. (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences.* Barcelona/Australia: Causal Productions. 2297–2300.

Cebrian, J. 2006. Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics* 34. 372-387.

Cucchiarini, C., Strik, H., Boves, L. 2002. Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech. *Journal of the Acoustical Society of America* 111. 2862–2873.

Derwing, T. M., Munro, M. J., Thomson, R. I. 2008. A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics* 29. 359–380.

Derwing, T. M., Munro, M.J., Thomson, R. I., Rossiter, M. J. 2009. The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition* 31. 533-557.

Derwing, T. M., Rossiter, M. J., Munro, M. J., Thomson, R. I. 2004. Second language fluency: Judgments on different tasks. *Language Learning* 54. 655–679.

Flege, J. E. 1995. Second language speech learning: Theory, findings and problems. In: Strange, W. (ed.), *Speech perception and linguistic experience. Issues in cross-language research*. Timonium, MD: York Press. 229–273

Flege, J. E. 2009. Give input a chance!. In: Piske, T., Young-Scholten, M. (eds.), *Input matters in SLA*. Clevedon: Multilingual Matters. 175–190.

Freed, B. F. 1995. What makes us think that students who study abroad become fluent? In: Freed, B. F. (ed.), *Second language acquisition in a study abroad context*. Amsterdam: John Benjamins. 123–148.

Fullana, N. (2006). The development of English (FL) perception and production skills: Starting age and exposure effects. In: Muñoz, C. (ed.), *Age and the rate of foreign language learning*. Clevedon: Multilingual Matters. 41–64.

Fullana, N., MacKay, I. R. A. 2009. Modelling Catalan/Spanish bilinguals'production of English extemporaneous and prompted speech in a formal learning context. Paper given at ACLA 2009 Conference, Ottawa, Canada, May 2009.

García-Lecumberri, M. L., Gallardo, F. 2003. English FL sounds in school learners of different ages. In: García-Mayo, M. P., García-Lecumberri, M. L. (eds.), *Age and the acquisition of English as a foreign language*. Clevedon, England: Multilingual Matters. 115_135.

Hahn, L. D. 2004. Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly* 38. 201–223.

Kormos, J. 2006. *Speech production and second language acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Kormos, J., Dénes, M. 2004. Exploring measures and perceptions of fluency in the speech of second language learners. *System* 32. 145-164.

MacWhinney, B. 2000. *The CHILDES Project: Tools for Analyzing Talk*. (3rd ed.) Mahwah, NJ: Lawrence Erlbaum Associates.

Mora, J. C. 2006. Age effects on oral fluency development. In: Muñoz, C. (ed.), *Age and the rate of foreign language learning*. Clevedon: Multilingual Matters. 65–88.

Mora, J. C. 2008. Learning context effects on the acquisition of a second language phonology. In: Pérez-Vidal, C., Juan-Garau, M., Bel, A. (eds.), *A portrait of the young in the new multilingual Spain*. Clevedon: Multilingual Matters. 241–263.

Munro, M. J. 2008. Variability in Cantonese speakers' production of English vowels. *Canadian Acoustics* 36. 140–141.

Munro, M. J., Derwing, T. M., Morton, S. L. 2006. The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition,* 28. 111–131.

O'Brien, I., Segalowitz, N., Freed, B., Collentine, J. 2007. Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition* 29. 557-582.

Piske, T., Young-Scholten, M. (eds.). 2009. *Input matters in SLA*. Clevedon: Multilingual Matters.

Rallo-Fabra, L. 2009. Discrimination of four English vowel contrasts by Catalan learners varying in language experience. Paper given at ASA09 Cross-Language Speech Workshop, Portland, Oregon, May 21-23.

Rossiter, M. J. 2009. Perceptions of L2 fluency by native and non-native speakers of English. *The Canadian Modern Language Review* 65. 395–412.

Sánchez, L., Jarvis, S. 2008. The use of picture stories in the investigation of crosslinguistic influence. *TESOL Quarterly* 42. 329–333.

Segalowitz, N. 2007. Access fluidity, attention control, and the acquisition of fluency in a second language. *TESOL Quarterly* 41. 181–186.

Trofimovich, P., Baker, W. 2006. Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition* 28. 1–30.

Phonetic cue-weighting in the acquisition of a second language (L2): evidence from Greek speakers of English.

Anastasia Giannakopoulou[1], Maria Uther[1], Sari Ylinen[2]*

[1] Centre for Cognition and Neuroimaging, Department of Psychology, Brunel University, Uxbridge, Middlesex, United Kingdom.

[2] Cognitive Brain Research Unit, Institute of Behavioural Sciences, University of Helsinki, Finland.

[*]anastasia.giannakopoulou@brunel.ac.uk

ABSTRACT

Speech sounds containing multiple phonetic cues are often the most difficult speech-sound contrasts for foreign-language learners, especially if certain cues are weighted differently in the foreign and native languages. Greek adult and child speakers of English were studied to determine on what basis they were making discrimination and identification judgements between English vowels. The Greek vowel inventory consists of five vowels that are distinguished by spectral changes, whereas the English vowel inventory involves both spectral and duration cues. The use of spectral and duration cues of English vowel minimal pairs in Greek speakers were studied using two forms of perception and discrimination tasks: one used natural stimuli and the other used matched vowel duration words to ‘force’ the use of spectral cues. Results of the pre-training task show performance was impaired for Greek speakers across both perceptual identification tasks compared to ceiling effects by the native English speakers. High-variability training sessions were used to enhance vowel perception. Following the use of a 2 week modified high-variability phonetic training program, performance improved for both Greek adult and child groups as revealed by post training tests. However the effects were most pronounced for the child Greek speaker group. The results are discussed in terms of current theories of spoken language acquisition.

Keywords: speech perception, cue-weighting, second language, perceptual training.

1. INTRODUCTION

A number of cues need to be weighted and integrated in order for a speech sound to be correctly identified (Holt & Lotto, 2006). Native speakers of a language learn to weight cues in way that those cues carrying more critical information are weighted higher than others (Holt & Lotto 2006). A second-language (L2) learner’s native language (L1), however, may inhibit correct cue weighting in the L2 by a process known as L1 transfer (Bohn 1995; Strange 1998). The acoustic cues that L2 learners use are often different compared to those used by L1 speakers, thereby leading to inferior performance on perceptual categorization tasks (Iverson et al. 2003). Previous research also suggests that the dissimilarity of cue weighting between an L1 and L2 could result in sounds being assimilated into the same L1 sound category (Best 1995). This study aims to explore whether the L1 affects L2 cue weighting in the case of Greek L2 speakers of English and whether specialised training could allow the L2 learners for correct cue weighting in perceptual identification and discrimination of L2 phonetic segments by shifting perception into relevant and critical cues.

The dimensions of English tense-lax vowel contrasts were examined here in cases where both temporal and spectral changes were involved in the distinction (e.g. /i/ vs. /I/). By comparison, Greek vowels involve a less complex interplay as only spectral dimensions are present in Greek vowel system and duration is not phonemically relevant (Fourakis et al. 1999). In a study by Gottfried and Suiter (1997), English L2 learners of Mandarin had difficulty learning lexical tone which is not phonemically relevant in their L1 but instead they had almost no difficulty learning Mandarin vowel quality, which is a feature they are used to using in their L1. For Greek L2 speakers of English, this would result in the ability to identify contrastive phonological segments on the basis of the spectral qualities of the L2 vowels, whereas duration could be an irrelevant or misleading cue. On the other hand, another view suggests that L2 speakers may as well attend to cues that are not existent in their L1. Indeed, according to Bohn (1995), duration can be easily introduced as a cue in vowel perceptual contrasts even though it may not be present phonologically in the learner’s L1.Therefore, Greek L2 learners of English might still access duration cues despite the fact that it would not result from L1 transfer. Given the various theories of speech acquisition that emphasize the role of age of acquisition (Kuhl et al. 2008; Flege 2002), we also examined performance of different L2 age groups to determine whether there were similar patterns of difficulty in adult vs. child learners and explore maturational differences in acquisition of L2 phonetic segments. Finally, we also examined perceptual improvement as a result of training using the high-variability training technique (Logan et al. 1991; Yamada 1993; Pisoni et al. 1994) used in several previous studies to improve performance of phonemic identification (e.g. Bradlow et al. 1997, 1999; Lively et al. 1993; Pisoni et al. 1994; Uther et al., 2007; Iverson & Evans 2009); Ylinen et al. 2010).

2. METHOD

Two experiments were designed: the first to investigate perception in two different age groups, and the second to explore how training can improve perceptual performance of L2 speech sounds.

2.1. Experiment 1

2.1.1 Participants

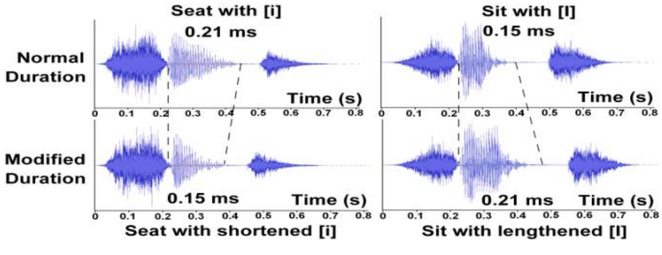
Adult groups: Twenty adult native speakers of Standard Modern Greek (8 female, 12 male) aged 20-30 (mean age = 25.4) were recruited from the University of Patras, Greece. They had all studied English as L2 and their level of proficiency was advanced (L2 English education mean = 7.8 years). Twenty monolingual English native speakers (14 female, 6 male) aged 19-26 (mean age = 21.4) served as controls. They were recruited from Brunel University, West London, UK. *Child groups*: Fifteen child native speakers of Standard Modern Greek (8 female, 7 male), aged 7-8 (mean age = 7.6), were tested. They had all studied English as L2 and their level of proficiency was basic (L2 English education mean = 1.2 years). Fifteen child monolingual native speakers of Standard English (6 female, 9 male) aged 7-8 (mean age = 7.4) served as controls. They were recruited from Brunel University and surrounding schools. All participants had normal or corrected to normal vision and none reported any history of a speech or hearing impairment.

2.1.2 Stimuli and apparatus

Forty-five minimal pairs (e.g. *sit* versus *seat*) with the English tense-lax vowel distinction were used. The auditory stimuli were pronounced by a male native speaker of English representing typical Southern British English pronunciation. A second set of minimal pairs was created where the normal vowel duration was digitally manipulated with Praat software (Boersma and Weenink 2004) so that for each minimal pair the intrinsically long /i/ vowel was shortened to match the duration of the intrinsically shorter /I/ vowel and vice versa (see Figure 1). Vowel modifications were resynthesized with the Pitch-Synchronous Overlap and Add (PSOLA) technique (Boersma and Weenink 2004). In these modified stimuli, the spectral quality of vowels was preserved even though duration cues were not available (see Figure 1).

Stimuli were presented on a laptop with E-Prime software (Schneider et al. 2002a,b). All auditory stimuli were binaurally presented through high quality headphones at a comfortable listening level (varying between 65-75 dB).

Figure 1: An example of the stimuli used. Top: The minimal pair ‘seat’ and ‘sit’ with normal duration (i.e., as ordinarily pronounced by a native English speaker). Bottom: The same minimal pair with modified vowel durations: [i] was shortened to correspond to the original duration of [I], and [I] was lengthened to correspond to the original duration of [i].



2.1.3 Procedure

Experiment 1 included two tasks with two conditions. The tasks and stimuli within each task were presented in random order. Participants were tested individually. They completed a language background questionnaire, gave written consent and participated in all tasks. Each session lasted approximately 45 minutes.

Perceptual Identification Task (PI): This task used 45 normal vowel duration and 45 modified vowel duration minimal pairs of English words arranged into two conditions (natural and modified duration stimuli). Each trial consisted of an auditory stimulus (i.e. the auditory form of one of the minimal pair counterparts) presented through the headphones and a simultaneous visual stimulus (i.e. the orthographic form of the minimal word pair, e.g. sit – seat) presented on the screen. In both conditions, each auditory word was presented once in random order, accompanied by the visual stimulus consisting of the orthographic representation of the corresponding minimal pair on the screen. Thus, 90 stimuli were presented for each condition, 180 in total. Participants were instructed to choose which one of the two words presented on the screen they heard through pressing a relevant key on the computer keyboard.

Auditory Discrimination Task (AB-X): The stimuli identical to those of the previous task were used but they were presented only aurally as a discrimination task, arranged in an ABX format (i.e. ABA or ABB). Each trial comprised a sequence of a minimal pair words (word A followed by word B or word B followed by word A) and a third stimulus being the exact repetition of either word A or word B. As for the previous identification task, two conditions were included: a natural stimuli condition and a modified duration condition (with stimuli matched in duration to the other word in the minimal pair). Participants were instructed to respond by pressing relevant keys on the computer keyboard whether the third word on each trial was same as the first word or same as the second word in the auditory sequence.

2.2. Experiment 2

2.2.1. Participants

Adult groups: Ten native adult speakers of Standard Modern Greek (7 female, 3 male) aged 20-30 (mean age = 24.6) were tested. They had all studied English as L2 and their level of proficiency was advanced (L2 English education mean = 8.6 years). *Child groups*: Ten child native speakers of Standard Modern Greek (5 female, 5 male), aged 7-8 (mean age = 7.9) were tested. They had all studied English as L2 and their level of proficiency was basic (L2 English education mean = 1.4 years). All participants had normal or corrected to normal vision and none reported any history of a speech or hearing impairment.

2.2.2 Stimuli and apparatus

The auditory stimuli used in the training sessions were pronounced by 4 different native British English speakers (2 male and 2 female speakers) considered to represent typical English pronunciation. Auditory stimuli were modified following same procedure as described for the auditory stimuli of Experiment 1 (for natural and modified duration stimuli alike) and were binaurally presented at a comfortable listening level (65-70 dB). The 4 speakers used for the training stimuli were different than the speaker who pronounced the stimuli for the pre- and post-training test. The equipment was otherwise identical to that used in Experiment 1.

2.2.3 Procedure

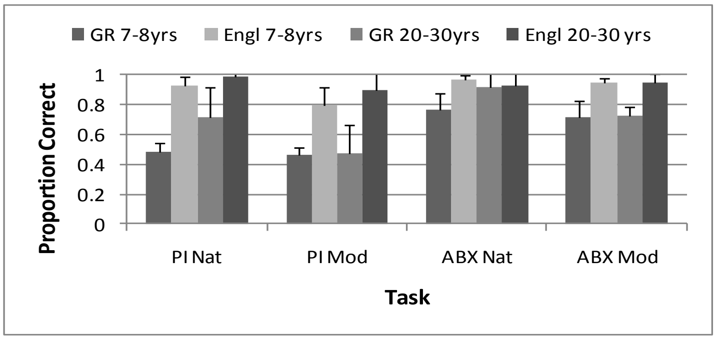
Experiment 2 involved three stages for the Greek (adult and child) participants: pre-test, training, and post-test. The pre- and post-test sessions were identical and consisted of the same tasks as described in Experiment 1. The training program was based on the high-variability perceptual training procedures developed by Logan *et al.* (1991), extended by Yamada (1993) and others (e.g. Bradlow et al. 1997, 1999; Lively et al. 1993; Pisoni et al. 1994; Ylinen, et al. 2010). High variability refers to the use of multiple speakers and multiple phonetic contexts. For this experiment, the training phase consisted of 10 x 30 minute sessions (one session per weekday) in a 2-week period. Like the pre-training testing paradigm, each trial consisted of an auditory stimulus (one of the words of a minimal pair) and the requirement to select a choice from a pair of visually presented minimal pair of words. Each training session had 304 trials (different combinations of speaker and word type) presented in random order. Unlike in the test sessions, the participants received feedback on each trial and had the option to replay the auditory stimulus if needed. 'Correction' trials were also included: these additional trials were given in the case of an incorrect response (with the visual positions of the minimal word pair randomly reassigned on the next trial to avoid guessing). Training involved two conditions: natural vowel training and modified vowel training. Five participants were trained in each condition. This aimed to show whether it was necessary to train on modified duration stimuli in order to achieve improvements in performance with the same stimuli.

3. RESULTS

3.1 Experiment 1

For the perceptual identification task, there was an effect of age with adult participants performing better compared to child participants (see Figure 2; $F_{(1, 66)}$= 30.255, p< .001). Greeks also performed worse overall compared to English speakers (see Figure 2; $F_{(1, 66)}$=365.234, p, .001).

Figure 2: Accuracy scores for Greek child, Greek adult, English child and English adult across all tasks.



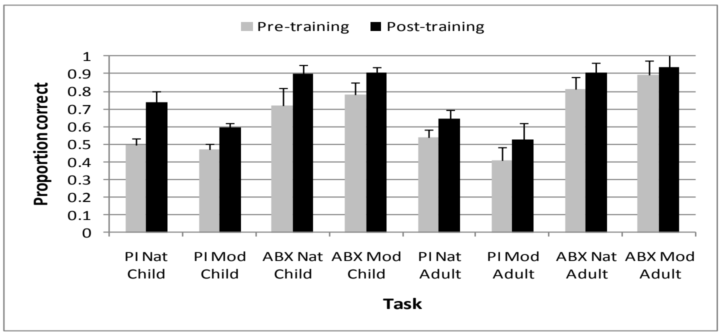
Identification of words with natural vowel length was better than those with modified duration vowels ($F_{(1,66)}$=110.668, p<.001) confirming the assumption that vowel duration was indeed used a critical cue for perceptual identification performance. For child participants there was less difference between natural and modified duration words compared to adult participants ($F_{(1,66)}$=13.190, p< .05). Results also showed that Greek adult and child participants' performance was impaired compared to both English groups. In comparison with the other groups, a larger difference in performance between natural and matched vowel duration words was observed in Greek adults ($F_{(1,66)}$=31.711, p< .001). They had higher scores for the natural vowel duration words suggesting that Greek adults tend to rely more on vowel duration compared to Greek children or English groups.

For the AB-X discrimination task, adults performed better compared to children ($F_{(1, 66)}$= 7.819, p< .001) and Greeks also performed worse than English speakers ($F_{(1, 66)}$=146.922, p< .001). Discrimination of words with natural vowel duration was overall better compared to words with matched vowel duration ($F_{(1, 66)}$= 8.05, p<.001) with Greeks performing significantly better with natural vowel duration words ($F_{(1, 66)}$=11.097, p<.05). Child participants revealed lower scores overall compared to adult participants with most pronounced difference between Greek child and adult participants ($F_{(1, 66)}$=14.654, p< .001). All groups performed better with the natural vowel duration words compared to matched duration words but the largest difference was revealed with Greek adults whose performance with words of matched vowel duration was impaired ($F_{(1, 66)}$=10.806, p< .001). No other main effects or interactions were significant.

3.2 Experiment 2

Both Greek adult and child groups reveal significant improvement following the training intervention for the perceptual identification task (see Figure 3; $F_{(1, 16)}$=67.493, p<.001) with Greek children showing higher improvement compared to Greek adults (see Figure 3; $F_{(1, 16)}$=6.411, p<.05). Perceptual identification of words with natural vowel duration was overall better compared to words with modified vowel duration ($F_{(1,16)}$=28.514, p<.001).

Figure 3: Training results for child and adult participants showing pre and post training effects for natural (Nat) and modified (Mod) duration training conditions.



The stimuli used to train each group affected participants' performance as groups that trained with modified duration stimuli improved more in comparison to groups trained with natural duration stimuli ($F_{(1,16)}=4.864$, $p<.05$). Overall, training improved perceptual identification for both natural and matched duration words ($F_{(1,16)}=5.858$, $p< .05$). For the AB-X discrimination task the training intervention also revealed significant improvement ($F_{(1,16)}=52.906$, $p<.001$). Discrimination of words with modified vowel duration was overall better compared to words with natural vowel duration ($F_{(1,16)}=8.022$, $p<.05$) which is a mirror opposite result to the perceptual identification task results. Both groups improved but child participants showed the most pronounced improvement in the post training tests ($F_{(1,16)}=7.770$, $p<.05$). Training overall improved discrimination performance for both stimulus types ($F_{(1,16)}=4.615$, $p<.05$). No other main effects or interactions were significant.

4. DISCUSSION

The primary goal of this study was to examine the identification and discrimination of L2 vowels with cues that were not phonemically relevant in the L1 and investigate whether targeted perceptual training could help to improve L2 cue weighting specifically for Greek L2 speakers of English. It is clear that L2 cues are weighted differently between Greek L2 speakers of English and native English speakers. Experiment 1 demonstrates that although native English speakers use spectral features as the primary cue and duration as secondary cue, Greek speakers of L2 English use duration as the primary cue while performance is impaired for those tasks that identification and discrimination is based solely on spectral cues. Interestingly, it appears that the Greek speakers use duration cues in the recognition of English vowels despite the fact that duration is not phonemically relevant in Greek. This holds true for both Greek adult and child groups although Greek adults show a greater tendency to rely on duration cues. Thus, cues that are not phonemically relevant in the L1 may be perceptually identified and treated as primary cues in the identification and discrimination of L2 vowel quality. The results also suggest that duration was used as a cue for non-native speakers, supporting the data reported by Bohn (1995) that even though a duration cue may not be phonemically relevant in the L1, it can still be used for perceptual identification in the L2.

The second finding, as revealed in experiment 2, is that high variability perceptual identification training can alter perceptual cue weighting. Specifically, targeted training where certain cues are removed (e.g. duration cues) and participants are 'forced' to attend to relevant (e.g. spectral) cues appears to be a highly successful approach to L2 learning. The perceptual training procedure reveals improvement across all tasks and for both Greek adult and child groups. It appears that the combination of multiple talkers used for the stimulus presentation as well as the minimal pair arrangement that involved an identification task type form a robust method for perceptual identification and learning. This method has been used effectively for Japanese L2 speakers of English (e.g. Magnuson et al. 1995; Uther et al. 2007) and Finnish (Ylinen et al. 2010) and from these data, we can see it also seems to help Greek L2 speakers of English. The stimulus types used throughout the training sessions (natural versus modified duration stimuli) also appear to affect post training performance. The post training perceptual identification task showed more improvement with the natural duration words compared to modified duration words whereas the opposite effect was true for the discrimination task. In this case, further experiments are needed to validate this result. Our final finding is that the training intervention also shows age differences when it comes to training child versus adult groups.

Although, training significantly improved performance for both groups, the degree of improvement was larger for the child than the adult group, even despite differences in years of English language education. This would suggest children are more easily susceptible to inducing plastic changes in their phonetic categories. These data could be interpreted in terms of speech learning models that suggest a consolidation of neural and perceptual phonetic categories with increasing age (Kuhl 2008; Flege 2002).

5. REFERENCES

Best, C.T. (1995), A direct realist view of cross-language speech perception. In W. Strange (eds), *Speech Perception and Linguistic Experience: Issues in cross-language research.* Timonium, MD:York Press, 171-204.

Boersma, P. & Weenink, D. (2004). Praat: doing phonetics by computer (Version 4.2.17) [Computer program]. Retrieved from http://www.praat.org/.

Bohn, O.-S. (1995). Cross language speech perception in adults: first language transfer doesn't tell it all. In: *Speech perception and linguistic experience: Theoretical and methodological issues in cross language research*. W. Strange, (Ed.), Timonium, MD: York Press. 279–304.

Bradlow, A.R., Pisoni, D.B., Akahane,-Yamada, R., Tohkura, Y. (1997). Training Japanese listeners to identify English /r/and /l/: IV. Some effects of perceptual learning on speech production. Journal of the Acoustical Society of America. 101(10), 2299-2310.

Bradlow, A.R., Yamada, R.A., Pisoni, D.B., Tohkura, Y. (1999). Training Japanese listeners to identify English /r/and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*. *61 (5), 977-985*

Flege, J. E. (2002). Interactions between the native and second-language phonetic systems. In P. Burmeister, T. Piske, & A. Rohde (Eds.). *An integrated view of language development*. *Papers in honor of Henning Wode*. Trier: Wissenschaftlicher Verlag. 217–244.

Fourakis, M., Botinis, A., Katsaiti, M. (1999). Acoustic Characteristics of Vowels. *Phonetica*. 56: 28-43.

Gottfried, T. L. & Suiter, T. L. (1997) Effect of linguistic experience on the identification of Mandarin tones, *Journal of Phonetics*, 25, 207–231.

Holt, L.L. & Lotto, A.J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *The Journal of the Acoustical Society of America*, 119, 3059-3071.

Iverson, P., Evans, B.G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *Journal of the Acoustical Society of America* 122(5), 2842-2854. ISSN: 0001-4966.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition, 87,* B47-B57.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M. & Nelson, T. (2008). Early phonetic perception as a gateway to language: New data and Native Language Magnet Theory, expanded (NLM-e). Philosophical Transactions of the Royal Society B, 363, 979-1000.

Lively, S. E., Logan, J.S., Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories. Journal of the Acoustical Society of America. 94(3), 1242-1255.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/: III. Long-term retention of new phonetic categories. Journal of the Acoustical Society of America, 96, 2076–2087.

Logan, J. D., Lively, S. E., and Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America*, 89**,** 874–886.

Magnuson, J.S., Yamada, R.A., Tohkura, Y., Pisoni, D.B., Lively, S.E., Bradlow, A.R. (1995). The role of talker variability in non-native phoneme training. *Proceedings of the 1995 Spring Meeting of the Acoustical Society of Japan* (pp. 393-394). Tokyo: Acoustical Society of Japan.

Pisoni, D.B., Lively, S.E., & Logan, J.S. (1994). Perceptual learning of non-native speech contrasts: Implications for theories of speech perception. In J. Goodman & H. C. Nusbaum (Eds.), *Development of speech perception: The transition from speech sounds to spoken words.* (pp. 121 – 166). Cambridge, MA: MIT Press.

Schneider,W., Eschman, A. & Zuccolotto, A., (2002b). E-Prime Reference Guide. Pittsburgh: Pychology Software Tools Inc. 2002b.

Schneider,W., Eschman, A. & Zuccolotto, A.. (2002a). E-Prime User's Guide. Pittsburgh: Pychology Software Tools Inc. 2002a.

Strange, W. (1998). Perceptual assimilation of American English vowels by Japanese listeners. *Journal of Phonetics*, 26, 311–344.

Uther, M., Uther, J., Athanasopoulos, P., Singh, P. & Akahane-Yamada, R. (2007). Mobile Adaptive CALL (MAC). A lightweight speech-based intervention for mobile language learners. Proceedings of Interspeech 2007 conference, Antwerp, Belgium, August 27-31, 2007.

Yamada, R. A. (1993). Effects of extended training on /r/ and /l/ identification by native speakers of Japanese. *The Journal of the Acoustical Society of America,* 93**,** 2391(A).

Ylinen, S., Uther, M, Latvala, A., Vepsäläinen, S., Iverson, P., Akahane-Yamada, R., and Näätänen, R. (2010). Training the Brain to Weight Speech Cues Differently: A Study of Finnish Second-language Users of English. *Journal of Cognitive Neuroscience*, Vol. 22, No. 6, Pages 1319-1332: 1319-1332.

Investigating training effects in the production of English weak forms by Spanish learners

Esther Gómez Lacabex & María Luisa García Lecumberri

University of the Basque Country
esther.glacabex@ehu.es
garcia.lecumberri@ehu.es

ABSTRACT

The present study investigated the production of English weak forms by Spanish speakers before and after specific phonetic training on perception and production. The productions of 34 Spanish learners of English as a foreign language were tested in two contexts: reading aloud and imitation. Results revealed a significant improvement in the production of weak forms after training in both reading aloud (gain score: 7.8%) and imitation tasks (8.5%), Whilst these results indicate that instruction had a beneficial effect, post-test production of weak forms was still only judged correct 50% of the time. This suggests that longer training maybe necessary. Significant differences were observed between the reading and the imitation tasks in the pre-test and post-test for both experimental groups, always favouring imitative production. Our results agree with previous findings on the facilitating relationship between perception and production training and provide further evidence for the strong effect of L1 interference in second language speech learning.

Keywords: perception, production, training, English weak forms.

1. INTRODUCTION

The positive effect of phonetic training on speech perception and production abilities has been widely attested in the literature. Based on the premise that second language speech perception leads production (Neufeld, 1988; Flege, 1995), many studies have looked into the effects of perceptual training on both perception (Logan, *et al.*, 1991; Lively, *et al*., 1993, 1994; Cenoz & García Lecumberri 1999; Iverson & Bronwen, 2007) and production (Rochet, 1995; Bradlow *et al.*, 1999; Wang, 2000; Aliaga-García & Mora, 2009). Other research has explored the role of audio/visual training on perception (Hazan *et al*., 2006; Hazan & Sennema, 2007) while fewer studies have investigated production training effects on production (Flege, 1988). Finally, a few studies have looked into the potential differences between perceptual and production training (Catford & Pisoni, 1970; Leather, 1990; Gómez Lacabex, 2009) with interesting results. Catford and Pisoni (1970) examined articulatory and perceptual training on exotic sounds with a group of English speakers and measured effects on both production and perception. While both regimes proved effective as the subjects improved in both perception and production, the authors found articulatory training to have a more significant effect on perception than perceptual training. Leather (1990) analysed the effect of training on Chinese lexical tones for production and perception. In this study, one experimental group received computer-based perceptual training whilst another group was trained to use computer-assisted visual feedback in their production. Results suggested that both training regimes had positive effects for perceptual and production abilities. Similar results were found by Gómez Lacabex (2009), who analysed perception and production of English vowel reduction (mainly schwa) by a group of Spanish learners of English after either perceptual or production training. Results did not reveal significant differences between the two groups tested. In sum, the few studies on perception and production training carried out so far have not found consistent differences between these two training approaches, which indicates that both perception and production seem to benefit from each other (‘mutually-facilitating relationship’, c.f. Leather 1990).

The present study further explores these research lines by investigating perception and production training effects on the production of English weak forms by Spanish speakers. Centripetal vowel reduction (cf. Harris, 2005) or the centralization of vowels in unstressed syllables does not affect all languages to a similar extent. Whilst languages such as English exhibit important vowel reduction processes (both phonological and phonetic), in other languages, such as Spanish, vowel reduction is only a phonetic phenomenon (Jaworski, 2008). Existing work on non-native production of vowel reduction has revealed that it is not automatically associated by learners to lack of stress (Flege & Bohn, 1989) and that advanced learners may show awareness of reduced vowels but may not always target the coarticulatory patterns which characterise them in production (Kondo, 1995; Yun & Jackson, 2006).

2. METHOD

2.1. Subjects

The subjects in our study were 34 Spanish teenagers (see Table 1 for distribution of age and gender in the two groups) learning English as a foreign language in a formal context. As well as the 3 hours per week of English instruction at school, they attended English lessons for 3 further hours a week at a language school, where they were exposed to native accents of English. Participants were divided into two groups: group A undergoing perceptual-inductive vowel reduction training and group B receiving productive-deductive vowel reduction training. None of the participants reported any listening or hearing impairment.

Table 1: Distribution of subjects, age and gender for the two groups.

	N	Age *(s.d.)*	*Gender*	
			Male%	*Female%*
A	17	15.6 (*1.1*)	58.8	41.2
B	17	15.8 (*1.3*)	29.4	70.6

2.2. Stimuli

9 sentences containing a grammatical word were presented in embedding sentences. The grammatical words included four auxiliary verbs (*have, has, can, was*), the conjunction *that*, the particle *there* in the existential construction, the preposition *for*, the pronoun *them* and the connector *and*. These grammatical words were presented in weak forms in a set of sentences and in strong forms in other sentences. For the imitation task (IMIT), the audio prompts were previously recorded in a sound-proof booth. The model speaker was an English female talker with a standard Southern British accent. The sentences were recorded with a MATLAB programme and saved as .wav sound files.

2.3. Procedure

Students productions were recorded before and after training in a quiet computer room with a custom designed MATLAB programme and saved as .wav sound files.

For the read aloud elicitation task (READ), subjects read the prompt presented orthographically on the screen in their PCs. For the imitation task (IMIT), subjects listened to the audio prompts with headphones and then repeated the sentences.

Students production were presented to a group of 5 native English speakers to be assessed auditorily. These were university postgraduates with little or no contact with Spanish or Spanish accented English. This auditory test was also run with a MATLAB programme. The interface presented a two-alternative forced-choice task: ‘weak’ and ‘strong’ in which judges were previously trained. Additionally, the choice ‘unassessable’ was also provided in case a judgement could not be provided due to noisy recordings, unintelligible pronunciation, wrong stressing, word/syllables missing etc. Judges listened to whole sentences but were asked and trained to focus only on the vowel sound in the word appearing on their screen (a grammatical word). The programme was designed so that judges could listen to all the productions of each token in one session.

2.4. Training

The experimental groups received two different types of training about the rules governing English weak forms. Experimental group A was presented with a variety of audio stimuli (perceptual training) and was asked to induce rules after having listened to several examples of weak and strong forms. Experimental group B was given the English weak form rules a priori and was given feedback in production practice sessions (deductive approach with production training). The introduction of weak form rules was part of a vowel reduction training programme which also included perception and production practice of phonological schwa in lexical words. The time devoted to weak forms in grammatical words was around 90 minutes distributed in three sessions of 30 minutes each on average (approximately 25% of the total training time of the programme).

3. RESULTS

Results are presented for production of weak forms in the reading aloud (READ) task, in the imitation task (IMIT) and for the comparison of the two tasks. Students' productions were coded as percentages of correct answers and analysed with SPSS 15.1. Wilcoxon non-parametric analyses were computed for intra-group data and U-Mann Whitney non-parametric tests were applied to inter-group analyses.

3.1. Reading aloud task

In the reading aloud task (READ) significant intra-group differences (between pre- and post-test results) (z = -2.82, p<.005) were found for groups A and B analysed together showing that training was effective. The two experimental groups were analysed separately to investigate the effects of treatment type (group A vs. group B). An inter-group analysis of the pre-test results showed no significant differences between groups, which indicated that both groups exhibited similar weak form production prior to training. Despite inter-group differences being larger in the post-test, they were still not statistically significant. Although both groups experimented an improvement in the post test, intra-group analysis revealed significant differences between pre- and post-test only for experimental group A (z = -2.59, p>05) as observed in Table 2. Group B also experimented gains but these were not significant.

Table 2: Mean % of correct productions in READ task, standard deviation *(s.d.)* and significance (*p-value*: ***significant at p<.0001, **significant at p<.005, *significant at p<.05) for weak forms.

9	*Mean %*			*p-value*	
schwa	*pre (s.d.)*	*post (s.d.)*	*gain (s.d.)*	*intra-group*	
A	12.42 *(11.7)*	22.88 *(15.9)*	10.46 *(14.4)*	**	*
B	11.76 *(12.7)*	16.99 *(20.8)*	5.22 *(15.3)*		
p-value inter-group					

3.2. Imitation task

In the imitation task, when experimental groups (A and B) were analysed together, significant intra-group differences between the pre- and post-test (z = -2.46, p<.05) were found, indicating again positive training effects overall. When exploring differences between experimental groups, inter-group analyses showed no significant differences in pre-test or post-test. As in the previous task, both groups showed improvements in the post-test. Intra-group analysis revealed a larger improvement for experimental group A, although the statistical significance (p = .058), was only tendential, as Table 3 reveals.

Table 3: Mean % of correct productions in IMIT task, standard deviation *(s.d.)* and significance (*p-value*: ***significant at p<.0001, **significant at p<.005, *significant at p<.05) for weak forms.

9 schwa	*Mean %*			*p-value*	
	pre (s.d.)	*post (s.d.)*	*gain (s.d.)*	*intra-group*	
A	29.41 *(18.4)*	40.52 *(18)*	11.11 *(20.4)*		.058
B	30.06 *(14.0)*	35.95 *(20.7)*	5.88 *(16.7)*	*	
p-value inter-group					

Although it was observed that experimental group A experienced a higher improvement than experimental group B in both the reading and in the imitation tasks, statistics for differences between groups in either testing condition were not significant; the wide standard deviations and the moderate number of subjects (17) may be partly responsible for this lack of significance.

3.3. Comparisons between the reading aloud and imitation tasks

Differences between tasks were also explored (Table 4). Significantly higher production scores for the imitation task were found for both A and B analysed together and separately in the pre-test (A&B: z = -4.57, p<.0001; A: z = -3.09, p<.005; B: z = -3.37, p<.005) and the post-test (A&B: z = -4.24, p<.0001; A: z = -2.88, p<.005; B: z = -3.19, p<.005). Interestingly, groups exhibited similar gain indexes in the two tasks, indicating that each group behaved similarly across tasks: experimental group A experienced a gain of around 10.7% in the post-test while experimental group B experienced a lower gain percentage (around 5.5%).

Table 4: Mean %, gain score and significance (*p-value*: ***significant at p<.0001, **significant at p<.005, *significant at p<.05) for production of weak forms in READ and IMIT tasks.

9 schwa	*Mean %*						*p-value*		
	READ			*IMIT*			*(READ vs. IMIT)*		
	pre (s.d.)	*post (s.d.)*	*gain (s.d.)*	*pre (s.d.)*	*post (s.d.)*	*gain (s.d.)*	*pre*	*post*	*gain*
A	12.42 *(11.7)*	22.88 *(15.9)*	10.46 *(14.4)*	29.41 *(18.4)*	40.52 *(18)*	11.11*(20.4)*	**	**	
B	11.76 *(12.7)*	16.99 *(20.8)*	5.22 *(15.3)*	30.06 *(14.0)*	35.95 *(20.7)*	5.88 *(16.7)*	**	**	

4. DISCUSSION

Pre-test results revealed that subjects exhibited some reduced productions already before training in both the reading and the imitation tasks (Tables 2 and 3). These weak productions in the pre-test condition manifest some awareness of the phenomenon on the part of the learners, which could be due simply to exposure prior to training.

Post-test results revealed a significant improvement in the production of weak forms after training in the reading task when exploring results in both experimental groups as a whole (Table 2) and when analysing both groups separately (gain score for A: 10.46% and for B: 5.22%). Similar positive training effects were found for the imitation task when the experimental groups were taken together (Table 3) and when they were analysed individually (gain score for A:11.11% and for B: 5.88%), (see Tables 2 and 3). These data prove the existence of a positive effect of the trainings administrated. However, it shall be noted that performance scores are somewhat low in both testing conditions. While low scores might have been expected in the pre-test, the rather moderate improvement shown in post-tests merits some comments. More training time may be needed to increase the training effectiveness. The time devoted to weak forms within the vowel reduction training programme was approximately 90 minutes divided in 3 sessions during the course of 3 weeks. Therefore it might not have been enough time for learning weak form rules, as results reveal that students were not very consistent in applying them in their productions. Another possible explanation could be the relative functional load which the students may have assigned to English grammatical word weakening. During their training regime, phonological vowel reduction had been presented first and shown in contrastive contexts (e.g. “pillars” vs. “pillows”), which emphasized that the incorrect use of a reduced vowel provokes a change in meaning. The fact that students were made aware that the use of a strong form does not often change the lexical but the semantic and pragmatic meaning (emphasis etc…) might have led students to give the rule low priority.

Data also revealed few differences between the two training regimes explored in the study. Both groups improved their production performance after the training. Group A, which received perception training combined with an inductive learning approach, experimented significant gains in their production of English weak forms. Group B, which had a production based approach (combined with a deductive learning approach) also experimented gains but these were not significant. Despite these differences, which seem to favour the use of a training regime based on perception and inductive learning techniques, no significant differences between groups were found in the post-test. This agrees with previous results indicating that training in one regime (either perception or production) benefits the other one, pointing to a mutually facilitating relationship between perception and production (Leather, 1990) .

Significant differences were observed between the reading and the imitation tasks in pre-tests and post-tests always favouring imitative production. Imitation techniques in speech production testing are used so as to verify the ability to model a specific sound (Logan & Pruitt, 1995). In the present study imitation had positive effects, since subjects were able to successfully produce some vowel reduction in weak forms after a perceptual model. The fact that scores were higher in post-tests also supports the idea that training helped develop this perceptual modelling ability. However, data reveal that this was not done in more than 50% of the cases (see Table 4) suggesting that, although the ability is present, it was not automatically activated and/or it was not sufficient to overcome other influences. These data agree with other studies which have concluded that linguistic interference is not fully avoided in imitation tasks (Nielsen, 2007; Alivuotila et al., 2007). The low scores in the reading task also suggest the presence of L1 orthographic influences and reading strategies (Koda, 1988) as in the learners’ L1, Spanish, there is a nearly transparent grapheme-phoneme correspondence. These learners may experience difficulties in developing reading strategies in languages with a less transparent correspondence, as is the case of English. This may be particularly true in the case of the English unstressed mid-central vowel, which may be represented with multiple vowel graphemes and grapheme combinations and which demands the activation of rules in the case of grammatical word production.

5. CONCLUSIONS

The present study showed moderate positive effects of phonetic training in second language acquisition of vowel weakening in weak forms. It explored potential differences between two different training regimes: perception (combined with an inductive learning approach) and production training (combined with a deductive learning approach). Few differences between the effect of either training on production were found, thus, supporting former studies which have revealed a facilitating relationship between the perception and production. Our study has also provided further evidence of the relevance of L1 interference in second language speech learning, which seems to be activated not only in reading tasks but also in imitation tasks.

6. REFERENCES

Aliaga-García, C., Mora, J. C. 2009. Assessing the effects of phonetic training on L2 sound perception and production. In M. A. Watkins, A. S. Rauber, B. O. Baptista (eds.), *Recent Research in Second language Phonetics/Phonology*. Cambridge Scholars Publishing. 2-31.

Alivuotila, L., Hakokari, J., Savela, J., Happonen, R. P., Aaltonen, O. 2007. Perception and imitation of Finnish open vowels among children, naïve adults, and trained phoneticians. *Sixteenth International Congress of Phonetic Sciences.*). Saarbrücken, Germany. 361-364.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., Tohkura, Y. 1999. Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception and Psychophysics, 61*(5), 977-985.

Catford, J. C., Pisoni, D. 1970. Auditory versus articulatory training in exotic sounds. *The Modern Language Journal, 54,* 477-481.

Cenoz, J., García Lecumberri, M. L. 1999. The Effect of Training on the Discrimination of English Vowels. *International Review of Applied Linguistics 37*(4), 261-275.

Flege, J. E. 1988. Using Visual Information to Train Foreign-Language Vowel Production. *Language Learning, 38* (3), 364-407.

Flege, J. E. 1995. Second-language speech learning: theory, findings and problems. In W. Strange (ed.), *Speech Perception and Linguistic Experience. Issues in Cross-Language Research.* Timonium, MD: York Press. 233-277.

Flege, J. E., Bohn, O.-S. 1989. An Instrumental study of vowel reduction and stress placement in Spanish accented English. *Studies in Second Language Acquisition, 11*, 35-62.

Gómez Lacabex, E. 2009. Perception and Production of Vowel Reduction in Second Language Acquisition. PhD Dissertation. Serie Tesis Doctorales. Leioa: Universidad del País Vasco.

Harris, J. 2005. Vowel reduction as information loss. In P. Carr, J. Durand, C. J. Ewen (eds.), *Headhood, Elements, Specification and Contrastivity: Phonological Papers in Honour of John Anderson*. Amsterdam: John Benjamins. 119-132.

Hazan, V., Sennema, A. 2007. The Effect of Visual Training on the Perception of Non-Native Phonetic Contrasts. *Proceedings of the Sixteenth International Congress of Phonetic Sciences*). Saarbrücken, Germany. 1585-1588.

Hazan, V., Sennema, A., Faulkner, A., Ortega-Llebaria, M., Iba, M., Chung, H. 2006. The use of visual cues in the perception of non-native consonant contrasts. *Journal of the Acoustical Society of America 119* (3), 1740-1751.

Iverson, P., Bronwen, G. E. 2007. Auditory Training of English vowels for first-language speakers of Spanish and German. *Proceedings of the Sixteenth International Congress of Phonetic Sciences.* Saarbrücken, Germany. 1625-1628.

Jaworski, S. 2008. *Lenition processes in English and other languages: A hierarchy of susceptibility to inertia.* Unpublished Philosophical Dissertation. Adam Mickiewicz University.

Koda, K. 1988. Effects of L2 Orthographic Representation on L2 Phonological Coding Strategies. *Journal of Psycholinguistic Research, 18* (2), 201-222.

Kondo, Y. 1995. Production of schwa by Japanese speakers of English: a crosslinguistic study of coarticulatory strategies. *Proceedings of the Sixteenth International Congress of Phonetic Sciences, 1* Stockholm. 302-305.

Leather, J. H. 1990. Perceptual and productive learning of Chinese lexical tone by Dutch and English speakers. *New Sounds 90: Proceedings of the 1990 Amsterdam Symposium on the Acquisition of Second Language Speech.* Amsterdam: University of Amsterdam Press. 72-89.

Lively, S. E., Logan, J. S., Pisoni, D. B. 1993. Training Japanese listeners to identify English /r/ and /l/. The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America, 94,* 1242-1255.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., Yamada, T. 1994. Training Japanese listeners to identify English /r/ and /l/. III Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America, 96* (4), 2076-2087.

Logan, J. S., Pruitt, J. S. 1995. Methodological Issues in Training Listeners to Perceive Non-Native Phonemes. In W. Strange (ed.), *Speech Perception and Linguistic Experience. Issues in Cross-Language Research*. Timonium, MD: York Press. 351-377.

Logan, J. S., Lively, S. E., Pisoni, D. B. 1991. Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America*, *89*, 874-886.

Neufeld, G. 1988. Phonological asymmetry in second-language learning and performance. *Language Learning*, *38* (4), 531-559.

Nielsen, K. Y. 2007. The Interaction between Spontaneous Imitation and Linguistic Knowledge. Working Papers in Phonetics. Department of Linguistics, UCLA. 125-137.

Rochet, B. L. 1995. Perception and production of second language speech by adults. In W. Strange (ed.), *Speech Perception and Linguistic Experience. Issues in Cross-Language Research*. Timonium, MD: York Press. 379-410.

Wang, X. 2000. Training Mandarin and Cantonese Speakers to identify English vowel contrasts: Long-term retention and effect on production. *Journal of the Acoustical Society of America*, *108* (5), 2653.

Yun, G., Jackson, S. R. 2006. Acquisition of English Reduced Vowels by Korean Speakers. *Proceedings of the XI Harvard International (Biennial) Symposium in Korean Linguistics.*

Acquiring angma – the velar nasal in advanced learners' English

Wiktor Gonet, Jolanta Szpyra-Kozłowska, Radosław Święciński

Institute of English, Maria Curie-Skłodowska University
gonet@umcs.pl, jolanta.szpyra-kozlowska@poczta.umcs.lublin.pl, radek@hektor.umcs.lublin.pl

ABSTRACT

The English velar nasal, known as angma, belongs to the most difficult consonants to master by Poles in spite of the fact that it is present in Polish as a realization of the dental nasal before velar plosives (e.g. in *tango* 'tango' pronounced as [ˈtaŋgɔ]). It is problematic for Polish learners, however, in other contexts, i.e. word-finally, as in *bring*, before vowels, as in *singer*, and before non-velars, as in *strongly*. In such cases speakers of Polish English commonly pronounce angma with the following velar plosive. It is the suppression of the latter segment that causes considerable learning difficulties.

This paper examines the acquisition of the velar nasal by 60 advanced/proficient Polish learners of English, students at the English Department of MCSU, Lublin, Poland. The realization of angma is scrutinized with a view to uncovering regularities in its acquisition by the participants, establishing their success rate and the degree of difficulty in the production of the velar nasal in each of the three problematic contexts. The obtained results are compared with those pertaining to students' acquisition of ash, schwa and unstressed unreduced vowels in order to characterize advanced learners' interlanguage in more detail. The presented observations carry important pedagogical implications for the phonetic training of Poles.

Keywords: acquisition of the velar nasal, advanced learners' interlanguage, Polish English.]

1. INTRODUCTION

While much research has been concerned with the acquisition of English pronunciation by beginning and intermediate foreign learners, relatively little is known about advanced and proficient learners' interlanguage (see, however, Nowacka 2008). In a series of studies (Gonet, Szpyra-Kozłowska & Święciński 2009a, b and c) we have undertaken an examination of phonetic progress and ultimate achievement of English Department students, representing a high level of language proficiency, in selected aspects of English phonetics particularly problematic for Poles. Thus, we have scrutinized their acquisition of unstressed unreduced vowels, ash and schwa, paying particular attention to the degree of difficulty involved in the realization of these segments in different phonological (segmental and prosodic) contexts.[1] This study is a continuation of our earlier research, this time with the focus on the velar nasal.

As noted by various researchers (e.g. Sobkowiak 1996:94), "*the velar nasal is among the hardest to master by Poles"* in spite of the fact that such a consonant is found in Polish, both in native words, e.g. *ręka* 'hand' [ˈrɛŋka], as well as in borrowings, e.g. *tango* 'tango' [ˈtaŋgɔ]. The problem with this segment lies in the fact that, as shown in the above examples, in Polish it occurs exclusively in the context of the following velar plosives while in English it has a wider distribution in that it is also found word-finally, as in *bring*, before vowels, as in *singer*, and before non-velars, as in *strongly.* It is exactly in these three contexts that the production of the consonant in question is problematic for Polish learners of English. A particularly difficult case concerns the items with two occurrences of [ŋ] in one word, as in *singing*.

As a matter of fact, the issue concerns not so much the articulation of angma itself as the suppression of the following plosive added to it due to two combined factors: distributional restrictions on the occurrence of this segment in Polish reinforced by a powerful impact of English spelling. Thus, typical of Polish English are forms such as [lɔŋk], [ˈsʲiŋgə], [ˈstrɔŋgli], [ˈrʲiŋɟiŋk][2] with the resulting impression of a strong foreign accent (Gonet 1982, Avery and Erlich 1994).

In this paper we examine the acquisition of the velar nasal in the problematic contexts by advanced to proficient Polish learners, i.e. by 60 English Department students. More specifically, the following issues are addressed:

- the participants' progress and success rate in the acquisition of angma;
- the degree of difficulty in mastering the velar nasal in each of the problematic contexts;
- a comparison of students' attainment in the production of angma, ash, schwa and unstressed unreduced vowels and its dependence on phonological context;
- pedagogical implications for the phonetic training of advanced Polish learners of English.

2. EXPERIMENTAL DESIGN

The experiment reported here was carried out in October and November of 2009. Below we present its design and the applied procedure.

60 students, both male and female, of the English Department at Maria Curie-Skłodowska University, Lublin, Poland took part in the experiment. They formed three groups which consisted of twenty 1st, 3rd and 5th year students, representing, roughly, upper intermediate, advanced and proficient learners respectively. Within each year the participants were selected randomly. The sample tested satisfactorily for homogeneity.

For the purposes of the experiment 60 English lexical items were selected. In 15 of them angma occurred in the word final position (e.g. *among*), in 15 before non-velars (e.g. *kingdom*) and in other 15 before vowels (e.g. *singer*). The last set comprised 15 words with two angmas: one medial before a vowel and one final (e.g. *singing*). The list of all the test items is placed in Appendix 1. The diagnostic words were subsequently used in 15 sentences (see Appendix 2).

The participants, unaware of the experimental goals, were given the list of diagnostic sentences to study for several minutes and then were asked to read them aloud at their own pace. Their performance was individually recorded. Thus obtained data were analysed auditorily, with the aid of acoustic analysis and a slow-rate replay. The applied procedure yielded 3600 tokens.

3. RESULTS AND DISCUSSION

In this section the experimental results are presented and discussed. The focus is first on the students' phonetic progress and final attainment in the production of angma and then on examining the realizations of this consonant in different phonological contexts. We will also compare the data concerning angma with those obtained in our earlier experiments on the acquisition of selected English vowels.

3.1. Progress and ultimate achievement in the production of angma

The percentage of correctly and incorrectly produced tokens by all the participants is shown in Fig. 1 while Figure 2 presents the correct realizations of angma by the three experimental groups.

Figure 1: Percentage of correctly and incorrectly realized tokens (in %, left).

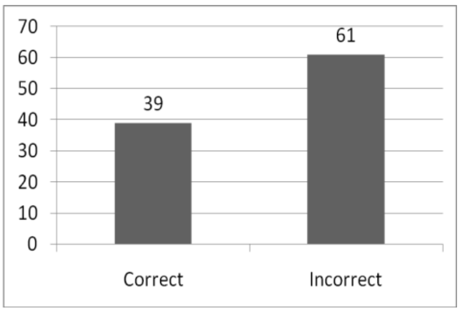
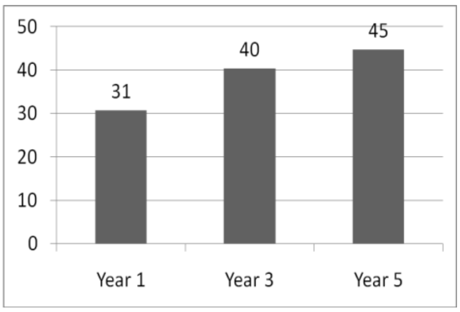


Figure 2: Correct realizations of angma by the three experimental groups (in %, right)



As shown in Figure 1, only 39% of all the tokens containing angma were pronounced properly by the participants while as much as 61% remained problematic for them.[3] Figure 2 provides the results obtained by the 1st, 3rd and 5th year students. It demonstrates that only 31% of the occurrences of angma were realized correctly by the first group, 40% by the second group and 45% by the most advanced students. This means a 9 per cent progress between Years 1 and 3, and only a further 5 per cent advancement between Years 3 and 5, with a very disappointing overall progress of 14%. The final outcome of 45% correct realizations of the velar nasal by the 5th year students can also be described as highly unsatisfactory.

3.2. Angma in different phonological contexts

The second goal of this study was to examine whether there is a meaningful connection between the degree of difficulty in producing angma and three problematic contexts of its occurrence: word-finally, before non-velars and before vowels. Figure 3 provides a summary of the relevant data.

Figure 3: Correct realizations of angma in 3 contexts (in %)

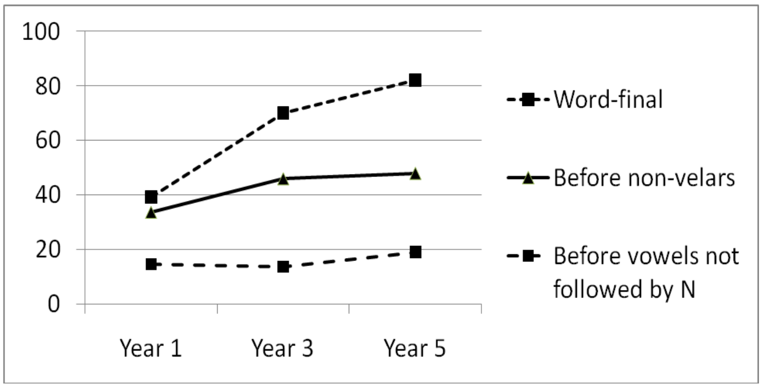


Figure 3 points to a significant contextual dependency in the realizations of the velar nasal by the three experimental groups; for all the participants angma is the easiest to pronounce in word-final position (the mean of 63,6% of correct responses), it is of medium difficulty before non-velars (mean = 42,6%) and the most problematic before vowels (mean = 16%). This dependence is smaller for Year 1 than for Year 3 and Year 5.

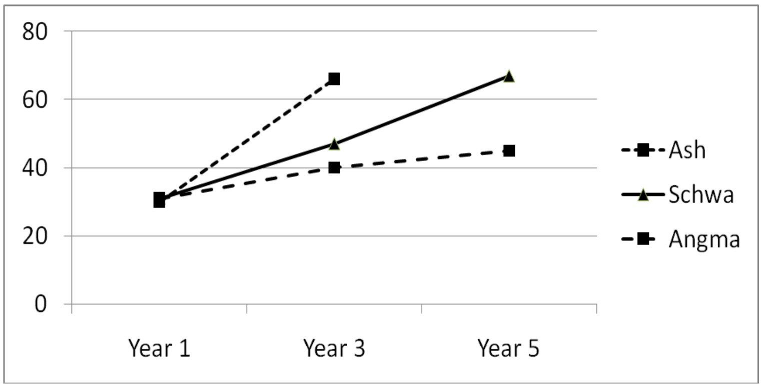
Moreover, there are important differences in the rate of progress in the production of the examined segment depending on the context of its occurrence. Thus, the students' performance with respect to word-final angma improved by 43% from Year 1 to Year 5, by 14% in the context before non-velars and only by 4% in the case of the prevocalic velar nasal (with a one per cent regression from Year 1 to Year 3 in the latter context). It seems that a much higher success rate in the production of word-final nasals in comparison with its remaining instances stems from a very high frequency of *–ing* forms as well as the focus of phonetic instruction on this particular case.

Interestingly, angma is easier to pronounce before vowels when another such consonant is found at the end of the same word, as in *singing*. In such instances the percentage of correct responses rises from 15% to 35% in Year 1, from 14% to 32% in Year 3 and from 18% to 39% in Year 5 (mean improvement = 20%). The same holds true of the second, word-final nasals whose realization improves from 39% to 58% Year 1, from 70% to 80% in Year 3 and from 78% to 83% in Year 5 (mean progress = 11%). It might be the case that such accumulation of angmas in a single word forces learners to concentrate hard on the proper articulation of these segments and results in increased correctness.

4. ADVANCED STUDENTS' ACQUISITION OF ANGMA, ASH, SCHWA AND UNREDUCED VOWELS – A COMPARISON.

In this section we intend to put the data concerning angma in a broader perspective by comparing them with the results obtained in the previous experiments on the students' acquisition of unstressed unreduced vowels, ash and schwa (see Gonet, Szpyra-Kozłowska & Święciński 2009a, b and c). The relevant figures are juxtaposed in Figure 4,[4] where percentages refer to the correct realizations of the segments in the test items.

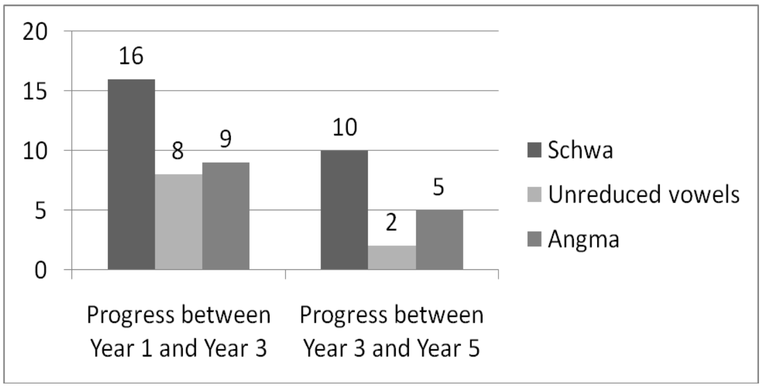
Figure 4: Results of 3 experimental studies on ash, schwa and angma.



An examination of the data in Figure 4 reveals a similar starting point in all three instances, with about 30% of correct realizations of all the segments by Year 1 students. In the case of ash and schwa, however, a marked overall progress of 33% and 36% can be observed between Year 1 and Year 5 (with a similar final achievement of about 70% in Year 5), whereas in the case of angma the improvement amounts to 14% only. This means that the proper production of the velar nasal in the three examined contexts is more problematic for advanced Polish learners than of the two vowels, contrary to a common view concerning the relative ease of mastering English consonants in comparison with vowels.

Another general interesting observation is that greater phonetic progress is made by the students within the first years of their university education than in its second part, as shown in Figure 5.[5]

Figure 5: Comparison of progress rate between Years 1, 3 and 5.



This regularity can probably be attributed to an intensive exposure of students to and their deep immersion in English in the first period of their university studies which is bound to have an immediate powerful impact on the quality of their phonetic performance. Another significant factor is the formal pronunciation training which students undergo within the first two years.

Table 1: Dependence of segment realization on phonological context

Phonetic issue	Dependence of pronunciation difficulty on phonological context		Strength of dependence
ASH	**Easier:** before palatoalveolar fricatives (e.g. *c**a**sh*) in monosyllables (e.g. *m**a**t*), in post-stress position (*rucks**a**ck*)	**More difficult:** before liquids (e.g. *sh**a**ll*) in polysyllables (e.g. *C**a**nada*) in pre-stress position (e.g. ***a**dverse*)	weak
SCHWA	word-medially (e.g. *m**a**jority*)	word-initially (e.g. ***a**ttend*) word-finally (e.g. *com**a***)	weak
Unstressed unreduced vowels	after stress (e.g. *ins**e**ct, handf**u**l*)	before stress (e.g. *c**a**rtoon*)	weak
ANGMA	word-finally (e.g. *so**ng***)	before vowels (e.g. *si**ng**er*)	strong

Let us now examine the relationship between the degree of difficulty involved in the production of the studied segments and the phonological context in which they occur.

Table 1 demonstrates that in all the examined aspects of Polish English phonetics there is a dependence between the ease/difficulty of segment articulation and the phonological context in which it occurs. In none of the instances that have been scrutinized is this dependence stronger than in the case of angma.

Another significant observation that has been made in our previous studies and which finds further support in the present work is that the dependence between phonological context and students' realization of problematic sounds is strongly related to their level of proficiency. Thus, for the least advanced freshmen all contexts appear to represent a similar degree of difficulty, but starting with Year 3 and continuing with older students, the sound's segmental and / or prosodic environment begins to play a more decisive role.

5. CONCLUSIONS

This study, while focusing on the intricacies involved in the production of the velar nasal, has also attempted to shed more light on the development of various aspects of pronunciation skills in Polish university students of English and to characterize some important properties of these learners' interlanguage.

In all the examined instances a steady phonetic progress in the realization of problematic sounds could be observed within five years of students' university education, most of which, however, took place within its initial period. The poorest progress occurred the in the case of angma (14%). This fact coupled with the final success rate of 39% only indicates that angma found in three contexts under examination is more difficult to master than the remaining aspects of English phonetics discussed in this paper.

Our experiment has also uncovered a strong dependence between the phonological context in which the velar nasal occurs and the degree of difficulty of its realization, with the following order of increasing difficulty: word-finally, before non-velars and before vowels. The presence of two angmas in a single word resulted in a considerably improved production of these consonants, which can be attributed, somewhat paradoxically, to an especially high degree of difficulty of such cases that require an extra articulatory effort on the learner's part.

An interesting observation can be made concerning context-dependence in the realization of English sounds and learners' proficiency: the more advanced they are, the greater the influence of the phonological environment in which a segment is found.

Pedagogical implications of this research are obvious: proper phonetic training of advanced and proficient learners should continue in some form throughout the whole period of their studies and take into account difficult phonological contexts of sounds' occurrence.

6. REFERENCES

Avery, P., Ehrlich, S. 1994. 1992. *Teaching American English Pronunciation.* Oxford: Oxford University Press.

Gonet, W. 1982. 'Legato speech', or the psychology of teaching phonetic fluency. *Lubelskie Materiały Neofilologiczne*, pp. 307-319.

Gonet, W., Szpyra-Kozłowska, J., Święciński., R. 2009a. Clashes with ashes. In Waniek-Klimczak (ed.)

Gonet, W., Szpyra-Kozłowska, J., Święciński, R. 2009b. The acquisition of Vowel Reduction by Polish students of English. In Waniek-Klimczak (ed.)

Gonet, W., Szpyra-Kozłowska, J., Święciński, R. 2009c. Unstressed unreduced English vowels in the speech of Polish students – a developmental study. In Waniek-Klimczak (ed.)

Nowacka, M. 2008. *Phonetic Attainment in Polish University and College Students of English. A Study in the Productive and Receptive Pronunciation Skills.* Unpublished Ph.D. dissertation. Maria Curie-Skłodowska Uniersity, Lublin, Poland.

Sobkowiak, W. 1996. *English Phonetics for Poles.* Poznań: Bene Nati.

Waniek-Klimczak, E. (ed.). 2009. *Issues in Accents II: Variability and Norm.* Cambridge Scholars Publishing.

APPENDIX 1

List of lexical items used in the experiment.

Word-final	Before non-velars	Before vowels	Sequence of 2 angmas (before a vowel & final)
1. among	*kingdom*	*Springer*	*singing*
2. nothing	*hangman*	*hangover*	*banging*
3. along	*belonged*	*youngish*	*hanging*
4. drinking	*songs*	*longish*	*springing*
5. looking	*Springfield*	*singer*	*ringing*
6. wrong	*pangs*	*hangout*	*flinging*
7. rang	*Wellington*	*Birmingham*	*upbringing*
8. Reading	*strongly*	*Buckingham*	*prolonging*
9. ring	*things*	*Bingen*	*clinging*
10. gang	*Langdale*	*coat hangers*	*belonging*
11. playing	*Arlington*	*ringers*	*swinging*
12. ping	*strings*	*Nottingham*	*slinging*
13. pong	*earings*	*slangy*	*stinging*
14. horse-riding	*surprisingly*	*stringy*	*stringing along*
15. young	*wrongly*	*banger*	*wringing*

APPENDIX 2

List of sentences used in the experiment

1. In the old days in most *kingdoms* the *hangman's* job *belonged* to highly prestigious professions.
2. She kept *wringing* her hands in despair when she thought of the way that *gang* was *stringing* her husband *along*.
3. When he lived *among* the bushmen there was *nothing* he missed more than "Jerry *Springer* Show."
4. After the night of *drinking, singing songs* and *banging* on the drums in a *Nottingham* bar they had a terrible *hangover* and *pangs* of conscience.
5. This *youngish-looking* man with *longish* hair was a *singer* in a terrible *hangout* in *Wellington* suburbs.
6.With his aristocratic *upbringing* he felt quite at ease in *Buckingham* Palace, where he received a medal and a royal *ring*.
7. I *strongly* agree that it was *wrong* to allow the children to do all that *hanging* from the ropes, *springing* from the furniture, *ringing* the bells and *flinging things* at each other.
8. They *rang* us up to tell us about their weekend trip to *Reading, Langdale, Birmingham* and *Arlington*.
9. *Surprisingly* we didn't know they enjoyed *horse-riding* and *playing ping pong*.
10. He thought of *prolonging* his stay in *Bingen* in order to buy new *strings* for his guitar, nice *earrings* for his girl-friend and some wooden *coat hangers*.
11. Because of the fire in our neighbours' house our eyes were *stinging* from the smoke.
12. The *ringers* of the bells kept *clinging* to the ladder of the church tower.
13. This *young* man from *Springfield* had *stringy* arms and *slangy* speech.
14. He drove his old *banger swinging* it from the left to the right and *slinging* mud all over the place.
15. They were *wrongly* suspected of *belonging* to a terrorist organization.

NOTES

[1] In this respect our studies differ from Nowacka's (2008) work in which Polish students' phonetic progress is examined without, however, taking into account contextual factors.

[2] The presence of [k] in some items is due to Polish word-final obstruent devoicing as well as to regressive voice assimilation in obstruent clusters.

[3] Incorrect realizations included the following: [ŋg, ŋk, ndž, n, ng, nk].

[4] The study concerning the vowel ash examined the performance of 1st and third year students only.

[5] This observation coincides with Nowacka's (2008) results.

Towards a quantitative model of Mandarin Chinese perception of English consonants

Jian Gong[1], *Martin Cooke*[1,2], *M. Luisa García Lecumberri*[1]

[1] Language and Speech Lab (laslab.org), University of the Basque Country, Spain
[2] Ikerbasque (Basque Foundation for Science)

ABSTRACT

Models which describe the perception of foreign language sounds typically do so using qualitative relations with sounds in the first language, but a more fine-grained account of learners' perceptual difficulties might be obtained via techniques from automatic speech recognition. Here, we employ generative statistical models of speech – Hidden Markov Models – which underly most work in speech recognition, to learn the sound systems of English and Mandarin Chinese, and use these as the basis for a quantitative model of the perception of English intervocalic consonants by Chinese listeners. This approach allows both the prediction of consonant identification preferences and the construction of a complete cross-language consonant distance matrix, which can then be compared to consonant categorisation and goodness ratings respectively. To evaluate the model, 30 native Chinese listeners with moderate English competence and residing in China categorised and rated English intervocalic consonants. A high degree of consistency was found between listeners and the computer model for sound categorisation as well as a clear and significant correlation between goodness ratings and distance measurements, suggesting that the technique has potential in predicting sound-level difficulties experienced by language learners. One key difference between listeners and the model is in the relative influence of acoustic and orthographic factors in sound categorisation: while model decisions and distances are based solely on acoustic information, clear evidence of orthographic interference can be seen in listener responses. We explored this issue by comparing the use of Chinese characters versus Pinyin graphemes on the task. Listeners who were presented with Pinyin showed some orthographic influences, while a dialect influence occurred in the character group.

Keywords: consonant categorisation, goodness, computer model.

1. INTRODUCTION

Learners' L2 sound perception is strongly influenced by their L1 sound system. Theoretical models of second language acquisition have been proposed to give predictions of the degree of success in L2 sound acquisition. The Perceptual Assimilation Model (PAM) (Best 1995) suggests that the perception of L2 sounds is based on their similarities to or distances from the closest counterpart sounds in the L1. Accordingly, the perceived distance between L2 and L1 sounds may determine the degree of discriminability of certain L2 sound pairs. The Speech Learning Model (SLM) (Flege 1995) hypothesises that only when the phonetic distance between the L2 sound and L1 sound is large enough for the learner to detect, a new phonetic category may be established. The larger the perceived distance between an L2 and an L1 sound, the more likely a new category for the L2 sound will be established during the learning process. Accordingly, both models agree that the perceived phonetic distance between L1 and L2 plays a crucial role in L2 perception. Many studies have been carried out to test the predictions made by the two models via cross-language mapping experiments and categorical discrimination tests (Strange et al. 1998, 2001; Guion et al. 2000; Lengeris and Hazan 2007). These studies show that although the two models have a certain degree of explanatory power, there are still some limitations. Since the PAM and SLM describe how L1-L2 sound similarity or distance affect non-native perception in a mainly qualitative manner, a better understanding of the relationship between the L1 and L2 sound systems might be obtained using a more quantitative approach. Our goal is to develop a computational account of L2 sound perception capable of predicting non-native confusions. Such a model would allow a controlled examination of the influence of factors such as the amount and quality of sound exposure as well as an exploration of how two or more co-existing sound systems interact in classifying speech sounds. Here we present an initial model of Chinese perception of English intervocalic consonants, and compare its responses with behavioural data from Chinese listeners.

2. COMPUTER MODEL

2.1. Corpus

The computer model was built using isolated vowel-consonant-vowel (VCV) tokens derived from material collected for the Interspeech 2008 Consonant Challenge (Cooke and Scharenborg 2008), which includes all 24 English consonants in vowel contexts derived from all nine combinations of /æ/ /iː/ /uː/ as first and second vowel, produced by 12 female and 16 male speakers. A similar Mandarin Chinese VCV corpus including all 24 Chinese consonants (/pʰ p tʰ t kʰ k tsʰ ts ʈʂʰ ʈʂ tɕʰ tɕ f s ʂ ɕ x m n ŋ l ɻ j w/) in the same set of vowel contexts was collected from 12 female and 17 male speakers, all native Chinese students studying at the University of Sheffield. Post-processing involved high-pass filtering to attenuate low frequency energy below 50 Hz from the tokens, followed by endpointing to remove silence. Tokens were downsampled to 25 kHz and normalised to have the same root mean square energy level. The corpus was screened manually to identify and remove tokens which were incorrectly produced or contained noise from, for example, key tapping. Following screening, a total of 3299 English and 3331 Chinese tokens were available for model training and testing.

2.2. Hidden Markov models

Separate models were built for English and Chinese sounds using speech material from the two corpora. Using standard Hidden Markov modelling (HMM) techniques and acoustic representations (Mel-Frequency Cepstral Coefficients) employed in automatic speech recognition, baseline recognition models were trained for each consonant and vowel in each language, resulting in two recognition systems, one for each language. Hidden Markov models essentially compute an optimal assignment or clustering of acoustic data to model state sequences. Here, each sound (consonant or vowel) was represented by three states, and each individual state learns a statistical distribution of acoustic information allocated to that state. The allocation of training data to states is determined by within-state similarity rather than uniformly sampling the interval during which the consonant or vowel is present. Thus, for example, while the model for a plosive is likely to contain a sequence of states representing formant transition, closure and burst, these segments are not manually-identified but emerge automatically from similarity criteria during the training phase. Trained HMMs can be used to categorise a new instance of a sound, whether or not that instance belongs to the language for which the models were trained. The categorisation result is obtained by choosing the HMM with the highest likelihood of generating the sound to be classified. The rightmost column of Table 1 provides classification results for English VCVs using HMMs trained on Chinese VCVs. We discuss these results in section 4.

2.3. Cross-language model distance

Cross-language sound similarity measures can be generated by taking all possible pairings of HMMs trained on English or Chinese sounds. Distances are not based on single exemplars of English and Chinese sounds but on the acoustic distributions learnt by the HMMs. Consequently, the distance measures we use involve the overlap between pairs of probability distributions rather than conventional pointwise distances such as the Euclidean or cityblock metrics. A number of distance measures employed in pattern recognition were used in current study to calculate all pairwise distances between the consonant HMMs of the two languages. These measures including Bhattacharrya distance, which is a measurement of the similarity of two probability distributions; Kullback-Leibler divergence, also called Relative Entropy in information theory, which is also a distance measurement between two probability distributions. In addition, Kullback-Leibler divergence with Monte-Carlo sampling was also used, which differs from the standard Kullback-Leibler divergence in using randomly-generated sample points of two probability distributions rather than probability density functions themselves (Mak and Barnard 1996; Sooful and Botha 2002; Cover and Thomas 2006; Hershey and Olsen 2007).

3. CROSS-LANGUAGE MAPPING EXPERIMENTS

Human perception data were collected to test and refine the computer model. Native Chinese listeners categorised English consonants in terms of their native sounds. To afford comparison with model distance measures, listeners also provided a goodness rating. Chinese is a well known language using the logographic writing system. Its characters are not directly linked to their pronunciations. However Chinese has a Romanisation system called Pinyin, which uses Roman letters to mark the sounds and to help the learner of Chinese to remember the pronunciation of the Chinese characters. Most Chinese children learn Pinyin in primary school at the same time as they learn Chinese characters. To examine whether the differing orthographic forms of Chinese characters or Pinyin symbols influence the categorisation or goodness ratings, one group of participants responded using Chinese characters while the other group used Pinyin (figure 1).

Figure 1: Screen shots of the experiment interface (left: character group; right: Pinyin group).

3.1. Stimuli

A subset of English and Chinese VCV tokens used to train the computer model were used as stimuli in the cross-language mapping experiments. 552 English tokens from 16 male speakers, including all English consonants except /ŋ/ (excluded due to phonotactic considerations) were used in the experiments. The initial vowel was one of /æ,iː,uː/ while the final vowel was /æ/. All tokens has end stress. 46 Chinese tokens selected from those produced by 12 male speakers were mixed into the English tokens. A further 23 Chinese tokens were used in a prior keyboard training session. The Chinese tokens used the same vowel contexts as the English. The average length of the final vowel /æ/ in Chinese VCVs (168 ms) is significantly shorter than the VCV-final English /æ/ (233 ms), as also found Guion et al. (2000). To prevent the use of final vowel length as a cue in the perception experiments, we also normalised the length of the final vowel of all the English and Chinese tokens to 150 ms, with the final 20 ms linearly ramped down to zero amplitude.

3.2. Participants

30 native Chinese speakers, 10 females and 20 males, were divided into two listener groups of 15 participants. One group used Chinese character symbols to record their perceptions, while the other group used Chinese Pinyin symbols. All were second year students (mean age = 20.6) at the Xi'an Technological University, China, studing non-linguistic courses, and none had lived outside China. Children start to learn English in China when they go to middle school and sometimes earlier. It is very difficult to find Chinese young adults who have never had English lessons. On average, our participants had begun to learn English at 12.6 years. Most of their exposure to English was in the classroom with Chinese teachers, and sometimes from other sources such as music and films. Most of the participants were from north-west China, which is the same dialect area as Mandarin Chinese. Pure-tone hearing tests at frequencies from 250 to 8000 Hz were carried out using a software audiometer. One participant from the character group was excluded due to problems in both ears at high frequencies. All participants received a small payment for taking part.

3.3. Procedure

The experiments were carried out in a quiet meeting room in Xi'an Technological University. Participants were tested individually and heard stimuli via Sennheiser HD650 headphones and an M-Audio Mobilepre external sound card. Stimulus presentation and response collection was controlled by a computer program.

Participants used the mouse to select their response category buttons on a virtual keyboard, with Chinese characters or Pinyin graphemes as the symbol of the VCV tokens on each button (Figure1). Participants were asked to first classify the token they heard as an instance of one of the Chinese consonant categories, then to move a slider bar under the keyboard to give a goodness rating of the sound (relative to the Chinese category they selected). The goodness rating used a 0-100 scale (0=bad, 100=good), although participants were only aware of the continuous sliding scale. Both a keyboard training session and a practice test were carried out before the formal test. Participants were asked to click each button to hear the corresponding Chinese token (1 for each button) to familiarise themselves with the Chinese consonants and their positions on the keyboard. The practice test contained two examples of each English consonant (46 tokens in all). In the formal test, 46 Chinese tokens (2 for each consonant) and 506 English tokens (22 for each consonant) were mixed and presented to the participants in 3 sessions (184 tokens randomly distributed for each session). Participants were asked to take a short break before they continued to the next session.

4. RESULTS

4.1. General description

The cross-language mapping results are listed separately for the Chinese character and Pinyin groups in Table 1. The cross-language computer model recognition results are also shown. Some English consonants, especially the plosives, were assimilated to certain Chinese categories with a very high frequency, and the goodness ratings for these English sounds were also very high, which suggests that the distances between these sound-pairs are relatively small. Other sounds, especially those English sounds which are known to lack Chinese counterparts, produced more confusions, and the goodness ratings were also lower. Inspired by Guion et al. (2000), a 'fit index' was calculated by multiplying the percentage and the goodness rating. The mean fit indices for the Chinese consonants used as an experiment control (section 3.1) were 69.8 (sd: 16.3) for the character group and 71.4 (sd: 16.7) for the Pinyin group. Based on the same standard deviation criterion as Guion et al. (2000), English consonants were classified into 'good', 'fair' or 'poor' instances of Chinese categories, as depicted in grey tones in the table. Four pairs of English consonants – /f v/, /θ s/, /ð z/ and /ʃ ʒ/ – where both sounds in each pair were assimilated to the same one or more Chinese categories, are expected to be difficult for Chinese listeners to discriminate.

4.2. Listener – Model comparison

Table 1 also shows that the model's cross-language recognition results are consistent with listeners' categorisation results. While listeners made fewer confusions, if we focus on the first few major confusions, the model made almost the same confusions as listeners. If the confusion ranking is considered, the biggest differences lie in the English consonants /ʧ ʤ ʒ/. Interestingly, in these cases the best ranked confusion for listeners was second ranked by the model and vice versa. Also, the percentage differences between the first two confusions in the model were smaller than for listeners. This suggests that listeners are able to exploit information which is poorly-represented or missing in the model's acoustic representation or topology.

As mentioned in section 2.3, the model permits an estimation of between-language consonant distance. Table 2 lists the correlations between model distances and listener goodness ratings. The two-way Kullback-Leibler divergence with Monte-Carlo sampling (KL_MC2) is the distance measurement with the most significant and highest correlation with the goodness rating [$r = -0.66$, $p < .001$]. The group presented with Chinese characters showed higher correlations with the model than the Pinyin group. Since model distances are purely acoustic, this finding points to orthographic influences from the Pinyin characters.

Table 1: Results of the cross-language mapping experiments and the cross-language computer model recognition. For each listener group, mean categorisation percentages and goodness ratings are provided as well as a fit index described in the text. Categorisation percentages from the model are also shown. Only the principal confusions (≥5%) are shown in the table. Rows highlighted in dark grey are 'good' exemplars, those in light grey are 'fair', while the rest are 'poor'.

English	Character			Pinyin			Model
	%	good	fit	%	good	fit	
p	pʰ(99%)	80	79.2	pʰ(98%)	82	80.36	pʰ(61%) p(20%) tʰ(9%) t(6%)
b	p(98%)	77	75.46	p (97%)	78	75.66	p (56%) t(23%) w(8%)
t	tʰ(86%) tsʰ(9%)	75 72	64.5 6.48	tʰ(82%) tsʰ(13%)	75 76	61.5 9.88	tʰ(23%) tʃ̺ʰ(20%) tɕʰ(17%) tsʰ(10%) tʃ̺(10%) ts(7%) t(6%)
d	t(95%)	75	71.25	t(98%)	76	74.48	t(68%) k(21%) tʃ̺(8%)
k	kʰ(95%)	78	74.1	kʰ(95%)	78	74.1	kʰ(52%) tʰ(16%) k(13%) tɕʰ(6%)
g	k(97%)	79	76.63	k(98%)	77	75.46	k(76%) t(9%) n(5%)
tʃ	tʃ̺ʰ(60%) tɕʰ(14%) tsʰ(13%) tʃ̺(7%)	71 70 69 71	42.6 9.8 8.97 4.79	tʃ̺ʰ(85%) tʃ̺(7%) tɕ(5%)	72 64 52	61.2 4.48 2.6	tɕʰ(55%) tʃ̺ʰ(23%) tʃ̺(15%)
dʒ	tʃ̺(63%) tɕ(25%)	75 70	47.25 17.5	tʃ̺(69%) tɕ(21%)	71 66	48.99 13.86	tɕ(38%) tʃ̺(32%) tɕʰ(11%) ɹ̺(10%)
f	f(97%)	77	74.69	f(99%)	80	79.2	f(61%) ʃ̺(14%) s(13%) x(9%)
v	w(74%) f(19%)	64 60	47.36 11.4	w(75%) f(20%)	68 46	51 9.2	ɹ̺(27%) f(19%) ʃ̺(10%) x(12%) s(8%) p(5%)
θ	s(50%) f(46%)	68 74	34 34.03	f(56%) s(41%)	70 63	39.2 24.6	f(35%) s(32%) ʃ̺(13%) x(11%)
ð	w(30%) ts(20%) ɹ̺(20%) f(13%) s(8%) tʃ̺(6%)	60 43 44 53 45 45	18 8.6 8.8 6.89 3.6 2.7	ts(45%) w(34%) f(11%)	58 56 39	26.1 19.04 4.29	ts(18%) ɹ̺(14%) ʃ̺(12%) f(11%) x(8%) t(8%) s(7%)
s	s(96%)	78	74.88	s(95%)	81	76.95	s(55%) ʃ̺(28%) ɕ(10%)
z	ɹ̺(42%) ts(33%) s(18%)	45 42 45	18.9 13.86 8.1	ts(74%) ɹ̺(13%) s(6%)	71 44 57	52.54 5.72 3.42	ɹ̺(34%) s(22%) ʃ̺(12%) ɕ(12%) ts(6%) j(5%)
ʃ	ʃ̺(83%) ɕ(14%)	80 66	66.4 9.24	ʃ̺(91%) ɕ(5%)	78 67	70.98 3.35	ʃ̺(69%) ɕ(22%) tɕʰ(7%)
ʒ	ɹ̺(68%) ʃ̺(14%) j(9%)	51 58 58	34.68 8.12 5.22	ɹ̺(78%) ʃ̺(10%)	56 63	43.68 6.3	ʃ̺(39%) ɹ̺(25%) ɕ(15%) tɕʰ(6%)
h	x(98%)	78	76.44	x(98%)	82	80.36	x(79%)
m	m(96%)	78	74.88	m(98%)	78	76.44	m(88%) l(5%)
n	n(88%) l(11%)	74 75	65.12 8.25	n(96%)	75	72	n(56%) m(27%) l(9%)
l	l(89%) n(10%)	79 71	70.31 7.1	l(97%)	82	79.54	l(69%) ɹ̺(12%) n(6%) w(6%)
r	ɹ̺(83%) w(11%)	49 46	40.67 5.06	ɹ̺(93%)	55	51.15	ɹ̺(61%) w(15%) l(8%) j(5%)
j	j(99%)	80	79.2	j(98%)	80	78.4	j(67%) ɹ̺(12%) w(9%)
w	w(96%)	78	74.88	w(96%)	77	73.92	w(80%) x(6%)

4.3. Character group – Pinyin group comparision

Table 1 suggests that most of the classifications by the two groups are similar. However, differences occurred for the English consonants /tʃ θ ð z ʃ ʒ n l r/**.** For instance, English /tʃ/ was heard as Chinese /tʃ̺ʰ/ on 85% of occasions by the Pinyin group compared to 60% for the character group, while figures for English /z/ heard as Chinese /ts/ are 74% and 37% respectively. Chinese characters form a logographic system, whose written form is not directly linked to the pronunciation, while Pinyin is a Romanised alphabetic system, which is used to represent the pronunciation of Chinese characters. For the English sounds /tʃ z ʃ r/**,** this "pronunciation-recall" aspect of Pinyin may have had an influence. In Pinyin, the Chinese consonants /tʃ̺ʰ ts ʃ̺ ɹ̺/ are written as "ch", "z", "sh" and "r", which have the same written forms as the English consonants /tʃ z ʃ r/. As all the participants were university students and knew some English (although none were fluent), and they knew the pronunciation of those Pinyin symbols in English. Although participants were not told the language of sounds they would hear, since English is the only foreign language they knew, it was very likely that they would connect the sounds they heard to English when making their choice and supplying a rating. For sounds such as /ð/ and /ʒ/, Pinyin may have had an orthographic influence on category decisions. Another interesting case concerns the /n/-/l/ confusions made by the Character group but not the Pinyin group. A closer look at the participants' individual data shows that most of the /n, l/ confusions came from 3 participants whose dialect belongs to the south west Mandarin region where this confusion is common.

Table 2: Correlations between goodness ratings and distance measurements for the Character and Pinyin groups. Both raw and normalised (z-score) goodness ratings are provided. Bha: Bhattacharrya distance; KL2: 2-way Kullback-Leibler divergence; KL_MC: KL divergence + Monte-Carlo sampling, KL_MC2: 2-way KL divergence + Monte-Carlo sampling.

Raw goodness rating		Distance measurements	Normalised goodness rating	
Character	Pinyin		Character	Pinyin
-0.29 n.s.	-0.20 n.s.	Bha	-0.31 *	-0.24 n.s.
-0.34 *	-0.12 n.s.	KL2	-0.36 *	-0.16 n.s.
-0.64 ***	-0.49 **	KL_MC	-0.64 ***	-0.48 **
-0.66 ***	-0.54 ***	KL_MC2	-0.67 ***	-0.53 **

5. DISCUSSION

The listener–model comparison shows clear similarities in sound categorisation and significant (negative) correlation between listeners' goodness ratings and model sound-pair distances, suggesting that acoustic clustering techniques used in automatic speech recognition may be valuable in cross-language studies. For example, it would be possible to predict which L2 sounds would be problematic for any given set of L1 listeners by training on speech material from the two languages. The modelling approach also has the potential to provide a more fine-grained classification of sound similarity than conventional qualitative models. However, listener-model differences in classification rates even for the categories which have a clear cross-language assimilation point to imperfections in the current computer model's ability to accurately represent those aspects of the acoustic signal which listeners have access to in decision making. The failings of the hidden Markov modelling framework are widely-acknowledged (e.g. HMMs are poor at duration modelling). Further, the acoustic representation used here is well-matched to HMMs but is almost certainly not that used by human listeners. Another limitation of the computer model is that it is trained on VCV tokens rather than natural speech. By acquiring speech using more natural material, listeners are exposed to more variety, for instance, in context, speech rate and accent. This limitation will be overcome in future work by training with more natural material. Possible orthographic influences revealed by Character-Pinyin differences highlights a methodological concern: as more young people in China start to learn English very early in their lives, use of Pinyin symbols in cross-language mapping experiments may become less reliable. However, Chinese characters are not free from problems, since they can have different pronunciations in different dialect regions, or even in the same dialect region by listeners with different backgrounds, as evidenced by the /n, l/ confusions which occurred for the Character group.

6. REFERENCES

Best, C. 1995. A Direct Realist View of Cross-Language Speech Perception. In: W, Strange. (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. York, Timonium, MD. 171–204.

Cooke, M.P. and Scharenborg, O. 2008. The Interspeech 2008 Consonant Challenge. *Proc. Interspeech*, *Brisbane, Australia*.

Cover, T.M. and Thomas, J.A. (2006). *Elements of Information Theory*, Second Edition. John Wiley and Sons.

Flege, J. 1995. Second-Language Speech Learning: Theory, Findings and Problems. In: W, Strange. (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. York, Timonium, MD. 233–273.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., Pruitt, J. C. 2000. An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *J. Acoust. Soc. Am.* 107(5). 2711–2723.

Hershey, J. R., Olsen, P. A. 2007. Approximating the Kullback-Leibler divergence between Gaussian mixture models. *Proceedings of ICASSP 2007, Honolulu, USA.* 317-320

Lengeris, A., Hazan, V. 2007. Cross-language perceptual assimilation and discrimination of southern British English vowels by Greek and Japanese learners of English. *Proceedings of the 16th ICPhS, Saarbrücken*. 1641–1644.

Mak, B. and Barnard, E. 1996. Phone clustering using the Bhattacharyya distance. *Proc. ICSLP'96, Philadelphia, USA*. 2005–2008.

Sooful, J. and Botha, E. 2002. Comparison of acoustic distance measure for automatic cross-language phoneme mapping. *Proc. ICSLP'2002, Denver, USA*. 521–524.

Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S., Nishi, K., and Jenkins, J. 1998. Perceptual Assimilation of American English Vowels by Japanese Listeners. *J. Phonetics* 26. 311-344.

Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S., and Nishi, K. 2001. Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners. *J. Acoust. Soc. Am.* 109(4). 1691–1704.

Phonetic attrition in Spanish with English as L3 in a natural setting

Argiñe González
MªLuisa García Lecumberri
Mª Yolanda Ruiz de Zarobe
Language and Speech Lab
(laslab.org)

University of the Basque Country, Spain

ABSTRACT

Studies investigating cross-language influences usually focus on the effect of the native language (NL) sound system on the second (L2) or third language (L3), but NL sound systems can also be affected by contact with other languages, a phenomenon known as attrition (Seliger and Vago, 1991). In the present study we focus on the influence of English (L3) on one of the NLs' sound systems (Spanish) in speakers with two NLs (Spanish and Basque). We analyze the degree of foreign accent in their L3 and in one of their NLs (Spanish), which is a sign of attrition. Several factors which might account for the various degrees of foreign accent are analyzed (age of arrival, length of residence, NL activation, identification with culture, degree of foreign accent in the other language). First generation Spanish/Basque immigrants were recruited in USA and interviewed in English and Spanish. Contrary to expectations, the speakers judged as having the strongest foreign accent in Spanish were not the ones who arrived in the US in childhood, but those speakers with less regular activation of their NL. The speaker judged as having the best pronunciation in Spanish was the one with the highest instrumental motivation to maintain it.

Keywords: attrition, foreign accent, factors.

1. INTRODUCTION

The influence of the native language (NL) sound system on a second or foreign language (L2 or FL) has been the object of a great body of research (see review in Strange, 1995, for example). Inter-linguistic influences may work both ways; the native language (NL) sound system can affect the sound system of the L2, but the NL sound system can also be affected by contact with the TL; this phenomenon is known as attrition (Seliger and Vago, 1991). Thus, phonetic attrition deals with the phonetic alterations in the NL caused, to a large extent, by the influence of the L2 on the NL. This usually takes place in the L2 environment, where speakers receive massive input from the L2 whilst contact with the NL is less frequent.

It has been found that one of the factors which may contribute to a lesser or greater degree of phonetic attrition of the NLs is the age factor (Köpke and Schmid, 2004), in the sense that the younger the speaker is when s/he arrives in the host country, the higher the chances of presenting NL phonetic attrition in their NL. There are significant correlations between both the amount of contact with the NL and "time elapsed since emigration"-also known as Length of Residence (LOR)-, with the extent of NL attrition a particular subject can suffer. However, these correlations are connected in that LOR only becomes relevant when there is not much contact with the NL; that is, NL phonetic attrition in the NL will occur over time in case the speaker does not keep much contact with his/her NL (De Bot *et al.*, 1991). On the other hand, the amount of contact with the NL or NL degree of activation has been found to be positively correlated with DFA in the L2 (Guion *et al*, 2000; Piske *et al*, 2001; Meador *et al*, 2000).

In the present paper, we focus on the phonetic attrition of the NL (Spanish) due to the influence of the target language (TL; English) in the TL environment (USA).

Taken into account the previously mentioned findings, we formulated the following research questions (i) is phonetic attrition in the NL always present in long-term immigrants in the TL environment (given that L2 input is maximized, whereas NL input may be minimized); (ii) is it possible to maintain balanced levels of phonological proficiency in both the NL (Spanish) and the third language (English) or, on the other hand, is proficiency in one language is inversely related to proficiency in the other language; (iii) can phonetic attrition in the NL be predicted by the dominance of a single factor (be it AOA, LOR, amount of language use, etc) or does it only occur by the collusion of several factors.

The factors we have analyzed in the present paper to see if they were related to phonetic attrition in the NL were: age of arrival (AOA) in the host country; length of residence (LOR) in the host country; degree of identification with the native culture (Spanish) and with the L3 culture (American); frequency of use of the NL (Spanish) and the L3 (English) and degree of foreign accent (DFA) in the NL and in the L3.

2. METHODOLOGY

2.1 The sample

For the present study, we recruited 13 first generation Spanish/Basque immigrants in the city of Reno (Nevada, USA), which is a traditional settlement for Spanish/Basque immigrants in the USA. They were interviewed and recorded in their three languages, namely two NLs (Spanish and Basque) and their L3 (English). In the present study, we concentrate on the NL (Spanish) and the L3 (English). Speakers also completed a questionnaire about language use, attitudes and biographical information (e.g. AOA, degree of identification with their native culture, etc.).

We have reproduced below two excerpts of the questionnaires regarding degree of identification with the host culture and with the native culture, and of that regarding frequency of use of both the NL and the L3 (both questionnaires are adapted from Lasagabaster & Huguet, 2007). The degree of identification with the NL/L3 questionnaire contained 7 questions. The frequency of use questionnaire had 18 questions.

For the degree of identification questionnaire we asked speakers to show their degree of agreement with statements such as the following: I like hearing Spanish spoken; The Spanish language is part of my cultural knowledge; if I have children, I would like them to be Spanish speakers regardless of other languages they may know.

In the frequency of language use questionnaire we asked speakers to indicate how often they performed activities such as the following in their different languages: Watching TV; Reading the press (newspapers, magazines); Talking to father/ mother /spouse, etc.

2.2 Accent judgements

2.2.1. Spanish accent judgements

Six native speakers of Spanish were recruited as native judges in Spain. All of them were university students with little knowledge of other languages. They listened to the interviews in Spanish in order to evaluate the attrition in this language in terms of DFA.

They rated the DFA of each of the speakers on a scale from 1 (no foreign accent) to 7 (very strong foreign accent). Apart from this, they were also asked to describe any phonetic characteristics which led them to their judgements for each individual speaker.

2.2.2 English accent judgements

Seven native speakers of American English were recruited as native judges. The recordings of the Spanish/Basque immigrants living in Reno were intermingled with recordings of three native speakers of American English and also with recordings of three Spanish/Basque students of English as a FL to provide extreme samples in the rating scale. Judges rated the DFA in English as in the ratings above for Spanish DFA.

3. RESULTS AND DISCUSSION

2.3 Results and discussion

In table 1 below, we can see some biographical data for the participants (age at the time of the recording, LOR, AOA), the results of both the Spanish and the English accent judgement results, the frequency of use of Spanish and English by these participants and their degree of identification with their native culture (Spanish) and with the L3 culture (American).

Following these ideas, we considered the fact that age influences might be mediated by the amount and regularity of language use (Flege, 1995), so that those speakers who use their NLs regularly will present a lesser degree of phonetic attrition in their NLs and a greater degree of interference with their L3.

The most relevant results are highlighted in bold. Regarding the frequency of use of the two languages, results were classified as either "frequent use of the language" or "infrequent use of the language". As we can see in the table below, all subjects reported a frequent use of English and most of them also reported a frequent use of the NL (Spanish), except for three participants (subjects 1, 3 and 10) who reported an infrequent use of their NL. Finally, the degree of identification with both the native culture (Spanish) and the host culture was provided through seven statements on this issue on a scale from 1 (strongly disagree) to 5 (strongly agree) (minimum score= 7; maximum score= 35. Given that the minimum score for each statement was 1=strongly disagree, and there were seven statements, the minimum score for the whole questionnaire regarding degree of identification was 7, whereas the maximum would be 35, in case s/he would completely agree with all the statements).

Table 1: Results of the questionnaires for each participant regarding age at testing, length of residence (LOR), age of arrival (AOA), frequency of use of Spanish and English (frequent vs. infrequent); degree of identification (DI) with the Spanish (NL) and American (L3) culture respectively (value ranges from 7=does not identify at all to 35= strongly identifies); mean degree of foreign accent judgements for Spanish and English (range 1-7)).

Subjects	AGE	LOR	AOA	Frequency of Spanish use	Frequency of English use	DI NL (7-35)	DI L3 (7-35)	DFA-NL (1-7)	DFA-L3 (1-7)
1	20	18	**2**	Infrequent	Frequent	**29**	27	**2.5**	**1.1**
2	60	54	**6**	Frequent	Frequent	**29**	30	4	**1**
3	64	44	20	Infrequent	Frequent	26	22	**4.3**	2.8
4	75	56	19	Frequent	Frequent	29	34	2	3.1
5	58	39	19	Frequent	Frequent	30	23	3.1	3.7
6	46	19	**27**	Frequent	Frequent	**29**	**26**	**1.8**	**4.8**
7	60	33	27	Frequent	Frequent	32	33	3.3	4.1
8	66	42	24	Frequent	Frequent	25	27	2.5	4.1
9	70	46	24	Frequent	Frequent	32	27	3.5	4.5
10	68	44	**24**	Infrequent	Frequent	**26**	**34**	**6.4**	**2.5**
11	63	36	27	Frequent	Frequent	29	22	2.8	5.8
12	45	21	24	Frequent	Frequent	28	32	2.1	3.2
13	61	33	28	Frequent	Frequent	24	**22**	**2.5**	**6.4**

All the participants in this study, who are first generation Spanish/Basque immigrants with a long residence in the host country, seem to have undergone some kind of phonetic attrition in their NL (Spanish). The participant who received the lowest DFA in Spanish was given a score of 1.8 (between 1= 'no foreign accent' and 2= 'near-native'), whereas the participant who received the highest DFA in his NL was placed nearly at the end of the scale with 6.4 (between 6= 'strong foreign accent' and 7= 'very strong foreign accent'). No participant obtained a "no foreign accent" mean judgement for his/her NL, so it seems that completely native phonetic proficiency in the NL is difficult to maintain for long-term immigrants in the L2 environment.
Against predictions, those two immigrants who were expected to present the highest level of NL attrition because of their early AOA (Flege *et al.*, 1995; Flege, 1999; Flege *et al.*, 1999; Flege *et al.*, 2006; Singleton & Ryan, 2004) in the host country (subjects 1 and 2) (e.g. De Bot *et al*,1991; Major, 1992) were not the ones who presented the highest degree of phonetic attrition in Spanish (highest rating in Spanish DFA), even though subject 1 reported an infrequent use of this language. Nevertheless, these two participants were found to have a high degree of identification with their native Spanish culture ((29), in both cases), higher than some other participants who arrived in the host country in adulthood. This could account for the good results they received in their DFA in Spanish (e.g. Moyer, 1999, Shumann, 1976).

The participant who presented the lowest degree of foreign accent (DFA) in Spanish (1.8) (subject 6), was very linguistically aware about the L1 and with a very high instrumental motivation (Moyer, 1999) to maintain it since she was a teacher of Spanish as an FL. Her degree of identification with the native culture was slightly higher (29) than her identification with the host culture (26). These factors could account for her phonetic maintenance of the NL (Spanish). In English, she was evaluated as having a more than moderate amount of foreign accent (4.8). Therefore, in her case, phonetic attainment in the L2 was inversely related to phonetic maintenance of the NL.

Subjects 4, 8 and 12 also present low DFAs in Spanish (2), (2.5) and (2.1) respectively. In the case of subject 4, he has the longest LOR, a frequent activation of the NL (Spanish) and also a high degree of identification with his native culture (29). Subject 8 also reported a frequent use of the NL. However, his degree of identification with his native culture (25) was not as high as in the case of subjects 4 and 6. In his case, his relatively late arrival in the host country (24) together with his frequent activation of the NL allowed him to maintain phonetic proficiency in this language. Finally, in the case of subject 12, three requirements which may favour the maintenance of phonetic proficiency in the NL are met; a relatively late AOA (24), frequent use of the NL and a high degree of identification (28) with his native culture.

The speaker who presented the highest degree of phonetic attrition in Spanish (6.4) (subject 10) was also the one who was evaluated as having the best pronunciation in American English (2.5) among the late arrivals (after the offset of the critical period), so in this case too, attainment in the L2 was inversely related to phonetic proficiency in the NL. This participant's AOA in the US was 24, higher than most of the other speakers, which contradicts the expectation of higher attrition being inversely correlated with AOA (Ventureyra *et al*, 2003). Explanations for his large L1 attrition and good DFA in English could be related to the fact that this speaker presented an infrequent use of the NL (Guion *et al*, 2000; Piske *et al*, 2001) and also one of the lowest degree of identification with his native culture, whereas he reported a frequent use of the L3 and also one of the highest rates in his degree of identification with the host culture (34) (e.g Moyer, 1999; Shumann, 1976).

Subject 13 shows the opposite pattern to that of subject 10. Subject 13 was rated 2.5 in his DFA (between 2 = near-native and 3 = less than moderate amount of foreign accent) in Spanish, whereas he was rated (6.4) in his DFA (between 6 = strong foreign accent and 7 = very strong foreign accent)) in English. As opposed to subject 10, this participant presented a frequent use of both the NL and the L3, but he presented a low degree of identification with both his native culture (24) and the host culture (22). This low degree of identification with the target culture could account for his poor phonetic performance in English as judged by the native judges (e.g. Moyer, 1999, Shumann, 1976) whereas his only incipient attrition in the NL maybe due to his late arrival and frequent use despite the low degree of identification with the NL.

Out of the 13 subjects, there are at least 5 cases for whom there is an important difference between DFA in one language vs. the other (speakers 2, 6, 10, 11, 12) and only one case (speaker 1) for whom there is native or near-native proficiency in the two languages. However, given the small number of participants in this study, these results may only be taken as tendencies until more speakers are analyzed allowing us to perform proper statistical comparisons.

Our data seem to suggest that NL attrition cannot be predicted by the dominance of a single factor (whether it is AOA, LOR, language activation, identification with the native and host cultures, etc.), but that it results from the collusion of different factors and factor combinations which may vary from speaker to speaker.

In order to further investigate these issues, we intend to obtain a larger speaker sample and Basque native judgements to compare them with the results for both the native Spanish and the English judgements. These additional data will allow us to explore in more detail the issue of whether it is possible to maintain balanced levels of proficiency in both the NLs and the TLs (Guion *et al*, 2000; Piske *et al*, 2001).

We also intend to investigate which are the factors that explain the level of phonetic proficiency attained in the L3 (Flege *et al*, 1995; Flege, 1999; Flege *et al*, 1999; Flege *et al*, 2006; Singleton & Ryan, 2004) and the presence of attrition in the NLs (De Bot *et al*, 1991; Seliger & Vago, 1991) by analyzing the questionnaires on language use and attitudes for the three languages and correlating DFA judgements.

Finally, we intend to look at the issue of whether one single factor may act as the most important one in predicting phonetic proficiency in the native language, whether phonetic attrition may only be predicted if several factors play together and whether some factors may offset the effect of others.

4. CONCLUSIONS

One of the most important conclusions we can get from this study is that AOA is not as determinant a factor in predicting phonetic attrition in the NL as expected (Köpke & Schmid, 2004). Since AOA is considered to be one of the main factors in the attainment of a native pronunciation in the L3 (Flege *et al.*, 1995; Flege, 1999; Singleton & Ryan, 2004) it has been proposed that the younger the speaker is when s/he arrives in the host country, the higher their chances of undergoing phonetic attrition in the NL (De Bot *et al*, 1991). This hypothesis does not appear to be confirmed by the current data.

On the other hand, the factor length of residence (LOR), which we expected to play some kind of influence on phonetic attrition in the NL, does not predict by itself either phonetic proficiency in the L3 nor phonetic attrition in the NL, although it has been observed to work along with other variables.

No single factor was seen to be able to predict phonetic attrition by itself. Other factors such as degree of identification with the NL culture (Moyer, 1999; Shumann, 1976) and frequency of use of the NL (Guion *et al*, 2000; Piske *et al*, 2001) seem to play an important role in our study in predicting NL phonetic attrition. In some cases, presenting a good pronunciation in one of the languages (NL or L3) seems to be related to having a bad pronunciation in the other language (e.g. subjects 10 and 13). However, near-native phonetic proficiency in both the NL and the L3 seems attainable (e.g. subject 1) if there is collusion of several favourable factors (early AOA and a high degree of identification with the NL and the L3. Further analysis with a larger sample is needed to substantiate the observed tendencies.

ACKNOWLEDGEMENTS

This work was supported by a grant to the first author from the University of the Basque Country (Real Decreto 63/2006, de 27 de enero).

5. REFERENCES

De Bot, K., Gommans, P., Rossing C. (1991) L1 loss in an L2 environment: Dutch Immigrants in France. In H.W. Seliger, R.M. Vago (Eds.) *First Language Attrition,* UK: Cambridge University Press. pp. 87-98.

Flege, J.E. (1995) Second language speech learning: Theory, findings and problems. W. Strange (ed.) *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*. Timonium, MD: York Press. pp. 233-277.

________ (1999) Age of learning and second language speech. In D. Birdsong (ed.). *Second Language Acquisition and the Critical Period Hypothesis* (pp. 101-131). London: Lawrence Erlbaum Associates, Publishers.

________, Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Tsukada, K. (2006) Degree of foreign accent in English sentences produced by Korean children and adults. *Journal of phonetics*, 34, 153-175.

_________, Munro, M.J., & Mackay, I.R.A. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, 95, 3623-3641.

_________, Yeni-Komshian, G., & Liu, H. (1999) Age constraints on second language acquisition. *Journal of Memory and Language*, 41, 78-104.

Guion , S.G., Flege, J.E., Loftin , J.D. (2000) The effect of L1 use on pronunciation in Quichua –Spanish bilinguals. J. Phonetics 28, 27-42.

Köpke, B., Schmid, M. (2004). The next phase. *First Language Attrition. Interdisciplinary perspectives on methodological issues*. M. Schmid, B. Köpke, M. Keijzer and L. Weilemar (Eds.)Studies in Bilingualism (SiBil), 28, John Benjamins Publishing Company. pp. 1-43.

Lasagabaster, D., Huguet, A. (2007) Multilingualism in European Bilingual Contexts: Language Use and Attitudes. Clevedon: Multilingual Matters, pp. 6-13.

Major, R.C (1992) Losing English as a first language. *The Modern Language Journal*, 76, 190-208.

Meador, D., Flege, J. E., Mackay, J. (2000) Factors affecting the recognition of words in a second language. *Bilingualism: Language and Cognition*, 3, 55-67.

Moyer, A. (1999) Ultimate attainment in L2 phonology. The critical factors of age, motivation and instruction. *Studies in Second Language Acquisition*, 21, 81-108.

Piske, T, Mackay, I.R.A., Flege, J.E. (2001) Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, 29, 191-215.

Seliger, H. W, Vago, R.M. (Eds) (1991) *First Language Attrition*. Cambridge, UK: Cambridge University Press.

Shumann, J. (1976) The Acculturation Model for second language acquisition. In R. Gringas (ed.) Second Language Acquisition and Foreign Language Teaching. Washington: Center for Applied Linguistics.

Singleton, D., and Ryan, L. (2004) (2nd edition) *Language Acquisition: The Age Factor*. Clevedon [etc.]:Multilingual Matters.

Strange, W. (1995). Cross-language studies of speech perception. In: Winifred Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, pp. 3-45. Baltimore: York Press.

Ventureyra, V. A. G., Pallier, C., Yoo, H. (Jan. 2005) The loss of first language phonetic perception in adopted Koreans. *Journal of Neurolinguistics*. 17 (1), 79-91.

The production of palatalized and unpalatalized consonants in Russian by American learners

Jane F. Hacking

University of Utah
j.hacking@utah.edu

ABSTRACT

The Russian phonological system is characterized by the opposition of a set of palatalized and unpalatalized consonants, a phonemic distinction not found in English. This paper reports on research into the production of this contrast by advanced American learners of Russian. Results show that the likelihood of Russian native speaker listeners' accurately perceiving the American talkers' intended target depends both on the position and type of consonant being produced. Words containing palatalized and unpalatalized /p/, /t/. /s/, /n/, /l/, /r/, in the environment both before the back vowel /o/ and in word final position were recorded by the Russian learners in a carrier phrase. Tokens were subsequently extracted from the carrier phrase and the resulting stimuli were presented to native Russian speakers as a forced-choice word identification task. Results show that the palatalization contrast was produced most successfully when the consonant appeared in pre-vocalic position. In word final position the results were generally poor, although accuracy did vary somewhat depending on which consonant was at issue. The paper elaborates on these results and proposes a scale of consonantal difficulty for American learners of Russian, appealing to the notion of sonority hierarchy and general work on the acoustic and articulatory properties of Russian palatalized consonants (Kavitskaya et. al. 2009; Kochetov 2004, 2006).

Keywords: second language acquisition, Russian palatalization.

1. INTRODUCTION

The opposition between palatalized and unpalatalized consonants is a fundamental feature of the Russian phonological system, and it poses a challenge for English speaking learners of Russian since no such phonemic opposition exists in their native language. This paper reports on research that addresses two questions. First, to what extent have advanced American learners of Russian mastered the production of this phonemic contrast? And second, is the ability of these learners to produce palatalized and unpalatalized segments dependent on type of consonant or the consonant's position? Results show that a consonant's position affects greatly the degree to which the American learner of Russian can produce the contrast at a level perceptible to the native speaker listener. Furthermore, the successful production of the phonemic contrast varies by type of consonant. The sonority hierarchy and general work on the acoustic and articulatory properties of Russian palatalized consonants (Kavitskaya et. al. 2009; Kochetov 2004, 2006) suggest inherent properties of the sounds' articulations that may help to explain the learners' performance.

2. METHOD

Six undergraduate American students of Russian performed a production task; they will be referred to as 'talkers' in this paper. All of them were male, between the ages of 22 and 30. Each of them had spent two years living in Russia and at the time of the recording was enrolled in a course on Russian phonetics and phonology. Six Russian native speakers served as 'listeners' for the experiment. Four were female and two were male. All were between the ages of 18 and 40. Four of the six were currently in the United States as Fulbright students or scholars, one was the spouse of a Fulbright scholar, and the sixth participant was an undergraduate student. They were paid for their participation in the study. All reported Russian as their native language, with one reporting as bilingual with Russian and Tatar. Four of the six report beginning to learn English between the ages of 6 and 10, one reported age 15, and the sixth participant began learning English at age 37. All were proficient speakers of English able to communicate easily to set up the experiment time and discuss procedures. All the listeners reported normal hearing.

2.1. Production task

The Russian learners were recorded in the Speech Acquisition Lab at the University of Utah reading words constituting minimal pairs. Words for the production task were selected to ensure that there was a pair containing each of palatalized and unpalatalized /p/, /t/, /s/, /n/, /l/, /r/. The words presented each consonant in two environments: before the back vowel /o/ and in word final position. For the speaking task, 24 words were presented to the Russian learners each in the carrier phrase: *Vot opjat slovo* _____ (Here again is the word _______). Each of the 24 words occurred three times in the task resulting in three recordings of each word. Words were subsequently extracted from the carrier phrase using Praat (Boersma 2001) and scaled for intensity.

1.2. Perception task

Tokens from the production task were presented to native Russian speakers as a forced-choice word identification task requiring the participant to press a left or right key on a keyboard depending on which word on the screen they thought they heard. The three recordings of each item produced by each speaker were presented twice, once with the visual cue on the right side of the screen and once with the visual cue on the left side of the screen. The resulting 864 word tokens (24 words x six tokens x six talkers) were presented in five blocks with four subject-controlled breaks between blocks. Data was presented and captured using the DMDX experiment presentation software (Forster & Forster 2003). The task took approximately forty-five minutes after which each participant completed a brief questionnaire to gather basic biographical information, details about study and exposure to English and other languages, and to ascertain whether he or she had any hearing difficulties.

3. RESULTS

I report the data in terms of how accurate the talkers' production of the consonants was. Accuracy here is understood to mean the extent to which the native speaker listeners' perceptions matched the intended target of the talker. Overall mean accuracy for all listeners was 0.75. My focus is how the items rank based on an 83-100% accuracy rating (that is, five of six or all six of the listeners' perceptions matched the talkers' intended target). Overall the listeners were highly consistent in their performance. Mean accuracy rates for each listener are as follows: Listener 1=.76; Listener 2=.75; Listener 3=.74; Listener 4=.74; Listener 5=.76; Listener 6= .75. A breakdown of individual items by accuracy, with a threshold of at least five of the six listeners correctly identifying the item produces a clear hierarchy as shown in Table 1.

Table 1: Percent of item tokens receiving accurate identification by five or six of the listeners.

Percent of tokens that 5 or 6 listeners correctly identified	Target word
100%	s^{j}ok, p^{j}or, t^{j}ok, nos, sok, tok, lot, kon, otrbos
97%	por, grob,
94%	n^{j}os, l^{j}ot
86%	top, udar
83%	ugol

78%	brat
47%	gr^job
28%	$ugol^j$
19%	kon^j
.06%	top^j
0%	$otbros^j$, $udar^j$, $brat^j$

These results show that the learners have either largely mastered the phonemic contrast or they have not, and whether or not this has occurred depends largely on the environment in which the contrast occurs. With the exception of gr^job, every item is either accurately identified at a rate of 78% or higher or is accurately identified less than 28% of the time. Furthermore, the way the items pattern is highly consistent. Items accurately perceived include all prevocalic contrasts (except gr^job), and all unpalatalized consonants in word final position. By contrast, all six items with a palatalized consonant in word final position had very low accuracy ratings. These results clearly indicate that production of palatalized consonants in word final position presents particular difficulty for English learners.

A closer look at the accuracy data for items with word final palatalized consonants suggests that of the six consonants tested, some presented a greater challenge to the learners than other. The palatalized lateral /l^j/ was most accurately produced, followed by the palatalized nasal /n^j/. The other four palatalized consonants /p^j/, /s^j/, /r^j/ and /t^j/ were either wholly or almost wholly inaccurately produced. These results suggest that the six consonants can be arrayed in terms of difficulty for these learners as follows:

←Most difficult--Least difficult→

/r^j/ /p^j/ /n^j/ /l^j/

/s^j/

/t^j/

4. DISCUSSION

Results of the perception task clearly show that that production of palatalized consonants in an environment without a following vowel presents considerable difficulty for these American learners. Furthermore, the data suggest a hierarchy of difficulty for the six consonants tested. The four consonants least accurately produced by the American learners were /r^j/, /s^j/, /t^j/ and /p^j/, that is, two stops: /t^j/ and /p^j/, one fricative: /s^j/, and one trill: /r^j/. The two more accurately produced consonants were the nasal: /n^j/, and the lateral: /l^j/. In the remainder of the paper I suggest possible explanations for this hierarchy and propose some directions for future research.

One possible approach to explaining the varied accuracy ratings for these palatalized consonants in word final position might be found in the sonority hierarchy, which would differentiate the tested consonants as follows:

←Least sonorous--Most sonorous→

/t^j/ /s^j/ /n^j/ /r^j/

/p^j/ /l^j/

Such an analysis might suggest that palatalization is somehow more compatible with sonority and hence easier for the second language learner to produce when the consonant falls on the more sonorous end of the hierarchy. These data are somewhat, but not completely consistent with such a proposal. The least sonorous of the three tested consonants, /t^j/, /p^j/ and /s^j/, are among the four least accurately produced by our talkers.

The more sonorous /n^j/ and /l^j/ displayed greater accuracy. However, /r^j/, which ranks with /l^j/ in terms of sonority was actually one of the least accurately produced by our talkers. Further explanation requires that we appeal to other aspects of the sounds' articulation.

The Russian stop consonants /t/ and /p/ share certain similarities with English /t/ and /p/ but also differ in significant ways. Most crucial for our purposes is the status of the final release. In English, this final release is optional, while in Russian it is not. As noted by Kochetov (2006: 116), "[t]hus, Russian utterance-final stops are always audibly released, with the releases accompanied by strongly aspirated either velarized or palatalized off-glides." His work shows that the release of stop consonants in Russian is the main acoustic cue for palatalized consonants (Kochetov 2004). The propensity for English speakers not to release stop consonants may well account for our talkers' failure to accurately produce palatalized stops in word final position. Further study of the acoustic data is needed to explore this hypothesis.

Kavitskaya et. al. (2009) present acoustic and articulatory data for native speaker production of Russian /r/ and /r^j/ in three different positions: word-initial, word-medial and word-final. They demonstrate that there are significant differences between the two in every position. First, /r/ shows higher amounts of trilling than /r^j/ in every environment. Second, the trilling vibration of /r/ is higher in all environments. The focus of their research was to offer possible explanations for the historical instability of /r^j/ across the Slavic languages and they conclude that the production of /r^j/ has the trill aspect which demands retraction of the dorsum, while the palatalization feature requires its fronting. In short, the articulation of /r^j/ requires "conflicting physical constraints on the tongue dorsum." (2009: 1). It seems reasonable to suspect that such conflicting articulatory constraints may well pose difficulty for learners of Russian, thereby accounting for our talkers' lack of accuracy for all items with word final /r^j/. Support for the inherent articulatory difficulty of /r^j/ may also help to explain the mixed accuracy ratings for the token gr^job. Accuracy for this item was markedly lower than for all other items with a prevocalic palatalized consonant.

The liquid consonants tested were the lateral /l^j/ and the nasal /n^j/. Accuracy ratings for each were higher than for the other four palatalized consonants. This may indeed be connected to the inherent sonority of the two sounds, but I want to propose as well an additional, metalinguistic, explanation for the particular accuracy of /l^j/. Advanced American learners of Russian, when queried about their pronunciation difficulties, always list 'soft l' as among the Russian sounds they find most challenging. Although they may refer to palatalization in general as a pronunciation difficulty they are aware of, they tend not to single out other palatalized consonants. I argue that it might be precisely this heightened awareness that accounts for these subjects' greater success with word final /l^j/. I propose that their experience primes them to notice and make an extra effort with /l^j/, particularly when it is orthographically realized with the soft sign (*mjagkij znak*) as it is in word final position.

This paper does not address differences among the talkers, although preliminary analysis of the data suggests that such differences are present. For example, all the accurate productions of word final /l^j/ were produced by the same two talkers. Further research will compare questionnaire data provided by the talkers in order to identify possible correlations between accuracy ratings and talker variables.

5. CONCLUSION

The results reported here show that the production of the palatalization contrast in Russian by American learners is a complex phenomenon. Successful production of the palatalized unpalatalized contrast is highly dependent on the consonant's environment. The data show that these proficient L2 speakers of Russian do produce palatalized consonants that are perceivable as such by native speaker listeners when the consonant appears before the vowel /o/. This was not the case when one of the six tested consonants occurs in word final position. While the learners were highly accurate in their production of unpalatalized consonants in this environment, their production of palatalized consonants was largely unsuccessful. However, accuracy ratings for palatalized consonants in word final position were not uniform in the data. Rather, certain consonants posed more difficulty than others. Sonority of the consonant as well as additional information about its articulatory and acoustic properties help to explain this variation in production accuracy.

6. ACKNOWLEDGEMENTS

This work was supported by a Faculty Development Grant from the College of Humanities, University of Utah, resources of the Speech Acquisition Lab, University of Utah, and the Undergraduate Research Opportunity Grant, University of Utah. I wish to thank Dr. Rachel Hayes-Harb and Lanise Thompson for their contributions to this project.

7. REFERENCES

Boersma, P. 2001. Praat, a system for doing phonetics by computer, Glot Int. 5, 341-345.

Forster, K. I., J. C. Forster. 2003. DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, and Computers*, 35. 116-124.

Jones, D., D. Ward. 1969. *The phonetics of Russian.* Cambridge: Cambridge University Press.

Kavitskaya, Darya, Khalil Iskarous, Aude Noiray, Michael Proctor. 2009. Trills and palatalization: Consequences for sound change. http://www.mproctor.net/docs/kavitskaya09_FASL.pdf.

Kochetov, Alexei. 2004. Perception of place and secondary articulation contrasts in different syllable positions: Language-particular and Language-independent asymmetries. *Language and Speech*, 47(4). 351-382.

Kochetov, Alexei. 2006. Testing licensing by cue: A case of Russian palatalized coronals. *Phonetica* 63. 113-148

Language-specific differences in the weighting of perceptual cues for labiodentals

Silke Hamann, Paul Boersma, Małgorzata Ćavar

University of Düsseldorf, University of Amsterdam, University of Zadar
hamann@phil-fak.uni-duesseldorf.de, paul.boersma@uva.nl, mecavar@unizd.hr

ABSTRACT

Cross-language perception provides insight into the use of perceptual cues to native segments and their application to segments in a different language. In the present study we test the perception of the three Dutch labiodentals /f, v, ʋ/ by listeners of German, English, Croatian and Polish in a forced-choice identification task. We test whether the perceptual boundaries on the auditory dimensions of harmonics-to-noise ratio and duration are more similar for listeners from the same language family (German and English versus Croatian and Polish) or whether these boundaries are more similar for listeners with the same number of labial categories in their native languages (German and Croatian with four labials versus English and Polish with five). Our findings show that the same number of labial categories results in similar perceptual boundaries along the two auditory dimensions, and that language family does not influence the location of the boundaries.

Keywords: cross-language perception, perceptual cues, labiodental, fricative, approximant

1. INTRODUCTION

Adult listeners often have difficulties perceiving a category contrast in a second or foreign language (L2) that does not exist or differs from the contrasts in their native language (L1). Several factors have been held responsible for this phenomenon, among them the auditory similarity between L1 and L2 categories (e.g., Best et al. 2001), which is often determined by the number of categories on the relevant phonetic dimension. Little attention has been paid to the role of language family in L2 perception.

In this study we investigate the question whether a genetic relationship between L1 and L2 influences L2 speech perception, or whether the perception depends solely on the number of phonological categories. We test the perception of the Dutch labiodentals /f, v, ʋ/ by native listeners of four languages from two language families: two Germanic languages, German and English, and two Slavic languages, Polish and Croatian. All four languages have the two labiodentals /f, v/ and the two bilabial plosives /p, b/. They differ, however, with respect to the presence of the labiovelar approximant /w/: while this sound is absent from German and Croatian, it is present in English and Polish. The labiodentals and further labials in the inventories of these four L1 languages and the L2 language Dutch are summarized in Table 1.

Table 1: Nonnasal labial consonants of the five languages investigated in the present study.

Language family	Language	Labiodentals	Other labials
Germanic	Dutch	f, v, ʋ	p, b
	German	f, v	p, b
	English	f, v	p, b, w
Slavic	Croatian	f, v	p, b
	Polish	f, v	p, b, w

This choice of languages allows us to make a preliminary comparison between the influence of the number of labial categories (four versus five) and the influence of language family (Germanic versus Slavic) on the perceptual cue weighting for labiodentals. We expect the influence of the labial inventory to be more important than the influence of language family, i.e. listeners of languages with a similar inventory (German and Croatian versus English and Polish) to behave more similarly than listeners of languages from the same language family (i.e. German and English versus Croatian and Polish) in the perception of the Dutch labiodentals. This expectation is based on the findings in the study by Boersma & Hamann (2008), where it was illustrated with computer simulations of diachronic data that any phoneme inventory with the same number of categories (one to five) along one auditory dimension must end up with the same stable system, i.e. with the same location of category boundaries on this dimension. Though Boersma & Hamann's study only looked at sibilant inventories along the auditory dimension of spectral mean, we transfer those findings in the present study to labiodental inventories that can be characterized by multiple auditory dimensions.

Acoustically, labiodentals are characterized by low-amplitude friction noise that is spread over the frequency range in a spectrum. The perception of this friction noise is influenced by the presence or absence of periodicity caused by vocal fold vibration. The more voiceless a sound is, the more fricated it sounds, and the more voiced it is, the more periodic its frication is and the less fricated it sounds. This perceptual correlation can be captured by the acoustic measure of harmonics-to-noise ratio (Yumoto et al. 1982, Boersma 1993). Hamann & Sennema (2005a) illustrate that the harmonics-to-noise ratio clearly distinguishes between the three Dutch labiodentals and between the two labiodentals in German. Their study further shows that Dutch and German /f/ have almost identical harmonics-to-noise ratios (around −1.5 dB), but German /v/ has a much higher ratio than Dutch /v/ (German 15.3 dB, Dutch 0.8 dB), coming close to the ratio for Dutch /ʋ/ (18.8 dB). In a perception study, Hamann & Sennema (2005b) find that German naïve listeners perceive Dutch /ʋ/ as their native /v/ in all of the cases, and Dutch /v/ as their /f/ in a considerable number of cases. In a perception experiment on the boundary differences between Dutch and German labiodentals along the dimension of harmonics-to-noise ratio, Hamann et al. (2007) find that the location of the perceptual boundary between the labiodentals /f/ and /v/ in Dutch and German differs. This leads us to use the harmonics-to-noise ratio as one of the auditory dimensions in the present study.

A further acoustic and auditory difference between fricatives and approximants, and between voiced and voiceless fricatives, is duration: fricatives are longer than approximants (e.g. Romero Gallego 1995 for Spanish) and voiceless fricatives are longer than voiced ones (e.g. Stevens et al. 1992 for English; Mees and Collins 1982 for Dutch; Jessen 1998 for German; Hamann & Sennema 2005a for the Dutch labiodentals).

Though formant transitions of the preceding and following vowels and the intensity of the consonant yield further possible auditory cues for the distinction of the three labiodentals, we concentrate in the present study on the two dimensions of harmonics-to-noise ratio and duration.

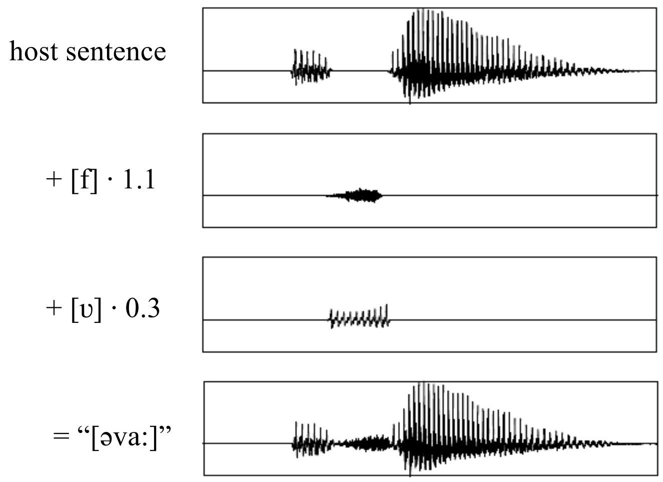
2. METHOD

2.1. Stimuli

The stimuli were synthesized on the basis of natural speech recordings from one male Dutch native speaker from the South of the Netherlands. Recordings were made in a sound-proof booth to a Pioneer PDR-555 CD recorder, using a Sennheiser MKH-105 microphone. The recordings included one token each of the three Dutch labiodentals /f, v, ʋ/ in a VCV context with a preceding [ə] and a following [aː] (from the sentence "Hoor je _a", *Do you hear _a*), with stress on the last vowel.

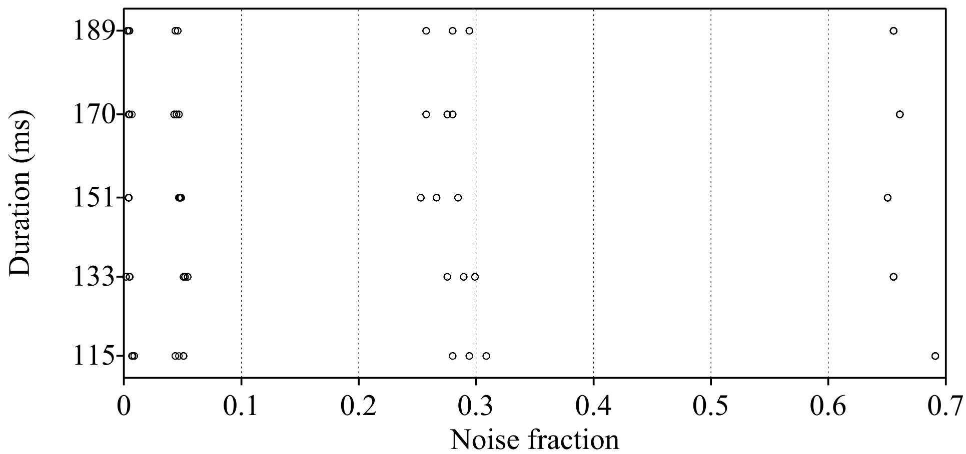
We manipulated the acoustic parameters of harmonics-to-noise ratio, duration, intensity and transition with the help of the Praat program (Boersma & Weenink 2009) in the following way. We cut the two labiodental tokens [f] and [ʋ] from their surrounding. The [f] was multiplied by the factor 0, 0.005, 0.25, 0.5, 0.75, 0.9, 1, 1.1, 1.8, 2.3, or 3. The [ʋ] was multiplied by the factor 0, 0.22, 0.3, 0.6, 0.7, 1, or 1.1. The two resulting sounds were added together and added to a host sentence, which we had created by using a recording of [əʋa:] and replacing the [ʋ] by silence. An illustration of this manipulation process is given in Figure 1.

Figure 1: Example of stimulus creation.



The stimuli resulting from this manipulation have the following values for harmonics-to-noise ratio: around −3.5 dB (close to a natural [f]), +3.5 dB (close to the natural recording of [v]), +13 dB, +22 dB (close to natural [ʋ]) and no ratio (close to natural [p]). Figure 2 shows how these values are dispersed in a two-dimensional plane. To achieve a moderate degree of perceptual equidistance the logarithmic harmonics-to-noise ratio in dB was first converted to its nonlogarithmic counterpart (HNR), then to a *noise fraction* between 0 and 1 (= HNR / (HNR + 1)).

Figure 2: Distribution of stimuli on the acoustic dimension of noise fraction (horizontal) and duration (vertical).



The duration of the stimuli was manipulated by lengthening the [ʋ] (by duplicating part of the original signal), shortening the [f] (by removing part of the original signal), and adjusting the duration of the silence phase in the frame sentence. Resulting duration values of the stimuli are 115 ms (the duration of the natural [ʋ]), 133 ms (natural [v]), 151 ms, 170 ms, and 189 ms (natural [f]).

The intensity of the stimuli was manipulated identical to the harmonics-to-noise ratio (the two could not be manipulated independently from each other). Intensity values of the stimuli are: 45 dB (identical to the natural recording of [f]), 50 dB (natural [v]), and 55 dB (natural [ʋ]).

The transitions were manipulated by removing two glottal waves from the transitions in the host sentence, yielding two transitional values: long transitions (from the recording of [əʋa:]) and short transitions (with glottal waves removed, mimicking an [f]-context).

The combination of all these parameter values results in 120 stimuli (four for voicing-to-friction ratio × five for duration × three for intensity × two for transition). Approximately 18 of these 120 stimuli do not sound like a labiodental fricative or approximant (all of them have a factor of 0.005 for [f] and a low factor for [ʋ]) but rather like a bilabial plosive [b] or [p]. This led us to include the bilabial plosives as possible answer categories (see 2.3 below). To avoid a range effect, we added 10 further stimuli with silence in the host sentence (five durational values × 2 transitions). This results in a total of 130 stimuli.

In the following we only report on the parameters of noise fraction and duration.

2.2. Listeners

The listeners of this experiment were a total of 94 participants with either German, English, Croatian or Polish as their native language. We tested 31 German listeners at the University of Potsdam. They were 20–41 years old. The English group consisted of 20 participants, 18–47 years of age, tested at University College London. The Polish group consisted of 23 participants, 20–36 years of age, and tested at the University of Warsaw. The Croatian group consisted of 20 listeners between 19 and 29 years of age, who were tested at the University of Zadar. The participants were mostly university students, though some were faculty members. No participants had lived outside of their country for longer than six months and none of them reported any hearing impairment.

2.3. Task

The task was a forced-choice identification task. Each of the 130 stimuli was repeated once, giving a total of 260 stimuli. This total set was randomized for each listener and presented via headphones. The set of answers included the labiodental fricatives (voiced and voiceless) and the labial plosives for all groups. The English and Polish groups had in addition the labiovelar approximant. German and Croatian listeners thus had four answer categories, the English and Polish listeners five. Orthographic representations of these labials in the native language of the listeners were presented on a computer screen. These answer categories are given in Table 2.

Table 2: Answer categories (in orthographic representation) for the five language groups. German and Polish <w> is /v/, and Polish <ł> is /w/.

Language	Answer categories
German	f w p b
English	f v p b w
Croatian	f v p b
Polish	f w p b ł

Participants heard one stimulus at a time and had to click on the consonant they thought they had heard. They could not listen to the stimulus a second time.

3. RESULTS

Of the 94 listeners we remove two outliers, one English and one Polish native speaker. These are the listeners who show a great lack of consistency in their answers: for more than 50 percent of the 120 target stimuli they gave a different response to the two replications.

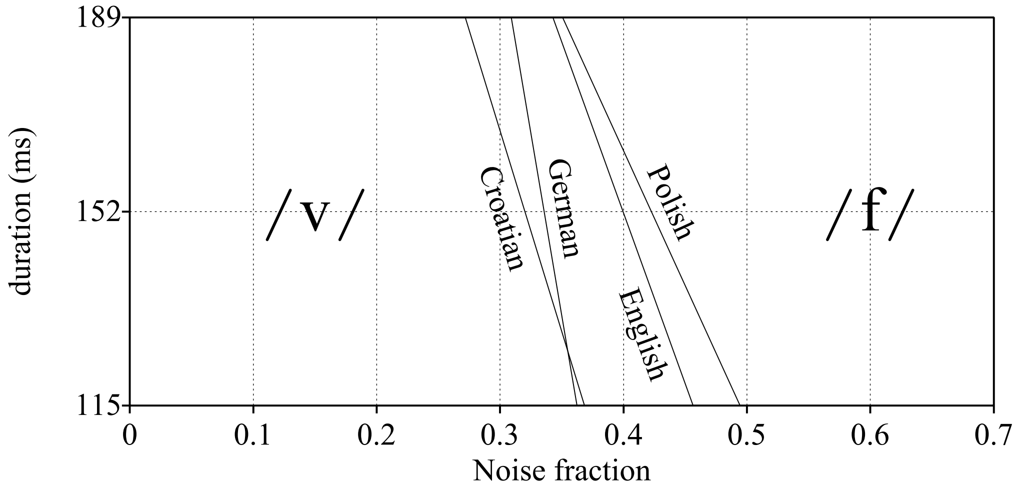
On the basis of the results of the identification test we perform a logistic regression analysis on the responses of each of the 92 speakers, with noise fraction and duration as the factors and the /f/–/v/ choice as the dependent variable. From the logistic regression coefficients of a speaker we define that speaker's *boundary location for noise fraction* as - (*intercept* + *duration coefficient* × 151 ms) / *noise fraction coefficient*, and that speaker's *angle between noise fraction and duration* as arctan2 (- *noise fraction coefficient*, *duration coefficient*). The four /f/–/v/ *boundary lines* in Figure 3 reflect the medians (over the speakers) of these two quantities for the four languages, as follows.

The boundary location for noise fraction is shown in Figure 3 as the noise fraction value where a boundary line intersects the horizontal line at the middle duration of 151 ms: for Croatian listeners it lies at 0.320, for German listeners at 0.336, for English listeners at 0.400, and for Polish listeners at 0.423. A Kruskal-Wallis test reveals that these boundary locations are not the same for all four languages ($\chi^2 = 25.663$, $df = 3$, $p = 0.000011$); in fact, the four languages seem to divide into two groups, namely Croatian and German on the one hand, and English and Polish on the other: the difference between the boundary locations for Croatian and English, as measured by Wilcoxon's rank sum test, is significant ($p = 0.013$), and so are the differences between Croatian and Polish ($p = 0.010$), German and English ($p = 0.004$), and German and

Polish (p = 0.0009), while the differences in boundary location between Croatian and German listeners and between English and Polish listeners are entirely nonsignificant (p = 0.685 and 0.418, respectively).

The angle between noise fraction and duration is shown in Figure 3 as the slope of the boundary line: for Polish listeners the slope is -514 ms, for English listeners -651 ms, for Croatian listeners -767 ms, and for German listeners -1390 ms (the steepest). The difference in angles is quite significant between Poles and Germans (p = 0.003), and less significant between Croatians and Poles (p = 0.044) and between Croatians and Germans (p = 0.056).

Figure 3: Boundaries between labiodental categories in German, English, Croatian and Polish.



As the angles of the boundaries in Figure 3 indicate, all four languages use noise fraction as a major cue and duration as a secondary cue to distinguish between /f/ and /v/: noisier and longer stimuli are more likely to be identified as /f/.

4. DISCUSSION AND CONCLUSIONS

The present study suggests that the size of the native labial inventory influences the perception of Dutch labiodental sounds to a greater extent than the genetic relationships between languages do. These findings are in accordance with the hypothesis that the size of the inventory determines the location of the category boundary between /f/ and /v/. This hypothesis was based on Boersma & Hamann's (2008) results for sibilants, which were differentiated on a single acoustic/auditory dimension. In the present study, at least two dimensions seem to play a role in the perception of labiodentals, namely noise fraction and duration. For the listeners in the four languages we tested here, the dimension of noise fraction was a major cue, and duration only a minor cue. This is probably due to the fact that noise fraction is a more salient and static (i.e. non-durational) cue. We therefore expect a similar perceptual preference for noise fraction in other languages as well. For segmental contrasts with several auditory dimensions and no phonetically-based preference for one of these dimensions, we expect languages to differ much more in their cue weighting and in the location of their perceptual boundaries, even if the segmental inventories in these languages have the same size. These expectations have to be tested in future studies.

Four languages from two language families, as well as one particular type of contrast, are of course not sufficient to make larger generalizations on the influence of language family on the perception of L2 sounds. Extensive future studies are required to support the conclusions of the present study.

5. ACKNOWLEDGEMENTS

This research was supported by NWO grant No. 016-064-057 to Silke Hamann, NWO grant No. 016-094-610 to Paul Boersma, and grant No. 269-2120920-0896 by the Croatian Ministry of Science and Education to Małgorzata Ćavar. We are thankful to Olga Głębicka, Paul Iverson, Melanie Preece-Pinet, Paweł Rydzewski, Anke Sennema, and Kordula Witjes for their support and/or help in testing the subjects.

6. REFERENCES

Best, C. T., McRoberts, G. W., Goodell, E. 2001. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America* 109. 775–794.

Boersma, P. 1993. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences (University of Amsterdam*) 17. 97–110.

Boersma, P., Hamann, S. 2008. The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology* 25. 217–270.

Boersma, P., Weenink, D. 2009. Praat: doing phonetics by computer (version 5.1.04) [Computer program].

Hamann, S., Sennema, A. 2005a. Acoustic differences between German and Dutch labiodentals. *ZAS Working Papers in Linguistics* 42. 33–41.

Hamann, S., Sennema, A. 2005b. Voiced labiodental fricatives or glides — all the same to Germans? In: Hazan, V., Iverson, P. (eds.) *Proceedings of the Conference on Plasticity in Speech Processing*. London: UCL. 164–167.

Hamann, S., Sennema, A., Boersma, P. 2007. Dutch labiodental perception in German L2 learning. Talk presented at *New Sounds 2007*, Florianópolis.

Jessen, M. 1998. *Phonetics and phonology of tense and lax obstruents in German.* Amsterdam: Benjamins.

Mees, I. & Collins, B. 1982. A phonetic description of the consonant system of Standard Dutch (ABN). *Journal of the International Phonetic Association* 12(1). 2–12.

Romero Gallego, J. 1995. Gestural organization in Spanish. An experimental study of spirantization and aspiration. PhD. diss. University of Connecticut.

Stevens, K, Blumstein, S., Burton, M., Kurowski, K. 1992. Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters. *Journal of the Acoustic Society of America* 91. 2979–3000.

Yumoto, E., Gould, W.J., Baer, T. 1982. Harmonics-to-noise ratio as an index of the degree of hoarseness. *Journal of the Acoustic Society of America* 71. 1544–1550.

Production of English interdental fricatives by Dutch, German, and English speakers

Adriana Hanulíková & Andrea Weber

Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
{A.Hanulikova; A.Weber}@mpi.nl

ABSTRACT

Non-native (L2) speakers of English often experience difficulties in producing English interdental fricatives (e.g. the voiceless [θ]), and this leads to frequent substitutions of these fricatives (e.g. with [t], [s], and [f]). Differences in the choice of [θ]-substitutions across L2 speakers with different native (L1) language backgrounds have been extensively explored. However, even within one foreign accent, more than one substitution choice occurs, but this has been less systematically studied. Furthermore, little is known about whether the substitutions of voiceless [θ] are phonetically clear instances of [t], [s], and [f], as they are often labelled. In this study, we attempted a phonetic approach to examine language-specific preferences for [θ]-substitutions by carrying out acoustic measurements of L1 and L2 realizations of these sounds. To this end, we collected a corpus of spoken English with L1 speakers (UK-English), and Dutch and German L2 speakers. We show a) that the distribution of differential substitutions using identical materials differs between Dutch and German L2 speakers, b) that [t,s,f]-substitutes differ acoustically from intended [t,s,f], and c) that L2 productions of [θ] are acoustically comparable to L1 productions.

Keywords: segmental substitutions, interdental fricatives, Dutch, German, English.

1. INTRODUCTION

One characteristic feature of speech produced by L2 speakers is its accent. Foreign accents result from a combination of subphonemic, segmental, and suprasegmental deviations from the target language. A common phenomenon at the segmental level is substitution, by which we mean the replacement of a specific L2 phoneme by another phoneme, usually one that occurs in the native phoneme inventory of the speaker. Substitutions can result, for example, from the lack of a native counterpart for a given L2 phoneme, and are often subject to variation, such as in the L2 production of English interdental fricatives. Since phoneme inventories of most European languages lack interdental fricatives, many L2 speakers of English have difficulties producing them correctly and often substitute them. German and European-French learners of English, for example, often replace the voiceless interdental fricative [θ] with [s], while Dutch and Canadian-French speakers are reported to prefer [t] (for an overview, see Brannen 2002). Phoneme-identification studies show that [θ] is perceptually most often confused with the acoustically similar [f] by native as well as by various L2 listeners, and less frequently confused with [t] or [s] (Brannen 2002; Cutler et al. 2004; Hancin-Bhatt 1994a; Miller and Nicely 1955; Tabain 1998). Given the acoustic similarity with [f], it is rather surprising that [f] is not the most common substitution in English L2 speech, not even when [f] is available in the L1 phoneme inventory of the L2 speakers. Note that substitutions of voiceless (and voiced) fricatives are not restricted to L2 speech; they also occur in dialects of English, with reported instances of [f] in Cockney (Wells 1982), and of [t] in Irish English (Hickey 2004). However, in contrast to L2 studies, the production frequencies of these dialectal substitutions across L1 speakers are seldom systematically studied or reported (see McGuire 2003).

Prior research has explored the causal relationship of variation in [θ]-substitutions across L2 learners with different L1 backgrounds, and has focused on the dissociation between perception and production (e.g. Brannen 2002; Hancin-Bhatt 1994b; Teasdale 1997), on phonological theories, universal factors, and language acquisition models (e.g. Flege and Davidian 1984; Picard 2002; Weinberger 1994; Westers et al. 2007). However, there does not appear to be a simple answer to the question why certain substitutes are chosen. While the phonological structure of the L1 certainly is an important factor in explaining different substitutions, other factors such as word-dependent characteristics or social factors and varying teaching curricula have probably an influence on L2 production as well.

Interestingly, even within one foreign accent different substitution choices are made, but these have been less systematically studied. Moreover, little is known about whether the substitutions are phonetically clear instances of [t,s,f] as they are often labelled. The purpose of our study was therefore to answer the following questions: a) what is the distribution of differential substitutions using comparable materials across L2 and L1 speakers, b) how do [t,s,f]-substitutes differ acoustically from the intended [t,s,f], and c) how do L2 productions of [θ] compare acoustically to L1 productions. Here we attempt a phonetic approach to examine language-specific preferences for [θ]-substitutions. A similar approach has been put forward by Teasdale (1997), who proposed that articulatory properties of [s] in the L1 are the best predictor of whether [t] or [s] would be chosen as the [θ]-substitute. Here we wish to elaborate on this idea by providing acoustic measurements of the substitutions as well as a comparison of these measurements between L1 and L2 speakers. We chose Dutch and German L2 learners of English, for which the acoustic properties of both [s] and [t] are different in their respective L1. Dutch [s] is less articulatorily tense and has graver friction than German (Mees and Collins 1982), and [t] in initial position is aspirated in German but unaspirated in Dutch (Keating 1984; Lisker and Abramson 1964). To this end, we collected a corpus of spoken English containing UK-English L1 speakers, as well as Dutch and German L2 speakers of English. The two groups of learners were selected because they not only differ in their predominant [θ]-substitutions (e.g. Westers et al. 2007 for Dutch; Hancin-Bhatt 1994b for German), but also in fine-acoustic details in fricative and stop production (Mees and Collins 1982; Rietveld and van Heuven 2001). The data obtained from the corpus were labelled and categorized. Moreover, acoustic measurements were taken to compare L1 and L2 [θ]-realizations, and to compare the L2 realizations of [θ]-substitutes [t,s,f] with intended [t,s,f]-realizations. In this study, only word-initial sounds are considered.

Studies on the acoustics of English fricatives have shown mainly four parameters that can distinguish fricatives: duration, spectral properties (e.g. centre of gravity, F2 onset, spectral peak location), amplitude (overall and relative noise amplitude), and transitions from the fricative into a vowel (e.g. Hughes and Halle 1956; Strevens 1960; Jassem 1962). While these measures can distinguish [s] from [f] and [θ], it seems that formant transitions and spectral peak location can provide additional information for less distinct fricatives such as [f] and [θ] (e.g. Harris 1958, Jongman et al. 2000, Tabain 1998). Which measure is best suitable can also depend on the use of real words versus syllables (e.g. Tabain 1998). The most informative cue for place of articulation of plosives is the distribution of energy in the release burst (e.g. Steven and Blumstein 1981), but other cues such as formant transition and spectral properties have also been reported (e.g. van Alphen and Smits 2004, for Dutch). In the present paper, we restrict the analysis of fricative intervals and plosive bursts to duration, center of gravity (COG), and amplitude.

2. METHOD

2.1. Materials and Procedure

A short story in English was constructed, containing numerous words with voiceless [θ] and words with [s], [f], and [t]. Participants were asked to read the story aloud at a comfortable speaking rate. Stereo recordings were made in a quiet room with a digital recorder at 44.1 kHz sampling rate with 16-bit resolution and were later transferred to a computer. The left channel was extracted for further processing.

For the analysis, we selected 18 content words with the voiceless [θ] in word-initial position (13 different words, occurring between 1 and 4 times in the story), and 10 [s]-, [f]-, and [t]-initial words each (in two cases for [t] and [f], the phoneme occurred in a stressed syllable-initial position within a word; altogether 27 different words, occurring 1 to 2 times). These words and their target phonemes were manually annotated. The spectrogram and the waveform were used to determine the onset and the cessation of the fricatives, and the onset and the offset of the burst. All [θ]-instances were then categorized by two trained research assistants (German learners' data by two assistants with L1 German, and Dutch learners' data by two assistants with L1 Dutch). Whenever there was disagreement about the category of a particular token, the categorization of a third coder (a trained phonetician) was decisive. English speakers' data were labelled and categorized by a trained phonetician and by one native speaker of English. In case of a disagreement, the opinion of the English native speaker was decisive. Before carrying out the acoustic analysis, all critical words were normalized for mean amplitude. Only [θ]-instances categorized as [t], [s], [f], or [θ] were measured, excluding few [θ]-instances that were either not fully produced, unclear, or substituted with other than the above mentioned substitutes. We used the PRAAT speech editor (Boersma 2001) to extract the duration, the amplitude, and the COG for each token. The weightening for COG was done by the absolute spectrum of the frequency (p=1). To calculate the average amplitude, the root-mean-squared (RMS) method was used, that is, the square-root of the mean of the squared amplitude of each point of a waveform.

2.2. Participants

The participants in the corpus study consisted of 37 native speakers of Dutch from the Radboud University in Nijmegen in the Netherlands (mean age 21.5, SD 2.2), 37 native speakers of German from the University of Cologne in Germany (mean age 22.5, SD 2.3; recordings of one participants were excluded due to technical problems), and 31 native speakers of English from the University Birmingham in England (mean age 19.4, SD 1.1). All participants took part in exchange for payment. The L2 participants were highly proficient in English. Dutch students had on average 7.6 years of formal English training, and German students had on average 8.8 years of formal English training. In an English multiple-choice vocabulary test (including many low frequency words), Dutch students scored on average 83% correctly, and German students scored on average 79% correctly (the difference in their scores did not reach significance). None of the German participants had lived in the Netherlands, and none of the Dutch participants had lived in Germany.

3. RESULTS

3.1. Categorization results

The categorization results across all items in Table 1 show how often [θ] was produced correctly or substituted with [s,t,f] or other phonemes (e.g. [tθ], [ts], [ʃ], or unclear), listed for each speaker group separately. The results show that all participants produced the English tokens with word-initial [θ] more often correctly than with a substitution, and that, unsurprisingly, L1 speakers substituted less frequently than L2 speakers. When comparing the two learner groups, German speakers produced significantly more words with substitutions than Dutch speakers did.

Table 1: Percentages of [θ]-productions per speaker group (percentages rounded up; numbers of occurrences are in brackets).

Speakers	[s]	[t]	[f]	[θ]	others
Dutch	5% (30)	23% (155)	3% (17)	62% (412)	7% (47)
German	29% (187)	7% (43)	5% (34)	51% (323)	8% (49)
English	0% (1)	0% (0)	12% (63)	88% (463)	1% (4)

Within the substituted instances, a significant difference between German and Dutch speakers was found: German learners predominantly substituted the English [θ] with [s] (71%, compared to 15% for Dutch speakers), while Dutch speakers predominantly substituted [θ] with [t] (77%, compared to 16% for German speakers). For both groups, the perceptually similar [f] occurred least frequently (13% for German speakers and 8% for Dutch speakers). It is worth noting that, overall, substitutions did not seem to be word-dependent, and that many participants produced more than one substitute type. For English speakers we found 12% of [f]-substitutions, which were mainly driven by three speakers. When excluding these speakers, the number of [f]-instances dropped to 5% and the number of [θ]-instances rose to 95%.

3.2. Measurements

The Figures below show the results from acoustic measurements across the three speaker groups for duration (Figure 1), RMS amplitude (Figure 2), and COG (Figure 3). To evaluate differences and similarities in the obtained values, t-tests were conducted across [θ]-realizations across the three speaker groups. Further t-tests within each of the speaker groups were aimed at a comparison between the accent-specific predominant substitutions ([t] for Dutch; [s] for Germans) and the realizations of the intended [t], [f], and [s] within and across L2 learners.

Figure 1: Duration in seconds (s) of the intended [t,s,f,θ], and of the [θ]-substitutions [t,s,f] (indicated by a following (th)).

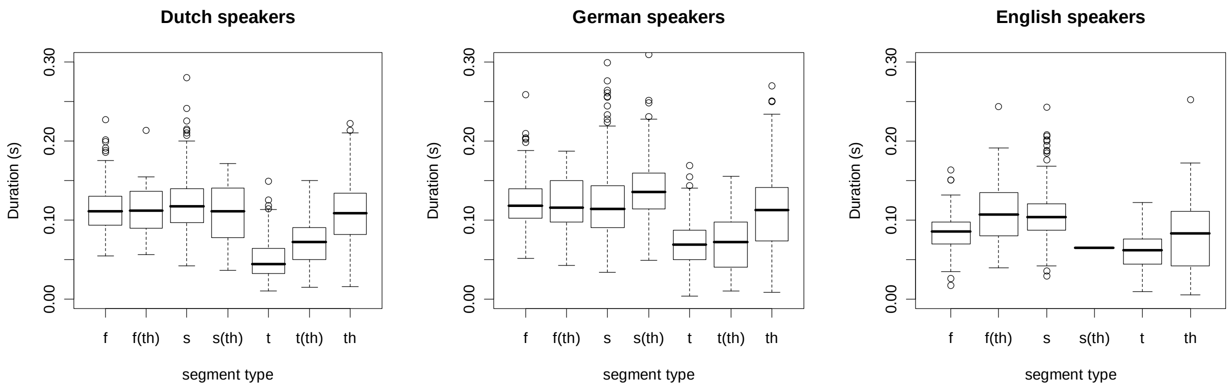


Figure 2: RMS in Pascal (Pa) of the intended [t,s,f,θ], and of the [θ]-substitutions [t,s,f] (indicated by a following (th)).

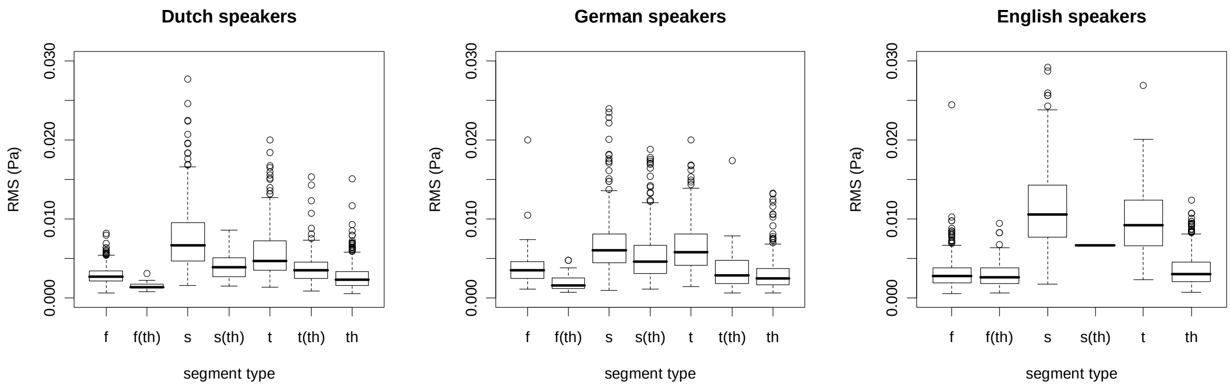
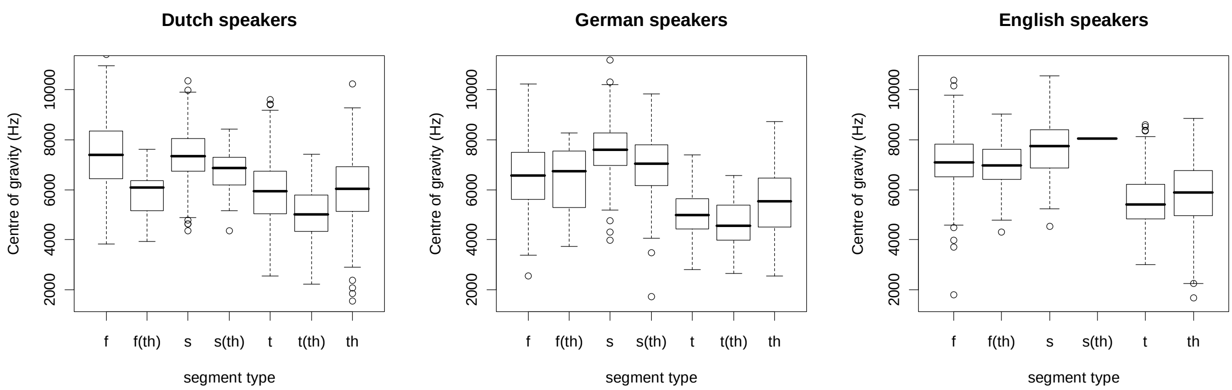


Figure 3: COG in Hertz (Hz) of the intended [t,s,f,θ], and of the [θ]-substitutions [t,s,f] (indicated by a following (th)).



The results showed that the [θ]-realizations of English L1 speakers differed from those of German L2 speakers in duration and RMS but not in COG. Dutch L2 speakers differed from English L1 speakers in duration but not in RMS or COG. Differences in duration between L1 and L2 speakers are not surprising, given that L2-speech rate is overall slower. Similarly, differences in amplitude could come about when L2 speakers encounter difficulties with a given speech sound and consequently lower their voice in amplitude. Importantly, the COG values did not differ across the groups, suggesting some evidence for target-like pronunciation of the English [θ] for L2 speakers. A comparison of the three measurements for [θ]-realization between German and Dutch speakers did not show significant differences (however, COG showed a weak tendency for a difference, $p = .084$). Further comparisons have shown that [s]-realizations did not differ between German and Dutch speakers, but the properties of [t] differed in all three measures. Given prior studies, we expected a difference between German and Dutch realizations of [s]. Because of a weak tendency for a difference in COG ($p = .097$), we further examined this issue by carrying out additional measurements that can help distinguish small differences in articulation (see Jongman et al. 2000). We found that the German [s] differed from the Dutch [s] in the kurtosis, standard deviation, skewness, and central moments.

A comparison of the intended [t,s,f] with the substitutes [t,s,f] was limited to the dominant substitutes within an L2 group (this was due to an insufficient number of responses for less frequent substitutes). Within the German group, [s]-substitutes differed from the intended [s]-realizations as well as from the correctly pronounced [θ]-instances in all measures. Similarly, Dutch speakers' [t]-substitutes differed from the intended [t]-realizations in all measures, and from [θ]-realizations in duration and COG, but not in RMS. This suggests that substitutions in L2 speech are on average not clear instances of the [t,s,f], as they are often labelled, and that they are neither clear instances of [θ].

4. DISCUSSION

The first question of this study concerned the distribution of substitution choices for the English voiceless interdental fricative [θ] by L2 and L1 speakers. The categorization results confirmed previous findings for L2 speakers: the dominant [θ]-substitute for German learners is clearly [s] while for Dutch learners it is [t] (e.g. Westers et al. 2007; Hancin-Bhatt 1994b). However, all three substitutions [t,s,f] occurred in the L2 productions of both learner groups, and the substitutes were not word- or speaker-specific. Importantly, L2 speakers produced native-like realizations of the fricative [θ] more often than any of the dominant substitutions. Since this probably depends strongly on the proficiency level of the L2 speakers, the numbers could reverse with lower proficiency. In contrast to L2 speakers, L1 speakers of English substituted [θ] (if at all) with [f]. It has been previously suggested that speakers of languages that articulate [s] further back and/or have a dental [t], are very likely to substitute the English interdental fricative [θ] with [t] (Taesdale 1997). This would indeed support Dutch preference for [t]-substitutes, because [s] is articulated further back in Dutch compared to German, which prefers [s]-substitutes. The present study further explored this proposal and found acoustic differences in L2-production of both [t] and [s] between German and Dutch speakers, supporting the phonetic explanation proposed by Taesdale (1997).

The second question concerned acoustic differences between [t,s,f]-substitutes and intended [t,s,f]. Given the nature of the natural elicitation method, we restricted the analysis only to dominant substitutes within an L2 group to ensure enough data points for a comparison. We found that not only did [t,s,f]-substitutes differ from the intended [t,s,f], they also differed from the [θ]-realization within each L2 group. This suggests that labeling conventions of [θ]-substitutions as [t,s,f] might not sufficiently characterize L2-productions, at least concerning its acoustics. Rather, [θ]-substitutions seem to show gradient properties, exhibiting acoustic properties that are often in between those of [θ]-realizations and [t,s,f]-realizations. However, perceptually these substitutes could still be perceived as good exemplars of [t,s,f]. To answer this question, results from a categorization experiment might be more telling, and we leave this issue to future studies.

The last question addressed an acoustic comparison of L2 and L1 [θ]-realizations. We found that German speakers did not differ from Dutch speakers, but differed from the English speakers in RMS and duration. Dutch speakers differed from the English speakers only in duration. Differences in amplitude and duration, however, are not surprising when comparing non-native with native speakers. Importantly, both L2 groups resembled the L1 group in the COG.

To conclude, this study showed that despite the difficulties that L2 speakers have with the English fricative [θ], more than half of the produced instances were target-like. Acoustically, the [θ]-substitutions were not clear instances of [t,s,f]. Articulatory differences between German and Dutch [t] and [s] were found and show a promising (phonetic) approach to future investigations of differential substitutions in L2 speech.

5. ACKNOWLEDGEMENTS

This work was supported by the Max-Planck-Gesellschaft (MPG). We would like to thank Laurence Bruggeman, Anne Blankenhorn, Sabrina Jung, Julia Lennertz, Simon Mack, Berit Meinert, Rachel Sheer, and Karina Visser for their assistance, and Frank Eisner for comments on an earlier version of the paper.

6. REFERENCES

Boersma, P. 2001. PRAAT, a system for doing phonetics by computer. *Glot Internatinal 5*. 341-345.

Brannen, K. 2002. The role of perception in differential substitution. *Canadian Journal of Linguistics – Revue Canadienne de Linguistique* 47. 1–20.

Cutler, A., Weber, A., Smits, R., Cooper, N. 2004. Patterns of English phoneme confusions by native and non-native listeners. *Journal of the Acoustical Society of America* 116(6). 3668-3678.

Flege, J.E., Davidian, R. 1984. Transfer and developmental processes in adult foreign language production. *Journal of Applied Psycholinguistic Research* 5. 323-347.

Hancin-Bhatt, B.J. 1994a. Segment transfer: a consequence of a dynamic system. *Second Language Research* 10(3). 241-269.

Hancin-Bhatt, B.J. 1994b. *Phonological Transfer in Second Language Perception and Production*. Doctoral Dissertation, University of Illinois.

Harris, K.S. 1958. Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech* 1. 1-7.

Hickey, R. (ed.). 2004. *A Sound Atlas of Irish English*. Berlin: Mouton de Gruyter.

Hughes, G.W., Halle, M. 1956. Spectral properties of fricative consonants. *Journal of the Acoustical Society of America* 28. 303–310.

Keating, P.A. 1984. Phonetic and phonological representation of stop consonant voicing. *Language 60*. 286-319.

Jassem, W. 1962. Noise spectra of Swedish, English, and Polish fricatives. *Proceedings of the Speech Communication Seminar, Stockholm, Royal Institute of Technology Speech Transmission Laboratory*. 1–4.

Jongman, A., Wayland, R., Wong, S. 2000. Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America* 108(3). 1252-1263.

Lisker, L., Abramson, A.S. 1964. A cross-language study of voicing in initial stops-acoustical measurements. *Word 20*. 384-422.

McGuire, G. 2003. The realization of interdental fricatives in Columbus, OH, AAVE. Paper presented at the *Montreal-Ottawa-Toronto Phonology Workshop*. Retrieved February 2009 from www.ling.ohio-state.edu/~mcguire/Interdentals]Handout.doc.

Mees, I., Collins, B. 1982. A phonetic description of the consonant system of Standard Dutch (ABN). *Journal of the International Phonetic Association* 12. 2-12.

Miller, G.A., Nicely, P.E. 1955. An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America* 27(2). 338-352.

Picard, M. 2002. The differential substitution of English /θ ð/ in French: The case against underspecification in L2 phonology. *Lingvisticæ Investigationes* 25(1). 87–96.

Rietveld, A.C.M., van Heuven, V.J. 2001. *Algemene Fonetiek*. Bussum: Coutinho.

Stevens, K., Blumstein, S. 1981. The search for invariant acoustic correlates of phonetic features. In Eimas, P.D. and Miller, J.L. (eds.), *Perspectives on the study of speech*. Hillsdale, NJ: Erlbaum. 1-38.

Strevens, P. 1960. Spectra of fricative noise in human speech. *Language and Speech* 3. 32–49.

Tabain, M. 1998. Non-sibilant fricatives in English: spectral information above 10 kHz. *Phonetica* 55. 107–130.

Teasdale, A.M. 1997. On the differential substitution of English [θ]: a phonetic approach. *Calgary Working Papers in Linguistics* 19. 71-85.

van Alphen, P., Smits, R. 2004. Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: the role of prevoicing. *Journal of Phonetics* 32. 455-491.

Weinberger, S.H. 1994. *Theoretical Foundations of Second Language Phonology*. Doctoral Dissertation, University of Washington.

Wells, J.C. 1982. *Accents of English 2: The British Isles*. Cambridge: Cambridge University Press.

Westers, F., Gilbers, D., Lowie, W. 2007. Substitution of dental fricatives in English by Dutch L2 speakers. *Language Sciences* 29. 477-491.

Few long-term effects of early immersion on L2 segmental timing in adulthood[1]

Tetsuo Harada

Waseda University
tharada@waseda.jp

ABSTRACT

This study investigated effects of childhood second language (L2) experience in additive bilingual education on L2 pronunciation in adulthood. Specifically, it acoustically compared the production of both voice onset time (VOT) and closure duration of singletons and geminates in Japanese by 9 English-speaking university students who had been enrolled in a Japanese immersion program in elementary school (early learners) to that of 9 English-speaking university learners of Japanese at the same proficiency level as the early learners with no exposure to Japanese in childhood (late learners). The closure duration ratio of geminates to singletons was also obtained. In addition, 8 native speakers of Japanese rated the degree of contrast between singletons and geminates produced by the L2 learners. The results show that although only the geminates produced by the early learners were rated more targetlike, they outperformed the late learners in neither VOT, the closure duration, the closure duration ratio nor the contrast rating of the singletons. Contrary to Knightly et al's (2003) hypothesis, the findings did not necessarily confirm a pronunciation advantage for the early learners over the late learners despite even a substantial amount of childhood input in the immersion program.

Keywords: childhood language memory, immersion education, early learners, VOT, closure duration

1. INTRODUCTION

Early exposure to L2 is reported to benefit L2 pronunciation in adulthood even if the amount of L2 input decreases dramatically after childhood (e.g. Knightly et al. 2003). Previous studies, however, focus on the acquisition of L2 sounds only in a naturalistic setting and do not give any clues to the effects of L2 input in an instructional setting. This study investigated the extent to which L2 input in a Japanese immersion program in an elementary school would benefit L2 pronunciation in adulthood despite a drastic decrease of Japanese input upon exiting the program.

Although early exposure to L2 has a greater impact than the length of exposure (e.g. Flege 1991), it remains unclear whether continued regular exposure is necessary to assist L2 pronunciation later in life. Childhood language memory may not be accessible if L2 input is discontinued early in life (e.g. Ventureyra et al. 2004). Other studies, however, suggest that childhood language memory is accessible in adulthood. For instance, Au et al. (2002) and Knightly et al. (2003) found that English-speaking adults who had childhood experience with Spanish produced voiceless stops (/p, t, k/) in Spanish with more nativelike VOT than late L2 learners. Furthermore, Oh et al. (2003) compared long-term benefits of childhood overhearing and speaking experiences. Both childhood speakers (i.e. people who had spoken Korean during childhood) and childhood hearers (i.e. people who had heard Korean during childhood but spoke only minimal Korean) outperformed typical L2 learners in phoneme perception, but only the childhood speakers excelled in the production of VOT over the late L2 learners. This finding suggests that childhood speaking experience may also be required to attain long-term benefits in L2 speech production.

Although no advantage in L2 pronunciation may arise from studying a foreign language in a traditional curriculum (e.g. a few hours a week) at an early age (Larson-Hall 2008), a substantial amount of L2 input provided in immersion education may be enough to trigger the formation of a phonological system. Harada (2007) found that although English-speaking children in an early total immersion program, in which many content subjects were taught in Japanese, produced Japanese voiceless stops with significantly longer VOT values than monolingual Japanese children, they distinguished phonetically between Japanese and English VOT regardless of places of articulation.

Whereas English has no phonemic distinction between single and geminate stops, the two types of stops in Japanese are contrastive as in *ita* ‘existed’ and *itta* ‘said.’ Harada (2006) analyzed the production of singletons and geminates by English-speaking children in the early total Japanese immersion program and found that their ratios of geminates to singletons were smaller than the native speakers’; however, the immersion children significantly differentiated between them.

When considering adults, it is unknown to what extent learners who were in a childhood immersion program could outperform typical late L2 learners at a university level. This current study compared the production of VOT and closure duration of singletons and geminates in Japanese by English-speaking university students who had been enrolled in a Japanese immersion program in childhood to that of a group of typical L2 learners at the university level. The study also tested whether or not the hypothesis (e.g. Au et al. 2002) that childhood language memory is still accessible in adulthood applies to the childhood exposure to L2 speech sounds in an immersion setting.

2. METHOD

2.1. Participants

For the production of VOT, 4 groups of paid participants consisted of 9 Japanese monolinguals (MJ), 9 American English monolinguals (ME), 9 English-speaking late learners of Japanese (LL) who were enrolled in a third or fourth year university-level Japanese course in the US, and 9 English-speaking early learners of Japanese (EL) who had been enrolled in a Japanese immersion program in elementary school in the US and whose amount of Japanese input dropped after they exited the program. For the production of single and geminate stops, the groups which participated were only MJ, LL and EL. The immersion program from which they graduated was early partial immersion, in which about 50% of the content courses were taught in Japanese from kindergarten or grade 1 to grade 5. In middle school (grades 6–8), 25% to 30% of the courses were instructed in Japanese, and in high school (grades 9–12), the Japanese input dramatically decreased to around 15% of instruction. In high school, some students enrolled in a regular Japanese course in a traditional curriculum. The early learners were exposed to Japanese only in the immersion class. Their teachers were either native speakers of Japanese or Japanese-English bilinguals who were born in a Japanese-speaking family in the US.

In addition, 8 native speakers of Japanese participated as paid informants to rate the contrast between the singletons and geminates produced by 13 early and late learners each (for the contrast rating, 4 additional learners were included in each group). They were undergraduate or graduate students in the School of Education at a private university in Tokyo and all spoke the Tokyo dialect except one rater, who was a speaker of the Kyushu dialect. Since his rating was highly correlated with the others’ (the average correlation coefficient with the other raters was .88, ranging from .84 to .91), it was included in the data set. Their ages raged from 23 to 40.

2.2. Procedures

The recording session consisted of a 20-minute face-to-face pronunciation elicitation test, and each participant was audiotaped while s/he interacted with the researcher in a quiet room. During each session the informant was shown pictures of objects which had been designed to elicit words including the target voiceless stop consonants. Each picture was presented on a Microsoft PowerPoint program.

The contrast rating was done in a computer lab, using an online questionnaire system with a sound file attached. The raters heard 384 target words in the carrier phrase (12 target words x 13 early learners + 12 target words x 13 late learners + 12 target words x 6 Japanese monolinguals) via headphones only one time. The raters were told that all the sentences had been spoken by either English-speaking learners of Japanese as a foreign language or native speakers of Japanese. An additional ten words were used for practice.

The raters were instructed to rate the degree of contrast between a singleton and a geminate while paying attention to the presence or absence of the geminate. They selected a rating on a 5-interval scale: 1 (clear singleton), 2 (ambiguous singleton), 3 (in between), 4 (ambiguous geminate), and 5 (clear geminate). The whole rating session lasted from 40 to 45 minutes.

2.3. Materials

The following criteria were taken into consideration in word selection: (a) the following vowel quality, (b) disyllabic words, and (c) the same accent pattern. Following a picture cue, the participants were asked to say a word, inserting it in the Japanese carrier phrase *sore wa ____ desu* 'That is ___' or in the English carrier phrase *I see a ___ in the picture*. The participants were asked to repeat each word in the VOT and geminate corpus three times, whereas they repeated the words in the singleton consonant corpus four times to balance out the number of the tokens in each category. Some of the words used in this experiment were *papa* 'papa,' *tako* 'octopus,' *kame* 'turtle' for Japanese VOT, *panda, taxi, candy* for English VOT, *happa* 'leaf,' *katta* 'bought,' *mikka* 'the third day' for Japanese geminates, and *papa* 'papa,' *kata* 'shoulder,' *saka* 'slope' for Japanese singletons.

In the contrast rating the six words just provided for the singleton and geminate category were used. The second and third repetitions of each word were selected, and each place of articulation for both types of stops was included. The rating session was divided into four parts with 10 blocks. Each block consisted of 10 randomized carrier phrases above containing one of the target words. The 10th block in each part had only six phrases. Each block began with a double beep generated by sinewaves and ended with a single beep. There were inter-trial intervals (ITTI) of 2 seconds, during which the raters evaluated the degree of contrast for each trial. They were instructed to take a brief break between the second and third blocks.

2.4. Data Measurement

The total number of tokens measured for the Japanese and English VOT was 729 each. The VOT of initial stops was measured to the nearest millisecond from the beginning of the release burst to the onset of voicing energy in F2 formants. Also, when the onset of F2 formants was not clear, VOT was measured from the beginning of the release burst to the first positive peak in the periodic portion of the waveform. A total of 6 mean VOT values for initial /p, t, k/ in English and Japanese, based on 9 observations each, were calculated for each learner of Japanese. For the monolinguals, a total of 3 mean VOT values were obtained.

For measurement of closure duration, the total number of tokens analyzed was 1377 tokens: 648 for single stops and 729 for geminate stops. The voiceless stops—[p], [pp], [t], [tt], [k], and [kk]—were identified by a gap on the spectrogram showing the stop closure. The beginning of the closure was defined by a cessation of the F2 of the preceding vowel. The end of the closure was defined by the burst of the following single or geminate consonant. When the burst was not visible, it was measured up to the beginning of frication. The waveform was used as secondary information. Mean closure duration values for /p, t, k/ and /pp, tt, kk/ for each participant were calculated based on 8 and 9 observations, respectively.

The 3 mean scores of the contrast rating for each speaker, one for each place of articulation, were obtained by adding the scores assigned by each rater and dividing them by the number of the raters. Also, to examine interrater reliability, Cronbach's coefficient alpha was calculated ($\alpha = .978$), which shows that the rating procedure was judged to be reliable across raters.

3. RESULTS

3.1. VOT

Figure 1 shows that both the early and late learners produced Japanese voiceless stops with longer mean VOT values than the Japanese monolinguals, while they tended to produce them with shorter VOT values than their English counterparts. The mean VOT values obtained for each participant were submitted to a Group/Language (6) and Place of Articulation (3) two-way repeated measures ANOVA, which yielded significant main effects but no interaction between them; group/language: $F(5, 48) = 21.850$, $p = .000$; place: $F(2, 96) = 95.425$, $p = .000$; Group/Language × Place: $F(10, 96) = 1.087$, $p = .380$. Post hoc Tukey tests revealed that the early and late learners' VOT values in Japanese were significantly different from those of the Japanese monolinguals (p = .000) regardless of the place of articulation. Furthermore, in the production of VOT in Japanese, the early learners did not outperform the late learners ($p = .679$).

Figure 1 The mean VOT values across the place of articulation for Japanese voiceless stops by the Japanese monolinguals (MJ), the early learners (ELJ), and the late learners (LLJ) and English voiceless stops by the early learners (ELE), the late learners (LLE), and the English monolinguals (ME). The error bars enclose ± one standard error.

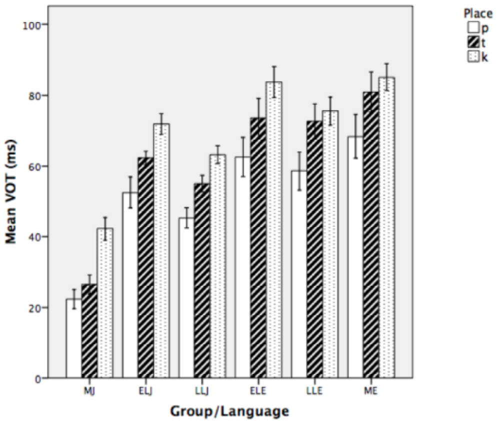
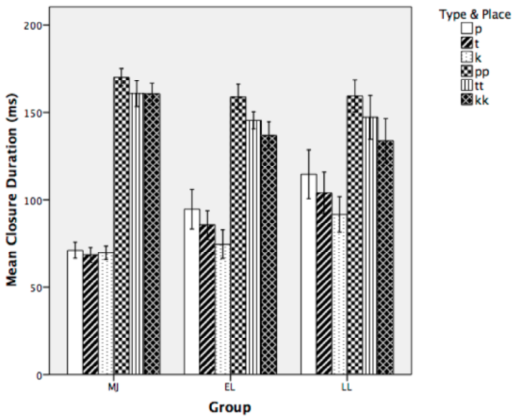


Figure 2 The mean closure duration of single and geminate stops across the place of articulation for the monolingual Japanese speakers (MJ), the early learners (EL), and the late learners (LL). The error bars enclose ± one standard error.



3.2. Closure Duration and the Ratio of Geminates to Singletons

Figure 2 illustrates that both the early and late learners produced singletons with longer duration values and geminates with shorter duration values than the monolinguals. The mean closure duration values for singletons and geminates were separately analyzed using a Group (3) × Place (3) two-way repeated measures ANOVA. The procedures yielded a significant main effect due to group and place for the singletons and due to place only for the geminates, and showed no significant interaction between group and place in either type of stops; Singletons: group: $F(2, 24) = 4.048$, $p = .031$; place: $F(2, 48) = 7.442$, $p = .002$; Group × Place: $F(4, 48) = 1.585$, $p = .194$ and Geminates: group: $F(2, 24) = 1.717$, $p = .201$; place: $F(2, 48) = 11.006$, $p = .000$; Group × Place: $F(4, 48) = .804$, $p = .528$. Post hoc Tukey tests revealed that although the late learners performed differently from the monolinguals in the production of singletons ($p = .024$), they did not differ from the early learners ($p = .280$), and that in the production of geminates both the early and late learners did not differ from the L1 speakers (EL: $p = .269$; LL: $p = .257$). This means that the early learners did not outperform the late learners in the production of both singletons and geminates.

Figure 3 shows that the MJ drew a contrast between the two types of stops with ratios of more than two times. However, the ratios for the early and late learners were smaller. A Group (3) × Place (3) two-way repeated measures ANOVA revealed a main effect due to group only but no interaction between group and place; group: $F(2, 24) = 3.739$, $p = .039$; place: $F(2, 48) = 3.104$, $p = .054$; Group × Place: $F(4, 48) = .761$, $p = .556$. Post hoc Tukey tests indicated that in the ratio of geminates to singletons the early learners did not differ from the L1 speakers ($p = .113$), but that there was no significant difference between the two groups of learners ($p = .882$). In summary, the early learners did not excel the late learners in the closure duration and the ratio of geminates to singletons.

Figure 3 The mean closure duration ratio of geminates to singletons across the place of articulation for the monolingual Japanese speakers (MJ), the early learners (EL), and the late learners (LL). The error bars enclose ± one standard error.

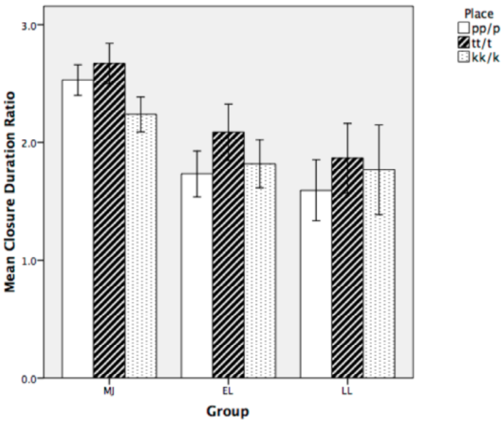
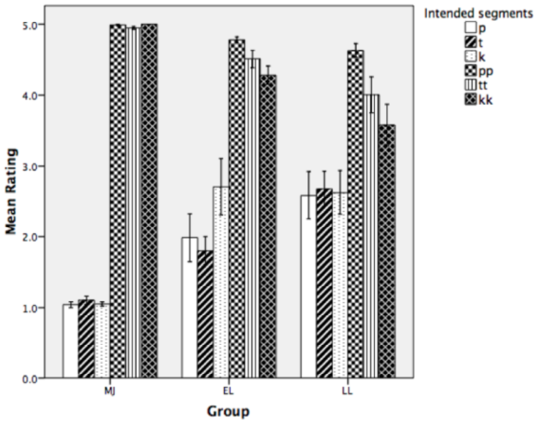


Figure 4 The mean rating of singletons and geminates across the place of articulation for the monolingual Japanese speakers (MJ), the early learners (EL), and the late learners (LL). The error bars enclose ± one standard error.



3.3. Contrast Rating

Figure 4 illustrates that the farther the singleton rating deviates from 1 and the geminate rating from 5, the less likely the intended segment is to be accurately identified as such. The mean ratings of singletons and geminates produced by the monolingual speakers were close to 1 and 5, respectively, which means that their tokens were accurately identified. The early learners were rated more highly than the late learners in the rating of both singletons and geminates (EL: singletons = 2.2, geminates = 4.5, LL: singletons = 2.6, geminates = 4.1). The mean ratings of the two types of stops for each speaker were separately submitted to a Group (3) × Place (3) two-way repeated measures ANOVA, which yielded a significant main effect due to group for the singletons and due to group and place for the geminates, and no interaction between group and place for either type of stops: Singletons: group: $F(2, 29) = 7.516$, $p = .002$; place: $F(2, 58) = 1.158$, $p = .321$; Group × Place: $F(4, 58) = 1.702$, $p = .162$ and Geminates: group: $F(2, 29) = 11.113$, $p = .000$; place: $F(2, 58) = 6.584$, $p = .003$; Group × Place: $F(4, 58) = 2.221$, $p = .078$. In the rating of singletons, both groups of learners significantly differed from the L1 speakers of Japanese (Tukey tests: EL, $p = .032$; LL, $p = .002$) and the early learners did not excel the late learners ($p = .474$). However, the geminates produced by the early learners were more accurately identified than those of the late learners ($p = .019$) and the former group's identification rate was not different from the Japanese monolinguals' ($p = .075$).

4. DISCUSSION AND CONCLUSION

The early learners outperformed the late learners only in the closure duration and contrast rating of geminates, but both groups did not differ in the production of VOT and the closure duration and rating of singletons. The findings suggest that even a substantial amount of input in immersion education during childhood provides only limited benefits for L2 pronunciation in adulthood. The long-term benefits of childhood L2 experience in a naturalistic setting hypothesized by Au et al. (2002) and their subsequent studies may not apply to childhood L2 exposure in an instructional setting as in immersion education. The benefit limited only to the production of geminates as opposed to singletons may be accounted for by one of the hypotheses in Flege's (1995) Speech Learning Model: the more dissimilar an L2 sound and the closest L1 sound are, the more likely one is to perceive the differences between the two sounds. In acquiring the contrast between singletons and geminates, English-speaking learners of Japanese must establish a new category of geminates, which they may find less challenging than to discern the slight difference between single stops in English and Japanese, and to fine-tune their already existing L1 category of /p, t, k/ for the development of the restructured L2 category of the singletons.

Then, what factors explain why a substantial amount of L2 input during childhood leads only to limited long-term benefits for L2 pronunciation in adulthood? As Piske et al. (2001) suggest, the amount of L2 use may be more crucial than the age of learning; thus the findings for this study can be attributed to a limited amount of input and output resulting from the traditional foreign language instruction they received in middle and high school.

Another possible factor is the amount of L2 input in an immersion program. The type of immersion education in which the participants were enrolled was a 50% partial immersion program, where 50% of instruction (i.e. about three hours a day) was conducted in Japanese from kindergarten through grade 5. In contrast, in a total immersion program, 100% of instruction is conducted in a foreign language in kindergarten and first grade, and around 80% in second and third grades (Johnson and Swain 1997). This implies that total immersion might provide better long-term effects on L2 pronunciation in adulthood.

An additional factor may be the distinction between childhood hearers and childhood speakers made by Oh et al. (2003), in which childhood speakers of Korean distinguished between three types of Korean stops (i.e. plain, tense, and aspirated), but childhood hearers did not. Children in immersion education may be similar to childhood hearers rather than childhood speakers. In immersion education, teachers always address their children in a target language, but children tend to converse with peers in English. L2 input may be limited to overhearing teacher-fronted discussions, experiencing basic interactions with peers (e.g. greetings, fixed expressions), and working in pairs for short periods. Tarone and Swain (1995) reported that peer-peer interactions in L2 decrease as children move into higher grade levels in immersion classes. This supports the assumption that immersion children may be similar to childhood hearers and explains why they did not outperform the late learners.

5. REFERENCES

Au, T. K., Knightly, L. M., Jun, S.-A., Oh, J. S. (2002). Overhearing a language during childhood. *Psychological Science, 13.* 238-43.

Flege, J. E. (1991). Age of learning affects the authenticity of voice onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America, 89.* 395–411.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research.* Timonium, Maryland: York Press. 233-277.

Harada, T. (2006). The acquisition of single and geminate stops by English-speaking children in a Japanese immersion program. *Studies in Second Language Acquisition, 28.* 601-32.

Harada, T. (2007). The production of voice onset time (VOT) by English-speaking children in a Japanese immersion program. *International Review of Applied Linguistics in Language Teaching (IRAL), 45.* 353-78.

Johnson, R. K., Swain, M. (1997). *Immersion education: International perspectives.* New York: Cambridge University Press.

Knightly, L. M., Jun, S.-A., Oh, J. S., Au, T. K. (2003). Production benefits of childhood overhearing. *The Journal of the Acoustical Society of America, 114.* 465-74.

Namai, K., Fukai, Y. (eds.). (2009). *Toward the fusion of language, culture and education: From the perspectives of international and interdisciplinary research.* Tokyo: Kaitakusha.

Larson-Hall, J. (2008). Weighing the benefits of studying a foreign language at a younger starting age in a minimal input situation. *Second Language Research, 24.* 35-63.

Oh, J. S., Jun, S.-A., Knightly, L. M., Au, T. K. (2003). Holding on to childhood language memory. *Cognition, 86.* B53-B64.

Piske, T., MacKay, I. R. A., Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics, 29.* 191-215.

Tarone, E., Swain, M. (1995). A sociolinguistic perspective on second language use in immersion classrooms. *The Modern Language Journal, 79.* 166-78.

Ventureyra, V. A. G., Pallier, C., Yoo, H.-Y. (2004). The loss of first language phonetic perception in adopted Koreans. *Journal of Neurolinguistics, 17.* 79-91.

NOTES

[1] This is an ongoing study supported by the Japan Society for the Promotion of Science (Grant-in-Aid for Scientific Research (C) 20520527). I thank Aya Kitagawa, a research assistant, for helping me analyze the data, and am grateful to the research participants for their time. Portions of this paper were presented at the 156th Meeting of the Acoustical Society of America, Miami, Florida, November 14, 2008 and the Second Language Research Forum (SLRF), Lansing, Michigan State University, November 1, 2009. This work is a revised version of the article published in Namai and Fukada (2009).

Choosing the optimal pitch accent location in Dutch by Chinese learners and native listeners

Xuliang He[1,2], Vincent J. van Heuven[3] and Carlos Gussenhoven[2]

[1]Nantong University, [2]Centre for Language and Communication, Radboud University Nijmegen
[3]Leiden University Centre for Linguistics
{x.he, c.gussenhoven}@let.ru.nl, v.j.j.p.van.heuven@hum.leidenuniv.nl

ABSTRACT

A perception experiment was conducted to study how well Chinese learners of Dutch identify the correct accentuation patterns in six categories of Dutch sentences. Thirty-six stimuli (6 sentences x 6 categories) were presented to 20 Dutch native listeners (NLD) and 20 Chinese learners of Dutch as a second language (CLD). In a forced-choice task, listeners had to decide which of two versions of each sentence was pronounced with optimal prosody, and to indicate how confident they were about their choice on a five-point scale. Per test item only one accent placement was prosodically optimal ('correct') as determined prior to the test by a panel of Dutch intonologists.

NLD correctness scores were significantly higher than those by the Chinese learners with high proficiency in Dutch (CLD-H), and the correctness scores of the latter were also significantly higher than those obtained for the low-proficiency learners (CLD-L). Along with the correctness scores, confidence ratings decreased from NLD to CLD-H to CLD-L. The results show that the different categories of accent placement do not present the same degree of difficulty to the Chinese learners.

Keywords: perception, accent, Dutch, Chinese learners of Dutch

1. INTRODUCTION

Speakers of non-native languages (L2 speakers) often have difficulties in producing acceptable stress and accentuation patterns, in part depending on the difference between L1 and L2. Incorrect prominence patterns often persist despite long exposure to the L2. It would appear that such 'stress deafness' is not merely the result of major stress typology differences (Gut, Trouvain and Barry 2007). One situation in which it has been observed is that of Chinese learners of English (Trouvain and Gut 2007). Germanic languages generally have sentence prosodic patterns that are the result of the inherent stress patterns of words and the rather complex rules for the placement of pitch accents on a subset of the stressed syllables. Variation in the distribution of pitch accents is strongly context-dependent, but even in reading tasks with isolated sentences, differences between native speakers and Chinese L2 speakers are striking (Chen Hua 2008). The question we attempt to address in this investigation is first, whether the apparent difficulties Chinese learners of Dutch (CLD) have with the correct placement of prominences in Dutch sentences is due to their inability to produce those patterns or rather in their ignorance of what an acceptable pronunciation of an isolated sentence is. The question, therefore, is whether they can recognize the correct location of sentence accents in a listening task. Additionally, given that the explanation for the presence or absence of a pitch accent on a given syllable may vary from morphology to information structure, our interest is in whether the acceptability of some accent placements are easier to establish by Chinese learners of Dutch than other accent placements. To this end, we have classified the accent placements according to the linguistic generalization that lies at their basis.

Accent is a place marker in the phonological structure where tones, known as pitch accents, are to be inserted (Goldsmith 1976, Hyman 1978). Pitch accents in a Dutch sentence are determined by lexical, phonological and morphological information, as well as by semantic and pragmatic factors (Gussenhoven ms). We chose six categories of accentuation problems in sentences for our participants to judge in the experiment.

The first category concerns primary word stress, which in Dutch falls on the antepenult, the penult, or the final syllable of the word if the penult is open, and on the penult or the final syllable if the penult is closed

(Gussenhoven 1999, Booij 1999). In the word *ooievaar* 'stork', for instance, in which the first and the last syllables are stressed, the primary stress falls on the antepenult *oo* and not on the last, *vaar*. Because final VVC syllables regularly take the primary stress, exceptional words like *ooievaar*, *Spanjaard* 'Spaniard', *olifant* 'elephant', may present difficulties to foreign learners, even though such words are both common and frequent. In the case of *ooievaar* (1a) is the correct reading.

(1) a. Er staat een OOievaar in de wei.
b. Er staat een ooieVAAR in de wei.
'There's a stork in the meadow.'

(2) a. Ze hebben een grote GROENTEtuin.
b. Ze hebben een grote groenteTUIN.
'They have a large kitchen garden.'

The second problem category comprises compound nouns. Of the two constituents in the compounds, the second loses its pitch accent. For example, the primary stress of *groentetuin* 'kitchen garden' is on the first component *groente* rather on the second *tuin*, which means that reading (2a) is correct.

The third problem is that of phrasal proper names. People's names generally consist of the given name, the surname and an optional surname prefix. The prefix is a function word (or multiple function words), which remains unstressed, and is considered a part of the surname. Some surnames, particularly those that were adopted as a result of the compulsory registration of the Dutch population during the French occupation of 1806-1813, are etymologically phrases, like *Vroegindewei*, which goes back to *Vroeg in de wei* 'early in the meadow'. The primary stress of a phrasal surname generally falls on the first syllable, but on the last if the same words are used as a common phrase. That is, (3a) is the appropriate reading.

(3) a. Dit is mevrouw VROEGindewei.
b. Dit is mevrouw vroegindeWEI.
'This is Mrs Vroegindewei.'

(4) a. Ze hebben een aantal nieuwe WERken aangekocht.
b. Ze hebben een aantal nieuwe WERken AANgekocht.
'They've acquired a number of new works of art.'

Gussenhoven (1983) made a distinction between eventive and non-eventive sentences, where eventive sentences involve the reporting of a change in the world. Following Schmerling (1976), the Sentence Accent Assignment Rule (SAAR) says that in eventive sentences, predicates lose their pitch accent if they are adjacent to one of their arguments. For instance, if a single accent falls upon *dogs* in the sentence *[DOGS must be carried]*$_{\text{EVENTIVE}}$, all relevant people will have to carry a dog. A non-eventive sentence would retain accents on both the argument *dogs* and the predicate *must be carried*, in which 'contingency' reading only those people who happen to have a dog are obligated to carry it (Ladd 2008). The fourth and the fifth categories are eventive and non-eventive sentences, respectively. In (4a), the argument of the verb *WERken* 'works' is accented and the verb *aangekocht* 'acquired' is deaccented. That is the reason why (4a), an eventive sentence, is correct. However, the (a) version is the correct reading in the case of (5), because *schade* 'damage' only exists potentially, so that the verb *MELden* 'report' is accented.

(5) a. U wordt verzocht eventuele SCHAde te MELden.
b. U wordt verzocht eventuele SCHAde te melden.
'You are requested to report any damage.'

Deaccenting for 'givenness' and accenting for 'newness' is illustrated in (6), where *jas* 'coat' is given information in the first sentence, and *zwarte* 'black' is the new information. That is, (6a) is correct, while (6b) is not.

(6) a. Ik heb wel een mooie bruine jas gezien. Maar ik zocht eigenlijk een ZWARTE jas.
b. Ik heb wel een mooie bruine jas gezien. Maar ik zocht eigenlijk een zwarte JAS.
'I did see a nice brown coat. But I was really looking for a black coat'

Mandarin Chinese has two types of accentuation: grammatical and logical (Gao 1984). Grammatical accents are determined by the structure of syntactic phrases and logical accents depend upon the meanings speakers intend to express. Chinese is a tone language, while Dutch is an intonation-only language. Inevitably, Chinese speakers of Dutch are influenced, to greater or lesser extent, by their L1 Chinese when they speak

Dutch. We know from the field of second language acquisition that the majority of second-language learners cannot acquire native-like oral ability, but less is known about the extent to which non-native speakers acquire the prosodic knowledge of the target language even in a situation in which they cannot produce it correctly. The answer to that question is important, because the cause of any mispronunciations will need to be attributed either to a lack of knowledge of the L1's prosodic structure or to an inability to pronounce such structures. The present perception experiment was conducted to study how well Chinese speakers of Dutch identify the correct accentuation pattern in the six categories of Dutch sentences. The first question we addressed is whether Chinese learners of Dutch are less often correct and less confident in their judgment of the appropriateness of accent patterns in Dutch than native Dutch listeners, and if their performance varies over the six accent placement categories. The second question is whether and to what extent Chinese speakers with higher overall proficiency in Dutch also do better on the accent judgment task than their counterparts with lower proficiency in Dutch.

2. METHOD

2.1. Materials

We obtained six categories of Dutch sentences with different accent patterns in each category (see section 1). Each category contains six sentences, so we get 36 sentences altogether in the corpus.

The recordings were made in a sound-treated booth at Radboud University Nijmegen. Each sentence was read with correct and incorrect accentuation by the third author or by a female Dutch phonologist. Two sets of stimuli included 72 [6 sentences x 2 (correct and incorrect) x 6 categories = 72] different stimuli with each sentence read with correct and incorrect accentuation by different speakers. The two sets had the same random order of sentences with the complementary stimuli, one mirrored the other. During the recordings, the third author checked to make sure the target words were read with proper accentuation. And we chose the best token of each pair of recordings as stimuli presented to the listeners.

2.2. Participants

A group of 20 native and 20 Chinese speakers of Dutch participated in the perception test. All Chinese participants (3 male, 17 female) were from the northern part of China, aged from 17 to 53. At the time of the experiment they had lived in the Netherlands for periods between three months and 22 years. The two subjects who had lived in the country for three months had studied Dutch in China for more than two years. This means that all Chinese participants had had sufficient exposure to the language to be able to do the experiment. The Dutch participants (4 male, 16 female) were self-declared native speakers of standard Dutch, aged between 18 to 54 years old. They were divided into two groups to do the forced-choice task. Ten Chinese subjects and ten Dutch subjects listened to stimuli in Set I and the other half in Set II.

The twenty Chinese subjects were asked to read a text of 42 Dutch sentences (our Production Experiment 3), and the Chinese group of speakers was divided into a higher ('Chinese Listeners of Dutch Higher': CLD-H) and a lower (CLD-L) subgroup on the basis of each subject's mean score over their segmental and prosodic proficiency scores judged by three experts. Pearson's correlations between the two scores between experts are $r_{pro1, pro2} = 0.97$, $r_{pro1, pro3} = 0.85$, $r_{pro2, pro3} = 0.88$; $r_{seg1, seg2} = 0.82$, $r_{seg1, seg3} = 0.76$, $r_{seg2, seg3} = 0.97$. This means three experts were highly constant to each other on the judgment of subjects' proficiency.

2.3. Procedure

Set I or Set II of stimuli was presented in individual sessions on a computer screen. Each participant was asked to listen to the stimuli and judge whether the reading was correct or not. The listener wore a GH632 headphone. Subjects first listened to 8 trial stimuli and then proceeded to the experiment proper. They could replay any stimulus before finalizing a judgment. The correct and incorrect tokens were randomized across categories. The listeners first clicked either of the two buttons marked 'correct' and 'incorrect', and then indicated his confidence in the judgment on a scale from 1 (poor confidence) to 5 (high confidence).

3. DATA ANALYSIS AND DISCUSSION

3.1. Analysis and discussion of correctness scores

We conducted a repeated measures Analysis of Variance (RM-ANOVA) using accent type (A: word stress, B: compound, C: proper names, D: eventive sentences, E: Non-eventive sentences, F: focus) as within-subjects variables and the different language groups as a between-subjects factor. The results show that correctness scores (Table 1) are significantly different ($F [5, 190] = 4.9$, $p < .05$) across accent types. Pairwise comparisons reveal that mean differences between types A and D, and between D and F are significant ($p = .014$, $p = .002$, respectively). There is no significant interaction between accent type and language group ($F [5, 190] = 1.4$, $p > .05$). The mean correctness scores (Table 2) for each stress type obtained by Chinese learners are significantly lower than those of the native Dutch group ($F [1, 38] = 39.4$, $p < .05$). This means that the native speakers of Dutch outperformed the Chinese speakers of Dutch. Patterns A and F are easier to identify, while D is the most difficult pattern.

Table 1: Mean correctness (Cor) and confidence (Conf) scores of Chinese (CSD) and Dutch subjects (NSD) broken down by accent type (A: word stress, B: compound, C: proper names, D: eventive sentences, E: non-eventive sentences, F: focus). For explanation of negative confidence scores see text.

Accent type	Language group	Mean		SD		N	Proficiency group	Mean		SD		N
		Cor	Conf	Cor	Conf			Cor	Conf	Cor	Conf	
A	CLD	3.95	1.45	1.60	2.17	20	CLD-L	3.20	.25	1.69	1.82	10
	NLD	5.80	4.38	.41	.50	20	CLD-H	4.70	2.65	1.16	1.84	10
B	CLD	4.10	1.50	1.25	1.81	20	CLD-L	3.70	.70	.95	.94	10
	NLD	5.15	3.38	.67	.96	20	CLD-H	4.50	2.30	1.43	2.14	10
C	CLD	3.90	1.38	1.29	1.79	20	CLD-L	3.10	.15	.99	1.03	10
	NLD	5.25	3.06	.79	1.04	20	CLD-H	4.70	2.62	1.06	1.52	10
D	CLD	3.55	.80	1.23	1.76	20	CLD-L	3.00	−.05	1.05	1.13	10
	NLD	4.75	2.73	.85	1.05	20	CLD-H	4.10	1.65	1.20	1.91	10
E	CLD	3.45	.71	1.23	1.74	20	CLD-L	2.90	−.18	1.29	1.53	10
	NLD	4.90	2.67	1.02	1.36	20	CLD-H	4.00	1.60	.94	1.51	10
F	CLD	4.45	1.98	1.05	1.59	20	CLD-L	4.20	1.32	1.03	1.17	10
	NLD	5.30	3.52	.80	1.11	20	CLD-H	4.70	2.63	1.06	1.73	10

Table 2: Overall mean correctness (Cor) and confidence (Conf) scores of Chinese and Dutch subjects

Groups	Mean		SD		N
	Cor	Conf	Cor	Conf	
Chinese	3.90	1.30	0.84	1.37	20
CLD-L	3.35	0.36	.55	.65	10
CLD-H	4.45	2.24	.72	1.27	10
Dutch	5.19	3.29	0.37	0.47	20

As Table 1 shows, native listeners of Dutch (NLD) obtained mean correctness scores above 5 points except for accent types D and E, whose mean scores are 4.8 and 4.9 respectively. The highest mean correctness score is 5.8 which goes to type A and the second highest (5.3) is type F. In the group of Chinese listeners of Dutch (CLD), the mean correctness scores are all below 5 points. The mean correctness scores of types B and F are comparatively higher (4.1 and 4.5, respectively), and the lowest is 3.5 which goes to type E. Both the CLD) and NLD groups have comparatively lower mean correctness scores for types D and E, and comparatively higher mean correctness scores for type F (the highest for CLD). The NLD group got the highest mean score for type A while this is not the highest for the CLD group. CLD got a comparatively higher mean score for Type B. This means that Chinese speakers of Dutch have problems with the identification of correct word stress. But their ability to judge the correctness stresses of compound words and focus information in sentences is higher than that of other types. Both groups have problems with the accentuation of enventive and non-eventive sentences.

We also conducted an RM-ANOVA using accent type as a within-subjects variable and different proficiency groups as between-subjects factor (for CLD listeners only). Correctness scores (Table 1) are significantly different across accent types ($F [5, 90] = 2.3$, $p = .05$). Pairwise comparisons reveal that accent types D and F differ significant ($p = .004$). There is no significant interaction between accent type and proficiency group ($F [5, 90] < 1$). The mean correctness scores (Table 2) for each stress type obtained by CLD-L are significantly lower than those of CLD-H ($F [1, 18] = 14.6$, $p < .05$). That means the CSD-H group outperformed the CLD-L group. Pattern F is easiest to identify, while Pattern D is difficult for Chinese listeners.

From Table 1, we also learn that CLD-H subjects got mean correctness scores above 4 points for all six accent types while CLD-L subjects obtained mean correctness scores around 3 points for types A, B, C, D and E while only accent type F got a significantly better score (4.2). Both CLD-L and CLD-H subjects got the highest mean correctness score for type F, and comparatively lower mean correctness scores for types D and E. Not surprisingly, CLD-L subjects have more problems with the correct identification of word stress. Both CLD-L and CLD-H have problems with the accentuation of eventive and non-eventive sentences (as do the native speakers of Dutch). All groups have significantly higher mean correctness scores for accent type F. Apparently, identifying the correct accentuation of sentences with focus information is relatively easy.

3.2. Analysis and discussion of confidence scores

For the analysis of confidence scores, each response was weighed positively if the accompanying judgment was correct. Otherwise, it was weighed negatively. We computed mean confidence scores per subject across accent patterns. We want to know how confident the subjects were when they made their judgments.

We again conducted an RM-ANOVA using accent types as within-subjects variables and different language groups as between-subjects factor. Confidence scores (Table 1) are significantly different across accent types ($F [5, 190] = 7.4$, $p < .05$). Pairwise comparisons reveal that the difference between accent types A and {C, D, E} are significant ($p < .05$), as is the difference between types {D, E} and F. There is no significant interaction between accent type and language group ($F [5, 190] =1.8$, $p > .05$). The mean confidence scores (Table 2) per stress type obtained by the Chinese subjects are significantly lower than those of the Dutch counterparts ($F [1, 38] = 37.4$, $p < .05$). This means that the native Dutch listeners were more confident than the Chinese listeners when they made their judgments. Both language groups were more confident when they judged types A and F and confident for types D and E.

As Table 1 shows, the mean confidence scores of Dutch group are all about 3 points except types D and E whose mean confidence scores are 2.7 and 2.7 (the lowest) respectively. NLD are most confident in their judgment of type A (4.4), followed by their judgment of type F (3.52). The mean confidence scores of CLD are above 1 point for all accent types except type D (0.8) and type E (0.7, the lowest).

An RM-ANOVA with accent type as a within-subjects variable and proficiency as a between-subjects factor show that confidence scores (Table 1) differ significantly across accent types ($F [5, 90] = 2.6$, $p < .05$). Only the difference between types F and D is significant (pairwise comparison, $p < .05$). There is no significant interaction between accent type and proficiency ($F [5, 90] < 1$). Mean confidence scores (Table 2) for each stress type obtained by the CLD-L are significantly lower than those of the CLD-H ($F [1, 18] = 17.4$, $p < .05$). This means that CLD-H were more confident in their judgments than CLD-L. Both language groups were more confident judging type F and less when judging type D.

Table 1 shows that CLD-H are most confident judging types A (2.7), C (2.6) and F (2.3), but less confident with types D (1.7) and E (1.6, lowest). Not surprisingly, CLD-L are not confident when judging type A (0.3). Like CLD-H subjects, they are not confident in their judgment of types D (−0.05) and E (−0.18).

Comparing mean correctness and mean confidence scores for the six accent types, we find that the two sets of scores correlate strongly. Nevertheless, for the Dutch group, the mean correctness score of type D is the lowest but the corresponding confidence score is not (the lowest confidence is for type E). This means though NLD got relatively higher mean correctness scores for type E than for type D, they are not sure whether their judgment was correct or not. Interestingly, for the Chinese listeners, whether CLD-L or CLD-H, the correctness scores and their confidence scores are in agreement with each other across all types of accentuation. It follows from the above comparisons, that all listeners took the experiment seriously and that their knowledge of Dutch accentuation patterns is truly reflected by their judgments.

4. CONCLUSIONS

Native speakers' correctness scores were significantly better than the scores by the high-proficiency Chinese speakers of Dutch, and the correctness scores of the latter were significantly higher than those of low proficiency Chinese learners of Dutch. This strongly suggests that greater proficiency in the language improves learners' ability to identify correct accent locations. Along with the correctness scores, confidence ratings decreased significantly from the native group to the high proficiency Chinese group and to the low proficiency Chinese group.

There were tantalizing differences between the linguistic categories, showing that some accent patterns are easier to judge − and thus learn − than others. Native speakers find it easiest to identify primary word stress (5.8 correct, where 6.0 is the highest possible score), but Chinese L2 speakers of Dutch find it easiest to judge the correctness of the focus condition (4.5). This reflects the fact that the location of primary and secondary word stress in the Dutch words we used is a language specific, arbitrary fact, whereas the focus structure of the sentences we used is given by the context. The Chinese learners thus derived the correct accentuation pattern in the focus condition from the pragmatics of the mini-dialogues that constituted the stimuli. Judging accentuation of eventive and non-eventive sentences is the most difficult for both native speakers (4.8) and for Chinese L2 speakers of Dutch (3.5). This is understandable, since both correct and incorrect accentuations in fact constitute natural accentuation patterns, given some adjustments in the context. Overall, with the exception of accentuation as a function of focus, our results show that the accentuation patterns of Dutch sentences are language specific and must be learnt.

ACKNOWLEDGEMENTS

This research was financially supported by the Netherlands Organisation for Research (NWO) as part of the programme 'Intonation in varieties of Dutch' (principal investigator Gussenhoven). The authors thank Judith Hanssen for permitting them to record her as the second speaker of our stimulus materials.

REFERENCES

Chen Hua 2008. *On Chinese EFL learner's English intonation patterns*. Shanghai: Shanghai Foreign Language Education Press.

Ellis, R. 1994. *The study of second language acquisition*. Oxford: Oxford University Press, 482-492

Gao Qiaoyang 1984. Grammatical Stress in Mandarin. *Language Teaching and Research*, 2, 65-76.

Booij, G. 1999. *The phonology of Dutch*. Oxford: Oxford University Press.

Goldsmith, J. 1976. Autosegmental tonology. Ph.D. Dissertation, M.I.T.

Gussenhoven, C. 1999. Illustrations of the IPA: Dutch. *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press. 74-77.

Gussenhoven, C. 1983. Focus, mode and the nucleus. *Journal of Linguistics* 19, 377-417.

Gussenhoven, C. (ms) Sentential prominence in English. To appear in Oostendorp, M.; Ewen, C.; Hume, B., Rice, K. (eds.), *Companion to Phonology*. Wiley-Blackwell.

Hyman, L.M. 1978. *Tone and/or accent*. In Napoli, D.J. (ed.), *Elements of tone, stress and intonation*. Washington: Georgetown University Press. 1-20.

Gut, U., Trouvain, J., Barry, W.J. 2007. Bridging research on phonetic descriptions with knowledge from teaching practice: The case of prosody in non-native speech. In Trouvain, J., Gut, U. (eds.), *Non-native prosody, phonetic description and teaching practice*. Berlin: Mouton de Gruyter. 3-25.

Ladd, D.R. 2008. *Intonational phonology*. Cambridge: Cambridge University Press.

Schmerling, S. 1976. *Aspects of English sentence stress*. Austin: University of Texas Press.

The contribution of accent distribution to foreign accentedness: causes and implications

Sam Hellmuth

University of York, UK
sh581@york.ac.uk

ABSTRACT

English and Egyptian Arabic (EA) display different patterns of pitch accent distribution. The distribution of accents in English can be formalised as phrase-level accent distribution, whereas in EA, accent distribution has been shown to be word-level accent distribution, with an accent occurring routinely on every content word, across a range of contexts and speech styles (Hellmuth 2006). In a production study, we test the hypothesis that the accent distribution pattern of EA will transfer into the L2 English of L1 EA speakers, then, in a perception study, we explore the potential effects of variation in the density of accent distribution on listeners' perceptions of L2-accented speech. The results indicate a mismatch between production and perception in L2 learner behaviour at the suprasegmental level, but apparently in the opposite direction to that observed at the segmental level (Sheldon and Strange 1982): advanced EA L2 learners' of English show evidence of L1 transfer in their production data, but pattern with English listeners in showing an affective interpretation of EA accent distribution.

Keywords: intonation, prosody, foreign-accentedness, production, perception.

1. INTRODUCTION

Egyptian Arabic (EA) displays a very different accent distribution pattern to that found in English. Although in English, in certain contexts, a speaker might realise an utterance with a pitch accent on every word, in most contexts, in naturally occurring speech, speakers realise pitch accents every few words at most. The distribution of accent in English is known to interact with information structure (Gussenhoven 1983, Selkirk 1984) and can be formalised in terms of phrase level accent distribution (Selkirk 2000). In contrast, Egyptian Arabic (EA) displays a default pitch accent on every content word, in a range of speech styles (Rifaat 1991, Hellmuth 2006), and can be formalised in terms of word level accent distribution (Hellmuth 2007). A similar rich accent distribution pattern has been described in other languages including Hindi (Harnsberger 1996, Patil et al. 2008) and Tamil (Keane 2006), and also, crucially, in varieties of English with a 'frequent accent' substrate such as Tamil English (Wiltshire & Harnsberger 2006).

This paper has two aims: i) to test the hypothesis that the rich accent distribution pattern of EA will transfer into the L2 English speech of advanced EA learners of English, and ii), to explore what effects this 'over-accentuation' might have on the interpretation of utterances by both L1 and L2 English listeners. Anecdotal evidence suggests that speakers of 'infrequent accent' languages like English may perceive speakers of 'frequent accent ' languages such as EA to be speaking in an angry or aggressive manner. A possible explanation of this anecdotal evidence is that the different accent distribution patterns in different languages map to a different function: in rich accent distribution languages, the pitch movement serves only to mark word-level prominence (Jun 2005, Hellmuth 2007), so that the mere occurrence of an accent provides no contribution to meaning; this contrasts strongly with the function of accent distribution in English which, as noted above, serves to mark argument/information structure. According to the Effort Code (Gussenhoven 2004:79), increased incidence of pitch movements will be interpreted paralinguistically as emphasis or insistence. If a speaker produces 'unnecessary' extra accents in their L2 English, due to transfer of a linguistic accent distribution pattern from their L1, there is a risk that this will be interpreted paralinguistically instead of linguistically.

We address these twin research questions by means of a production study (in section 2) and a perception study (in section 3).

2. PRODUCTION STUDY: DOES L1 ACCENT DISTRIBUTION TRANSFER TO L2?

Experiment 1 documents accent distribution patterns in the L2 English speech of two female speakers of EA, who recorded IViE corpus stimuli, with comparison to the accent distribution pattern observed in parallel utterances in the speech of two female L1 speakers of Southern British English from the IViE corpus.

2.1. Method

Two female speakers of EA who are advanced L2 learners of English recorded the IViE corpus stimuli (http://www.phon.ox.ac.uk/IViE). We present here an analysis of accent distribution in the read speech sentences and map task only. Both speakers were postgraduate students at the University of York, aged in their early thirties, who were born in Cairo and had lived there continuously until they moved to the UK. They had been in the UK for 15 and 18 months respectively and neither had previously lived in an English-speaking country. Both had studied English at school in Egypt since the age of approximately 4 years and have a level in English of IELTS 6.0 or higher. The speakers differ in that one (FD) attended English medium schools but the other (FR) attended Arabic medium schools. Recordings were made in a soundproof recording studio using B+K 4001 condenser microphones directly to digital format at 16bit 44.1Khz. The sound files were resampled to 16 bit 22.05KHz prior to analysis to reduce processing time in Praat.

For comparison, two female speakers from Cambridge were selected from the IViE corpus (speakers f1 and f3) and their read speech sentences and map task dialogue were analysed in directly parallel fashion to the L2 speakers' data. For the L2 speech data, prosodic transcriptions were made using IViE notation by the author and by a second trained transcriber and the final transcription of the data relied on both transcribers' annotations (intertranscriber agreement was approximately 84%). For the L1 speakers' data the author transcribed the data in parallel fashion then made comparison with published IViE transcriptions to arrive at a final transcription (agreement with the published data was approximately 86%). Since the dependent variable of interest in the present study is accent density, if the two input transcriptions differed over presence/absence of an accent, the transcription without an accent was maintained, so that the accent count represents the most conservative analysis of the data.

The number of function words and content words was counted in each portion of the data: the 44 read speech sentences comprise 444 words in total (260 function words, 184 content words); the L2 speakers' map task dialogue contained 333 words (158 function words, 175 content words); the L1 speakers' map task dialogue contained 371 words (162 function words, 207 content words). Finally, the proportion of words realised with a pitch accent, by L1 and L2 speakers respectively, was calculated.

2.2. Results

The results of the transcription study are shown in Figure 1 below. We observe a higher incidence of accents in the EA speakers' L2 English than in that of the two L1 speakers. The difference in accent density between L2 and L1 speakers is somewhat more pronounced in the map task dataset than in the read speech dataset. The difference in mean values of accent density between the two groups is significant for content words, in both datasets (read speech: $t=3.484$; $df= 75.9$; $p=0.001$; map task: $t=4.905$; $df=79.4$; $p<0.001$), but not for function words (read speech: $t=0.994$; $df= 86$; $p=0.323$; map task: $t=0.692$; $df=84$; $p= 0.491$). A pair of sample read speech utterances, illustrating the denser distribution of accents in L2 speech than in L1 speech, is provided in Figures 2 and 3 below (which also illustrate the transcription labelling scheme).

2.3. Summary

The production study results suggest that the accent distribution pattern observed in EA, whereby a pitch accent is realised on every content word, does transfer into the speech of even advanced L2 EA learners of English. In section 3 below we explore whether this transfer creates potential for misinterpretation of EA speakers' L2 English.

Figure 1: Percentage of words realised with an accent, by word type and by speaker group (L1 vs. L2), in read speech sentences (left) and in map task dialogues (right).

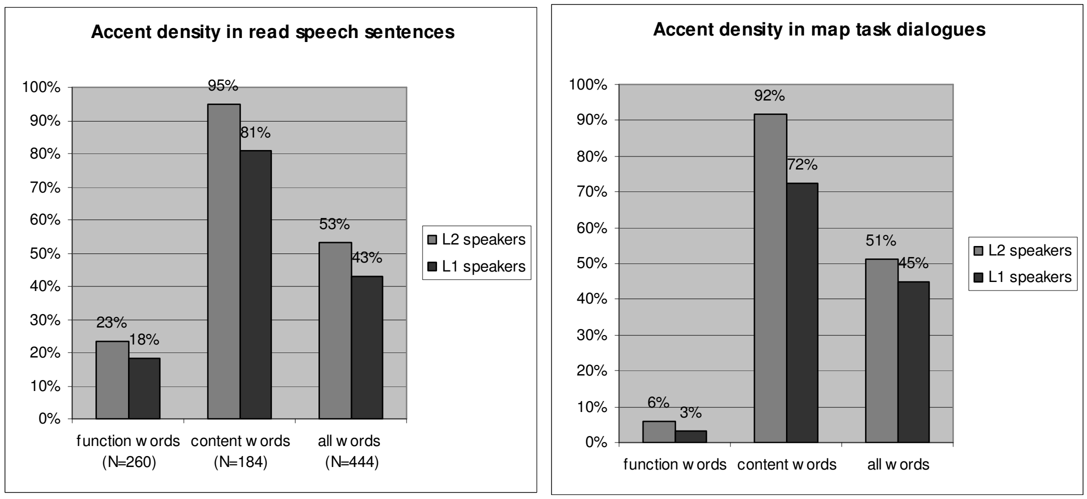


Figure 2: Sample pitch trace showing accentuation of all content words in a yes-no question by an L2 speaker (speaker FD).

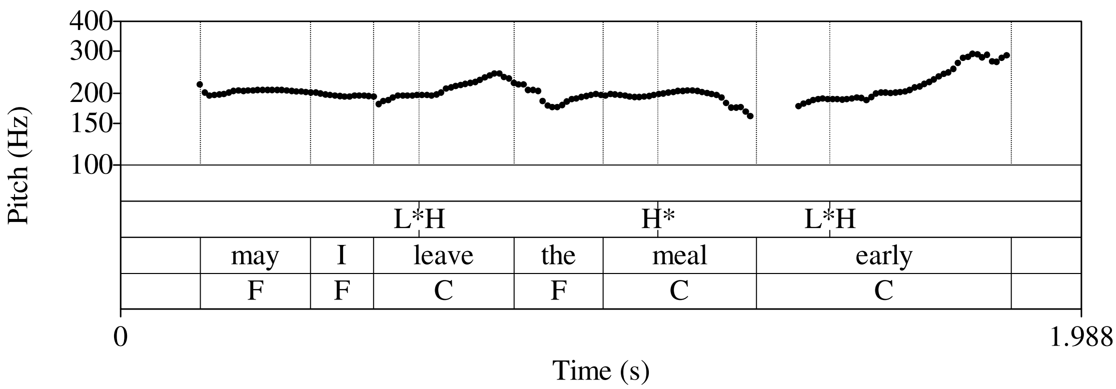
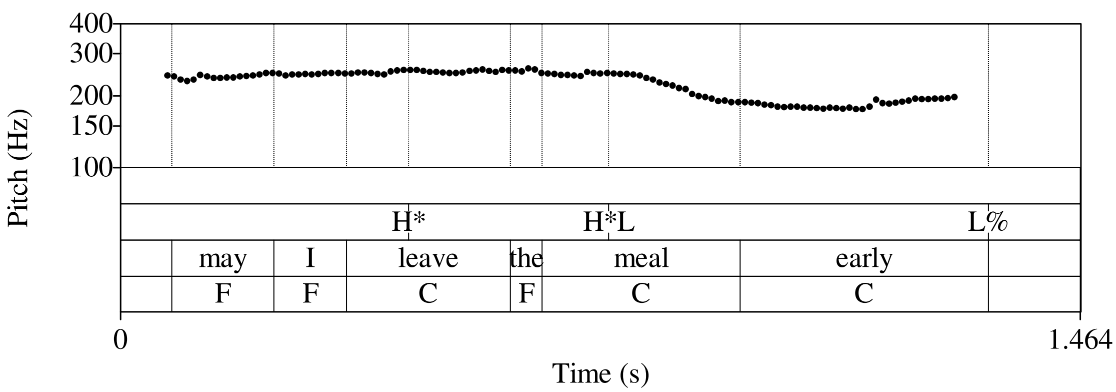


Figure 3: Sample pitch trace showing realisation of a yes-no question by an L1 speaker from the IViE corpus (speaker f1).



3. PERCEPTION STUDY: DOES L1 TRANSFER OF ACCENT DISTRIBUTION MATTER?

Experiment 2 explores how increased density of accent distribution, in EA and in English, is interpreted by listeners. As set out in section 1 above, anecdotal evidence suggests that speakers of phrase-level accent distribution languages, like English, may perceive speakers of word-level accent distribution languages, such as EA, to be speaking in an angry or aggressive manner. Gussenhoven (2004:71ff.) proposes a set of biological codes which influence both linguistic and paralinguistic interpretation of tonal events: the Frequency Code affects overall pitch range such that higher pitch is interpreted as being produced by a smaller speaker, and the Production Code affects declination, such that high pitch will be interpreted as utterance-initial, and low pitch as utterance-final. The Code that is relevant for our present purposes is the Effort Code which affects f0 excursion and what Gussenhoven terms the "incidence of movements" (p79). According to Gussenhoven, greater effort will result in either greater f0 excursion or more pitch movements. The informational interpretation (signalling some attribute of the message itself), of either of these instantiations, will be as more or less 'emphatic', whilst the affective interpretation (signalling some attribute of the speaker) will be 'insistent' or 'surprised'.

Since both English and Arabic use f0 excursion to signal focus (e.g. Ladd & Morton 1997, Hellmuth 2006) we expect that the two sets of listeners will give similar 'emphasis' ratings to utterances with differing accent density (in both English and EA). In contrast, for 'insistence', our hypothesis is that ratings will vary across the two sets of listeners: we expect density of accent distribution to have no affective value for EA listeners, since every content word is routinely accented in EA, whereas we expect English listeners to perceive utterances with a greater incidence of pitch accents as more insistent.

3.1. Method

We manipulated the incidence of pitch movements in two sample utterances, in EA and English, and they were rated separately for degree of emphasis and degree of insistence by 4 English listeners and 4 EA listeners (with advanced L2 English). The two target sentences were in each case the opening sentence in a read narrative: the Cindarella story from the IViE corpus and recordings of "Guha and the banana seller" (Abdel Massih 1975) from Hellmuth 2006. The incidence of pitch movements in each utterance was systematically varied by manipulating the f0 contour to increase or decrease the number of accents, as required, using the PSOLA resynthesis function in Praat.

For the EA stimuli, the base stimulus contained 10 content words and was realised with 10 pitch accents; the number of accents was systematically reduced in five additional steps (10 > 8, 7, 6, 4, 3). For the English stimuli, the base stimulus also contained 10 content words, but was realised in the original with 6 pitch accents; the number of accents was increased/decreased in five additional steps (10, 8, 7 < 6 > 4, 3). In each language the range was from 10 to 3 pitch accents per utterance. The base stimulus and a sample manipulated accent distribution contour for each language are illustrated in Figures 4-5 below. Four L1 English listeners and four L1 EA listeners rated each of the six stimuli in each language twice each on two scales: firstly, 'emphatic↔non-emphatic', then, in a separate task, on the scale 'insistent↔non-insistent'.

3.2. Results

The results of the ratings for degree of emphasis and degree of insistence, from English and EA listeners are shown in Figure 6 below. For 'emphasis', as predicted, there was no apparent effect of accent density on ratings, neither for English not EA listeners; there is however an overall effect of stimulus language, in that both sets of listeners rate utterances not in their own language as more emphatic than those in their own language, and the difference is greater for the Arabic listeners. For 'insistence', a similar effect of language is found (listeners rate utterances in their own language as less insistent than those not in their own language) but as predicted, there is an effect of accent density on the degree of 'insistence' perceived: English listeners interpret increased accent density in Arabic utterances as more insistent (and there is a similar though weaker effect for English utterances). Contrary to expectations however, the EA listeners show a similar trend, rating Arabic utterances as more insistent when they contain a greater number of accents.

Figure 4: Pitch trace showing English base stimulus (EN4, black) with fully accented stimulus (EN1, grey) for comparison.

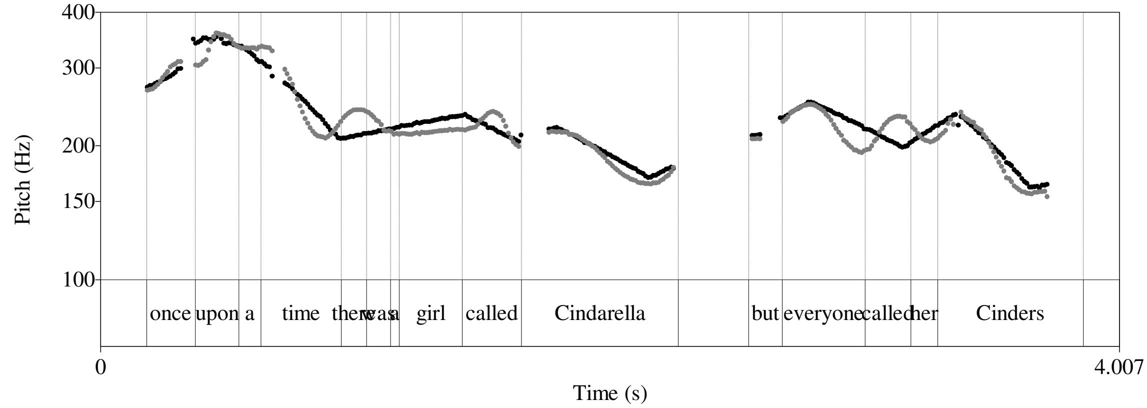


Figure 5: Pitch trace showing EA base stimulus (EG1, black) with de-accented stimulus (EG6, grey) for comparison.

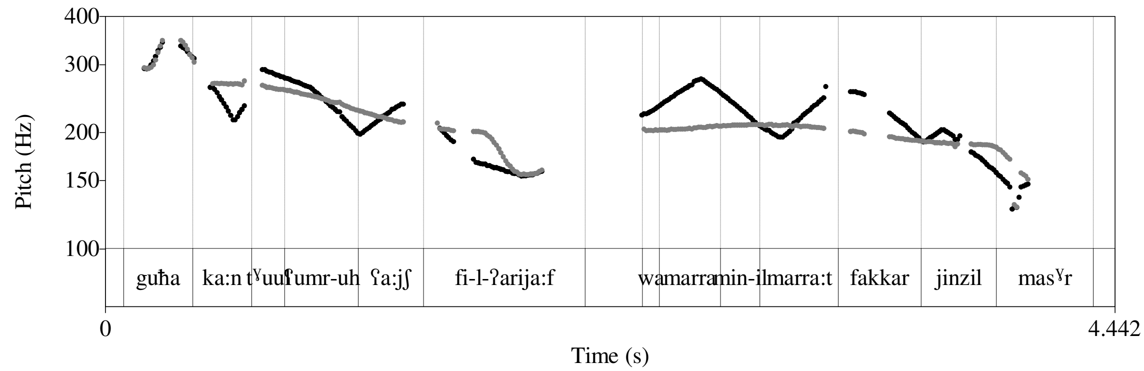
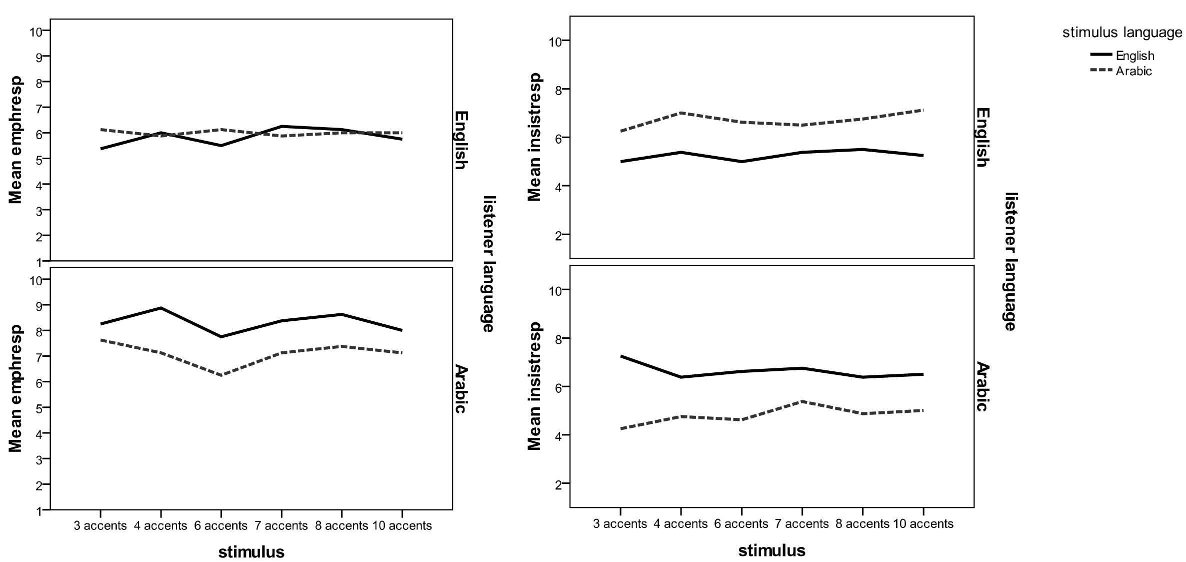


Figure 6: English and EA listeners' ratings for emphasis (left) and insistence (right) of English (solid)/EA (dotted) stimuli.



3.3. Summary

As predicted, the informational value of accent distribution, as emphasis, is similar for both English and EA listeners, which we ascribe to use of increased f0 excursion (an alternative instantiation of the Effort Code) in both languages for marking of contrastive focus. The affective value of accent distribution varies however, such that English listeners interpret utterances with a greater number of accents as more insistent than parallel utterances with fewer accents. Contrary to expectations, the EA listeners responded similarly, also rating greater accent distribution in EA utterances as more insistent. Since all our EA listeners were advanced learners of L2 English, resident in the UK, we interpret this as evidence that L2 learners may develop an awareness of the basic difference in accent distribution in the two languages, and an affective interpretation of accent density becomes available to them.

4. CONCLUSION

Prior research has established that accent distribution is a parameter of cross-linguistic prosodic variation (Hellmuth 2007) and is thus expected to be a candidate for potential L1 transfer effects into L2 learners' interlanguage. We have demonstrated here that accent distribution does transfer into the L2 English of advanced EA learners of English, and in addition have shown that density of accent distribution may be interpreted affectively by English listeners as increased insistence, due to differences in the instantiations of the Effort Code between EA and English. Nonetheless, we find that although L2 learners produce a greater density of accents, they appear to be aware in perception that dense accents are not English-sounding (and rate them as insistent). A limitation of the perception study is that the stimuli were manipulated in only one acoustic correlate of prominence (f0) so that listeners may be picking up on other correlates which remained unchanged (duration, intensity, vowel quality and articulatory force). In future work we hope to examine the effects on both affective and informational interpretation of variation in metrical prominence (marked by non-tonal and tonal cues) and accent distribution per se (marked by tonal cues only).

5. ACKNOWLEDGEMENTS

Thanks are due to the speakers of Egyptian Arabic who participated in the production and perception studies, to the English listeners who participated in the perception study, and to Kyle Day who acted as second transcriber for the EA production study data; we are grateful to the authors of the IViE corpus for use of excerpts of the corpus in both studies. The author wishes to thank audiences at the University of Edinburgh and the University of Newcastle for comments on earlier versions of this paper.

6. REFERENCES

Abdel-Massih, Ernest T. 1975. *An introduction to Egyptian Arabic*. Ann Arbor: University of Michigan.

Gussenhoven, C. 2004. *The phonology of tone and intonation.* Cambridge: CUP.

Harnsberger, J. 1996. Towards an intonational phonology of Hindi. (*5th Conference on Laboratory Phonology* July 7th 1996).

Hellmuth, Sam. 2006. *Intonational pitch accent distribution in Egyptian Arabic*. Unpublished PhD thesis, SOAS.

Hellmuth, Sam. 2007. The relationship between prosodic structure and pitch accent distribution: evidence from Egyptian Arabic. *The Linguistic Review*. 24(2). 289-314.

Jun, Sun-Ah. 2005. Prosodic typology. In Jun, Sun-Ah (ed.) *Prosodic Typology: The Phonology of Intonation and Phrasing*. 430-458. Oxford: OUP.

Keane, E. 2006. Prominence in Tamil. *Journal of the International Phonetic Association*. 36 1-20.

Ladd, D. R., R. Morton. 1997. The perception of intonational emphasis: continuous or categorical? *Journal of Phonetics*. 25(3). 313-342.

Patil, Umesh, Gerrit Kentner, Anja Gollrad, Frank Kügler, Caroline Féry, Shravan Vasishth. 2008. Focus, word order and intonation in Hindi. *Journal of South Asian Linguistics*. 1(1). 55-72.

Rifaat, Khalid. 1991. *The intonation of Arabic: an experimental study*. Unpublished PhD thesis, University of Alexandria.

Selkirk, Elisabeth O. 1984. *Phonology and syntax: the relation between sound and structure*. Cambridge, Mass.: MIT Press.

Selkirk, Elisabeth O. 2000. The interaction of constraints on prosodic phrasing. In Horne, Merle (ed.) *Prosody: theory and experiment*. 231-262. Dordrecht: Kluwer.

Sheldon, A., Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: evidence that speech production can precede speech perception. *Applied Psycholinguistics* 3, 243–61.

Wiltshire, Caroline, J. Harnsberger. 2006. The influence of Gujarati and Tamil L1s on Indian English: a preliminary study. *World Englishes*. 25(1). 91-1004.

Native Korean Speakers' Perception on Arabic Pharyngeal Contrast

Seung-ah Hong and Priyankoo Sarmah

Hankuk University of Foreign Studies
hongseungah@hufs.ac.kr and pri@hufs.ac.kr

ABSTRACT

The present study examined the native Korean speakers' perception of Arabic pharyngealized and non-pharyngealized consonant contrasts. The goal is to present how L1 speakers in different learning stages would accommodate L2 contrasts absent in their L1 phonology. A perception experiment was conducted on the participants in three different groups of Korean speakers differing in their familiarity with the Arabic phonemes. In an ISI of 1200ms the most experienced learners of Arabic showed worse performance in discriminating Arabic pharyngealized vs. non-pharyngealized contrasts. It was also observed that all participants across the groups have propensities of perceiving the aforementioned contrast in two distinct ways, while one subgroup had problems discriminating the /d / vs. /dˁ/ contrasts the other had problems with the /s/ vs. /sˁ/ contrasts.

Keywords: Arabic, Pharyngealization, Korean, L2, Perception

1. INTRODUCTION

Inability to discriminate among closely related foreign language sounds by native speakers of a particular language is not an unusual phenomenon. It has been widely reported that adults have difficulty in perceiving novel contrasts that are not available in their native phonology (Lisker and Abramson 1970, Werker and Tees 1984 etc.). On the contrary it has also been reported that infants in the first year of their life are able to discriminate between closely related speech sound contrasts (Werker et al. 1981, Werker and Tees 1984 etc.). However, the more a human is exposed to their L1; their ability to detect novel phonetic contrasts significantly decreases. The question that has been lingering is do adults completely lose the sensitivity for novel phonetic contrasts? Werker and Tees (1984) argues that it is rather the change in the focus of attention due to the altering of hearing strategies rather than losing the ability itself. Therefore, there might be a possibility of regaining the sensitivity by certain process.

Our goal in this paper is to see how adult native speakers of the Korean language perceive and discriminate between the pharyngealized and non-pharyngealized contrasts of the Arabic phonemes. Keeping this goal in mind we conducted a series of perception tests on three groups of Korean speakers- novice Korean speakers who have no experience with the Arabic language (G1), Korean learners of Arabic who have studied the language for less than a year (G2) and Korean learners of Arabic who have more than four years of experience with the language (G3). The results of the perception tests were compared among the three groups and reported in the sections to follow. The following sections outline the methodology, results and implications derived from the current study.

2. METHODOLOGY

2.1. Stimuli

A male native Arabic speaker from Kuwait was recorded producing the pharyngealized and non-pharyngealized contrasts. The speaker was asked to pronounce the each stimulus in a C/i/ structure in Modern Standard Arabic (MSA). The C comprised of four different types of Arabic pharyngealized consonants and their non-pharyngealized counterparts: /t, tˁ, d, dˁ, ð, ðˁ, s, sˁ/. Al-Masri & Jongman (2003) reported that the pharyngealized consonants in Arabic lower the F2 of the adjacent vowels. It is also reported that the degree of F2 lowering was less in [i] or [u] than [a] vowel. In this study, [i] has been chosen over [u] since it was noticed in our study that in the context of pharyngealized consonants [u] tended to give rise to an allophonic variant [o]. As [u] and [o] might have been considered to be two separate phonemes by the native Korean participants, the decision was taken in favor of the C/i/ structures. The vowels in the stimuli had been normalized for their duration (within each pair) and for intensity (at 75dB).

2.2. Participants

Thirty one (31) native speakers of Korean participated in this study (13 male and 18 female). Their ages ranged from 19 to 35 years. They were classified into three groups according to their proficiency in Arabic. The first group (G1) consisted of native Korean speakers with no knowledge and exposure to the Arabic language. The G1 group consisted of (10) participants with a mean age of 25.5 (*SD*= 4.4). The G2 group comprised of eleven undergraduate students with a mean age of 19.9 (*SD* = 0.3) who have taken Arabic lessons for less than a year. The participants in the G3 group consisted of ten advanced learners of Arabic with a mean age of 26.1 (*SD*= 2.1). The participants in G3 have had a minimum of four year university level study in Arabic with some of them having lived in Arabic speaking countries. None of the participants reported any history of hearing or speech disorders and they reported to be in normal physical condition at the time of the test. They were compensated with a gift voucher of five thousand Won (USD 4.5) for their participation.

2.3. Procedure

An ABX discrimination test was conducted on the participants of this study where the participants heard a block of three stimuli that were organized as described in Table 1. This type of 2AFC task is highly acknowledged for its reliability in works on L2 perception (Best et al. 1988, Best and Strange 1992 etc.).

Table 1: Sets of stimuli used in the current study

Blocks	Stimuli Type	Stimuli Sets
I	AAB	/t-t-tˤ/, /ð-ð-ðˤ/, /d-d-dˤ/, /s-s-sˤ/
II	BAA	/tˤ-t-t/, /ðˤ-ð-ð/, /dˤ-d-d/, /sˤ-s-s/
III	BBA	/tˤ-tˤ-t/, /ðˤ-ðˤ-ð/, /dˤ-dˤ-d/, /sˤ-sˤ-s/
IV	ABB	/t-tˤ-tˤ/, /ð-ðˤ-ðˤ/, /d-dˤ-dˤ/, /s-sˤ-sˤ/

Each stimuli set was repeated six times resulting in a total of 96 trials (4 contrastive pairs x 4 trial types x 6 repetitions) given to the participants with an inter-stimulus-interval (ISI) of 1200 milliseconds. Each set was randomly presented on a laptop computer using Praat (Boersma and Weenink 2009) MFC interface. The participants listened to the sounds through a pair of headphones and once a participant finished listening to a set of stimuli, she was asked to identify the distinct phoneme by clicking a relevant option on the computer screen. Each participant took about twenty minutes to complete the experiment. Their responses were collected onto a spreadsheet and subjected to further analyses. For statistical analysis, ANOVA and Bonferroni post-hoc tests were conducted on the responses.

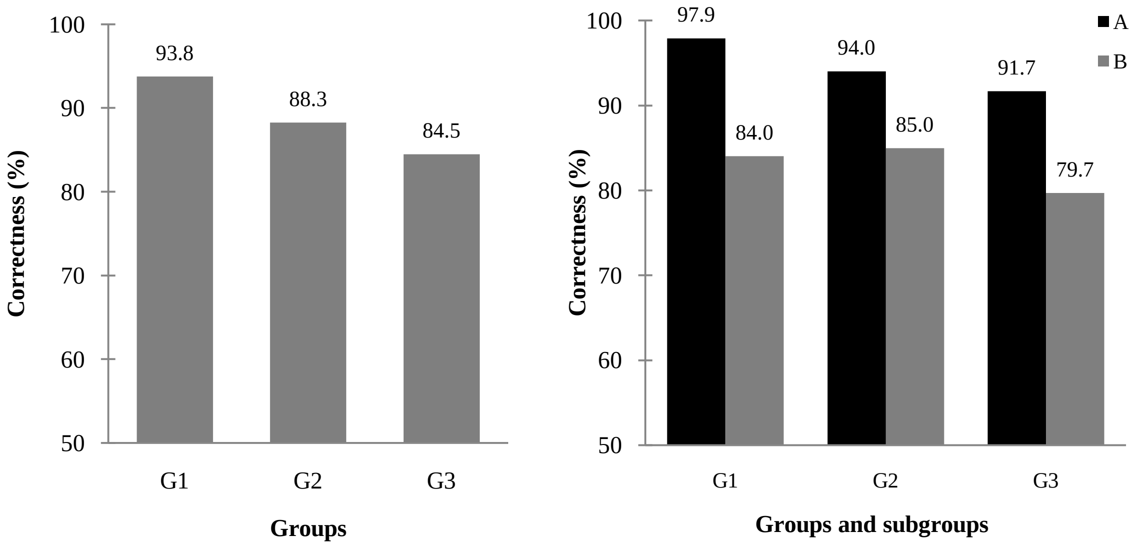
3. RESULTS

3.1. Performance by the Groups

3.1.1. Between groups

The participants across all three groups (G1, G2 and G3) showed high level of correct discrimination. Across all groups 88.81% of responses were correct (SD= 31.52). However, as seen in Figure 1, the average correct identification of the stimuli showed a significant group effect [F (2, 2973) = 21.28, p <0.001]. A Bonferroni post-hoc test confirmed that the performance of the naive group (G1) is significantly better than the experienced groups (G2 and G3) [p <0.001]. Also, the experienced groups showed significant difference in correctness from each other [p < 0.01].

Figure 1: Performance of different groups and their subgroups



3.1.2. Within groups

When compared among the four sets of stimuli, i.e. /t/-/tˤ/, /d/-/dˤ/, /ð/-/ðˤ/ and /s/-/sˤ/, a significant difference of performance was seen in the /d/-/dˤ/ and the /s/-/sˤ/ sets (see section 3.2 for more details). Speakers in each group fell into two distinct categories depending on their ability to distinguish the pharyngealized vs. non-pharyngealized contrasts in the /d/-/dˤ/ set. While one subgroup performed better on these contrasts almost (90% correct) the other performed below the 65% correctness mark on the /d/-/dˤ/ contrasts. Hence, we divided the data into two subgroups A and B. Subgroup A consisted of participants who had less difficulty in discriminating the /d/-/dˤ/ contrasts while subgroup B consisted of participants who had more difficulty in perceiving the /d/-/d□/ contrasts. An ANOVA test confirmed that the performance of subgroup A [$\bar{x}$=95.21, SD=21.36] was significantly different than that of subgroup B ($\bar{x}$=82.81, SD=37.73) [F (1, 2974) = 119.45, p <0.001].

In the subgroup A, there were significant differences among the three groups in terms of their correctness in distinguishing the pharyngealized and non-pharyngealized contrasts [F (2, 1437) = 11.44, p <0.001]. A Bonferroni test confirmed that within subgroup A, G1 performed significantly better than the experienced groups (p <0.01) while the performance of G2 was not significantly different from that of G3 (p >0.02). For subgroup B, performance of G1 does not differ significantly from G2 (p >0.017), but the performance of G2 and G3 differ significantly (p <0.01). Also, the performance of G1 and G3 was significantly different (p <0.01).

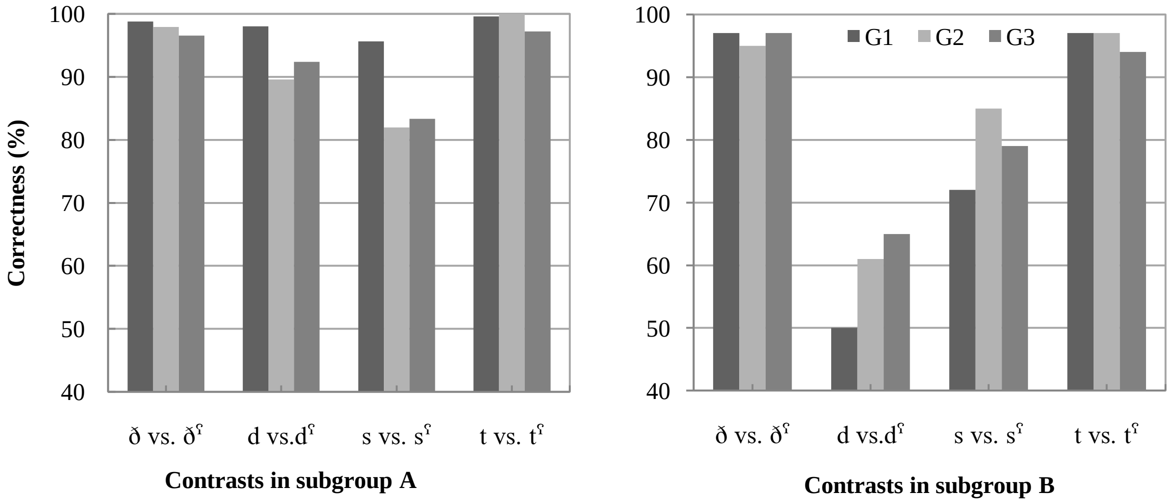
3.2. Performance by the Contrasts

Performance of the four contrasts across the groups was significantly different [F (3, 2972) = 97.34, p <0.001]. A post-hoc Bonferroni test indicated that the correctness in perceiving the /t/-/tˤ/ contrasts did not differ significantly from the correctness in perceiving the /ð/-/ðˤ/contrasts (p > 0.017). However, the correctness scores of /t/-/tˤ/ and /ð/-/ðˤ/contrasts differed significantly from the /d/-/dˤ/ and the /s/-/sˤ/ contrasts.

As mentioned in the previous section, responses from the participants were further divided into two subgroups namely, A and B. The performance of these two subgroups is shown in bar graphs in Figure 2. In Figure 2 it is noticed that the participants in subgroup A received the lowest correctness scores on the /s/-/s^{ʕ}/contrasts while the participants in subgroup B received the lowest correctness scores on the /d/-/d^{ʕ}/ contrasts. A one-way ANOVA conducted to compare the performance of the subgroups indicated that there was no group effect in the correctness scores for the /t/-/t^{ʕ}/ contrasts [F (5, 738) = 1.78, p >0.05]. For the /ð/-/ð^{ʕ}/ contrasts, only G1 and G3 subgroups showed significant difference [p < 0.017].

However, performance on the /d/-/d^{ʕ}/ contrasts revealed a significant between groups effect [F (5, 738) = 34.36, p <0.05]. The correctness score for subgroup A and subgroup B differed significantly, however, within a subgroup, there was no significant difference [F (1,742) = 158.56, p < 0.05]. This was also the case for the /s/-/s^{ʕ}/contrasts, where it showed significant difference among groups [F (5,738) = 7.347, p <0.05] caused by the significant difference between the two subgroups [F (1,712) = 12.78, p <0.05].

Figure 2: Performance by the four different contrast types

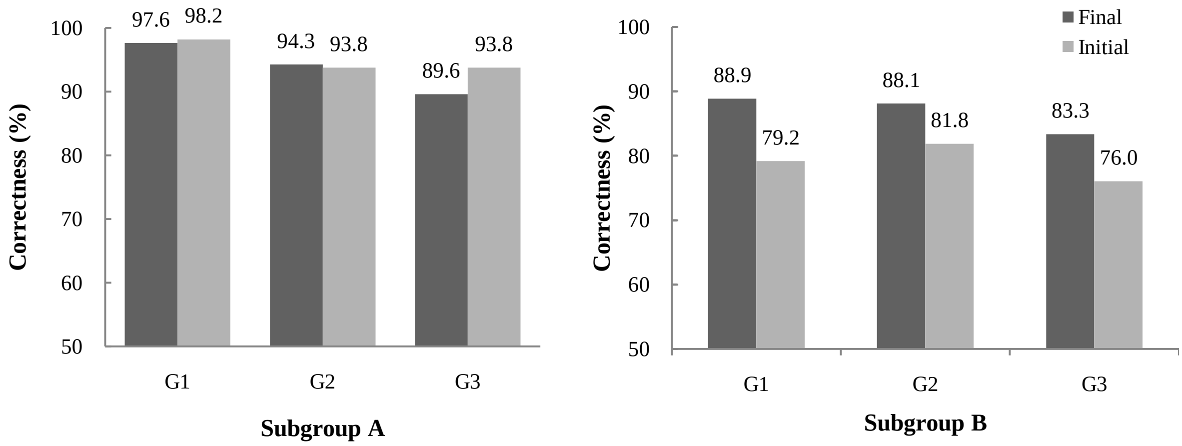


3.3. Performance by the Target Stimulus Place

The effect of the target stimulus place on the performance of the participants was analyzed- depending on whether the target stimulus occurs initially (as in the ABB and BAA type of trials) or finally (as in the AAB and BBA type of trials).

The participants performed better on the sets where the target stimuli occurred in the final position ($\bar{x}$=90.39, SD=20.9) than in the ones where they occurred in the initial position ($\bar{x}$=87.23, SD=23.1). However, an ANOVA test confirmed that this difference in correctness is not significant (p >0.05). The correctness scores were then compared by the two subgroups- A and B (see Figure 3), it was noticed that in subgroup A, there was no significant effect of target stimulus place on the correct responses (p >0.05); whereas in subgroup B target place did have an effect on performance (p <0.05).

Figure 3: Performance by target stimulus places



4. DISCUSSION

The current study showed that native speakers of Korean perceive the Arabic pharyngealized-non-pharyngealized distinstinction differently, based on their proficiency level in Arabic. The conclusions drawn from the results are discussed in the following subsections.

4.1. The naive group performs better

Among the three groups of Korean native speakers that participated in this experiment, it was noticed that the participants from the naive group (G1) was significantly better than the two experienced groups (G2 and G3) in discriminating the pharyngealized and non-pharyngealized contrasts in Arabic. Again the two experienced groups did not differ significantly in distinguishing the said contrasts. Even though this result seems to be counter-intuitive, it can be explained that the listeners in the naive, inexperienced group are probably using some kind of “phonetic mode” processing in discriminating the contrasts in Arabic.

Werker and Tees (1984) showed that the inexperienced speakers could discriminate among non native contrasts when the ISI was 500-ms however; their ability to correctly distinguish among the non-native contrasts was limited by a 1500-ms ISI. They suggested that at 1500-ms of ISI, the “phonological mode” of processing was initiated whereas at the 500-ms ISI their “phonetic mode” of processing was initiated. They also suggested that at the “phonetic mode” of processing the linguistic experience of the speakers would not influence their ability to discriminate among non-native contrasts. Burnham and Francis (1997) showed that while discriminating among some tone contrasts in Thai, the native speakers performed better in the 1500-ms ISI conditions than in the 500-ms ISI conditions.

In the current study, the ISI in the experiment was set at 1200-ms. Hence, we propose that even though the experienced speakers of the Arabic language want to implement a “phonological mode” to process the Arabic sounds, definitely an ISI of 1200-ms is not enough for them to exploit the “phonological mode” and arrive at the correct distinction. However, the non-experienced Korean speakers would solely depend on their “phonetic mode” of processing to distinguish the Arabic phonemes- in that case an ISI of 1200-ms is enough for them to correctly distinguish between the Arabic pharyngealized vs. non-pharyngealized contrasts.

4.2. Performance on the /d/-/dˤ/ contrasts

The results of current experiment also suggested that there are two distinct subgroups (A and B) of participants within each group (G1, G2 and G3). Subgroup A and B differed in their performance on distinguishing the /d/-/dˤ/ contrasts. Subgroup B was significantly worse than subgroup A in distinguishing the /d/-/dˤ/ contrasts (see Figure 2). If we look at the distribution of participants that constitute these two groups (see Table 2), we notice that most of the participants in the experienced groups (G2 and G3) had difficulty in distinguishing between the /d/and/dˤ/ in Arabic. While 70% of the naïve speakers correctly distinguished the /d/ from the /dˤ/ with above 90% correctness scores, only 38% of the experienced speakers could do so with a correctness score of 90%. In other words, it

Table 2: Correctness in percentage on the /d/-/dˤ/ contrast with number of participants within brackets

	Subgroup A	Subgroup B
G1	98% (7)	50% (3)
G2	90% (4)	61% (7)
G3	92% (4)	65% (6)

is noticed that the more the Korean speakers' get experienced with the Arabic language, the more difficult it is for them to tease apart the /d/-/dˤ/ distinction in the language. Again, as mentioned in the previous section, the experienced speakers inability to discriminate the /d/-/dˤ/contrast can be attributed to their attempt at processing the difference adopting a "phonological mode". However, it would not explain why the /d/-/dˤ/ category stands out distinctly among the contrasts the experienced learners fail to distinguish.

4.3. Effect of the target stimulus position

The current study showed that the place of the target stimuli (the odd stimuli) does have an effect in correctly distinguishing a pharyngealized consonant from its non-pharyngealized counterpart. While participants in the subgroup A did not show any such effect, the participants in subgroup B showed a significant effect of the position of the target stimuli. In case of subgroup A, the participants in the G1 and G3 groups performed marginally better when the target stimuli occurred in the initial position in the trial sets. However, in case of subgroup B, participants in all three groups performed better when the target stimuli appeared in the trial final position. This can be due to the participants' inability to retain (and subsequently adopting the "phonological mode") the information in the trial sets due to the short ISI of 1200-ms.

5. REFERENCES

AlMasri, M. & Jongman, A. 2003. Acoustic correlates of emphasis in Jordanian Arabic: preliminary results. *Proceedings of the 2003 Texas Linguistics Society Conference*, 96-106.

Best, C., Sithole, N. M., & McRoberts G. W. 1988. Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology 14* (3), 345-360.

Best, C. & Strange, W. 1992. Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics 20* (3), 305-330.

Boersma, Paul and Weenink, David. 2009. Praat: Doing Phonetics by Computer (Version 5.1.19) [Computer Program]. Retrieved October 21, 2009, from http://www.praat.org

Lisker, A. & Abramson, L. 1970, Discriminability along the voicing continuum: Cross-language tests. *Proceedings of the 6th International Congress of Phonetic Sciences, Plague* (pp. 569-573)

Werker, J. F., Gilbert, J. H., & Humphrey K. 1981. Developmental aspects of cross-language speech perception. Child Development 52, 349-355.

Werker, J. F. & Tees, R. 1984. Perceptual flexibility: maintenance or recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology 38* (4). 579-590.

Phonological aspects of the *Longman Communication 3000*

Dániel Huber

Université Rennes 2, France
huberd@freemail.hu

ABSTRACT

The paper examines the implications of the *Longman Communication 3000* word list for the teaching of English pronunciation at university level. In particular, I examined to what extent selected university curricula in France teach the pronunciation of this minimum of words and what modifications might be suggested to better serve the needs of students in equipping them with the pronunciation of contemporary, current vocabulary. It will be shown that this word list offers a wide range of useful examples to illustrate phonological phenomena, but teachers of English pronunciation/phonology might need to consider, for expository purposes, examples other than those found in standard handbooks of English phonology. I will show that using *Longman Communication 3000* does not compromise the teaching of English phonology in any major way, nor does it restrict the vocabulary to be taught. Quite the contrary, I am simply suggesting a conscious approximation of what is phonologically interesting/important to what is to be taught first.

Keywords: Spoken corpora, pronunciation teaching, word frequency

1. INTRODUCTION

The 2009 edition of the Longman Dictionary of Contemporary English (hence, LDOCE), following the earlier edition, has a valuable corpus-based list of the 3000 most frequent words in written and spoken English (excluding proper names): the *Longman Communication 3000* (LDOCE 2009: 2044-2059; hence, "the *Longman* list"). The present paper investigates the possible implications of this word list for the teaching of English pronunciation at university level, in particular English phonology and phonetics courses in the French higher education. I will argue that this list can be used effectively to choose appropriate words to illustrate a given phonological phenomenon in class, and this way we can better prepare our students to be able to determine the expected pronunciation of words not on the *Longman* list. I will show that using the *Longman* list does not compromise the teaching of English phonology in any way, nor does it restrict the vocabulary to be taught. Quite the contrary, I am suggesting a conscious approximation of what is phonologically more "interesting" (based on widely-used standard handbooks like Giegerich 1992, Guierre 1987, etc.) to what is to be taught in class (based on the *Longman* frequency list). Consequently, the purpose of this paper is not to expell lexical items from the stock of received words on the grounds that they are not listed. Quite the contrary, I am effectively proposing to increase the lexical input of pronunciation teaching both by refreshing (or up-dating) the standard stock of words used to illustrate phonological phenomena and by suggesting to cover the phonology of all the words on the *Longman* list before going on to the study of more complex, more challenging (hence, phonologically more exciting) words.

The *Longman* list has a stated applied linguistic purpose: "it shows students of English which words are the most important for them to learn and study in order to communicate effectively in both speech and writing" (LDOCE, 2009: 2044). This will naturally include their correct pronunciation(s). The present paper intends to draw attention to the phonological aspect of this list in some detail and addresses two questions in particular. First of all, I will examine whether current (French) university curricula achieve the goal of teaching the correct pronunciation of this minimum of words. To put it differently, do we actually teach the pronunciation of these words to our students? To this end, I relied on the lexical coverage of workbooks in use between 2007-2009 or currently in vigour at Université Paris 3–La Sorbonne Nouvelle and at Université Rennes 2 (Lagadec 2007, 2008, Scheuer 2009, Tang 2005, Vince et al. 2007a, Vince et al. 2007b). It is to be noted that by "teaching the correct pronunciation of this mimimun of words", I mean, perhaps simplistically, "equipping students with the theoretical means to determine the correct (or expected/regular) pronunciation of an English word (stress placement, vowel quality, reductions, etc.)". I do not examine here whether or not acquisition has actually taken place. The second question is to what extent the phonological knowledge required to correctly pronounce those words is explained in courses on English pronunciation. That is, what are the most important phonological phenomena to teach once we have this list at our disposal? The answer that emerges is that the curricula consulted cover a substantial subset of this minimum of words and often even surpass it in certain lexical domains, but on closer inspection this might be the result of mere coincidence rather than conscious curriculum development. This discrepancy seems to emerge from the fact that teaching English pronunciation tends to rely on the phonological literature rather than on what words really occur in contemporary speech. Therefore, I will argue that a conscious readjustment or approximation of the examples used may have to be considered. All the more so because such frequency lists have been conceived to serve precisely such educational needs (Leech et al. 2001: ix-x). On a more abstract level, these questions are essentially about to what extent university pronunciation teaching meets the communication needs of contemporary language learners/users. The paper argues that these words must be explicitly included in the curriculum and, of course, university courses must go well beyond them.

2. WHAT IS IN THE *LONGMAN* LIST?

The *Longman* list consists of "words", somewhat loosely understood and never explicitly defined. This looseness is particularly apparent if one approaches this list from the much fuller statistical treatment of the British National Corpus (on which LDOCE is partly based) in Leech et al. (2001). On the one hand, there are lemmas in the *Longman* list which subsume all forms of a verb like "make" – present simple 3SG, past, present and past participles–, which are counted separately in Leech et al. (2001) since all these forms do not have the same frequency of occurence. Obviously, there is no suggestion, in the *Longman* list, that only the form *make* should be learnt, but not *makes*, *making* or *made*. Nevertheless, the *Longman* list has lemmas that are inflected or derived like *amazing*, *fantastic* or *literally* where *to amaze*, *fantacy* and *literal* are clearly not undestood to be included because they are not listed while one readily finds other morphologically related pairs such as *deep–deeply*, *fail–failure*, *feel–feeling*, where both terms are. These restrictions may have some consequences for what words to include in a curriculum.

Additionally, it has to be remembered that the *Longman* list is the collation of two lists in fact: the 3000 most frequent words in (general) spoken English and the 3000 most frequent words in (general-purpose) writing. They are further classified: S1, S2 S3 stand for "among the 1000, 2000, 3000 most frequent words in speech", respectively; similarly, W1, W2, W3 for frequencies in writing. Moreover, these frequency indicators are also printed in the dictionary entries so one can immediately see the frequency of a word by looking it up. This is particularly helpful in the case of homographs, where it does matter which meaning is frequent and which is not: *lead* (of "to lead") is frequent, *lead* (metal) is not listed; similarly, *row* S2 W2 corresponds to "line of people/things", but not to *row* "dispute". Although there is considerable overlap, as could be expected, between the 3000 most frequent written and spoken words, the *Longman* list eventually runs to over 3600 words. This is considerable: "Analysis of the Longman Corpus Network [of 390 million words] shows that these 3000 most frequent words in spoken and written English account for 86% of the language" (LDOCE, p.2044). As statistics in Leech et al. (2001) reveal, the words in the *Longman* list represent word frequency down to about 10 words per million for the spoken data and to 20 per million for the written corpus. I will not discriminate between these two word lists since university students can be reasonably expected to master the pronunciation at least of all these words. Finally, further details of the nature of the list must be mentioned: While the *Longman* list is a word frequeny list, whose primary purpose is educational, it derives from corpora, including the British National Corpus of some 100 million words, whose purpose is not specifically pedagogical. Furthermore, the contemporariness of the BNC means that 93% of all texts in it date fom the period 1985-1994 and the spoken data are no earlier than 1991.

3. IMPLICATIONS FOR TEACHING ENGLISH PRONUNCIATION

3.1. Strong and weak forms

All grammatical words (pronouns, conjunctions, determiners, all simple prepositions, auxiliary verbs) are listed, and practically all of them in all their uses are among the 1000 most frequent lemmas in both spoken and written English. This is not surprising. Nevertheless, there are some interesting details. *Yours*, *yourself* and *ourselves* are S1, but less frequent in writing, W3, W2 and W3, respectively. From among the 3SG feminine pronouns, only *she* and *her* are S1 W1, *hers* is S3 W3 and *herself* is S2 W1. *Mine* and *ours* are only among the 1000 most frequent spoken words, but not on the written word list at all. *Theirs* is listed as S3, but it is not on the written word list. The import of these statistics, especially *mine*, *ours* and *theirs*, is that they show that all grammatical words occur with very high frequency in the spoken language and not necessarily in the written language, meaning that their strong and weak forms must be mastered in oral English.

3.2. Graphophonemics, homophones, minimal pairs

One of the practical uses of the *Longman* list is that is enables teachers to choose words which they can readily use in class to illustrate regular phonological phenomena as well as certain "irregularities". In this respect, it is interesting to remark that even LDOCE (2009) does not make full use of the *Longman* list: the Pronunciation Table on the inside of the front cover has *cheer* to represent /tʃ/ and *peculiar* for /iə/, while neither figures among the 3000 most frequent words (*cheese* and *familiar* would be better). That Table also has numbers like *zero* for /z/ and *four* for /ɔː/, which do not occur in the *Longman* list (as well as *loch* for /x/, which stands for a non-English phoneme, hence not expected to figure in the *Longman* list). In this section, a brief overview is given of the applicability of the *Longman* list to letter-to-sound rules.

Graphophonemics can be approached on the basis of the *Longman* list quite fruitfully. Regular vowel letter-to-sound rules can be illustrated on numerous series. Exceptions to the general rules will include *any*, *attorney*, *break*, *bury*, *busy*, *many*, *work*. Some classes of exception represented in the *Longman* list are <ew> as in *sew*, <ow> as in *below*, *blow*, *borrow*, *fellow* (adj), *flow*, *follow*, *grow*, *narrow*, *shadow*, *soul*, *window*, *yellow*. Apparently, *after*, *craft*, *daft* (S3 in BrE only) and *draft* represent lengthening better than *haft*, *shaft*. Nice examples for a regular tense vowel before another vowel include *client*, *diet* and *giant*. While all the workbooks consulted had *most* of these examples, *all* these should be included.

Consonant letter-to-sound rules do not generally pose big problems in English. Common peculiarities like *architect*, *character* for <ch>=/k/ appear in the list (but not *choir*). Examples for silent consonant letters are numerous: *answer*, *autumn*, *bomb*, *bright*, *calm*, *castle*, *debt*, *design*, *high*, *hour*, *knee*, *whole*, *wrap*. Pairs contrasting /θ/ and /ð/ include all <th>-initial grammatical words (*this*, *thus*, *then*) vs. *theme*, *thin* (but not *thumb*), for medial position *bother*, *brother*, *father*, *leather*, *mother*, *weather* and *other*, *rather*, but not *feather*, *heather*, and for final position *bath*, *beneath*, *breath*, *cloth*, *growth* but not *myth* and *breathe* but not *loathe*. For the contrasts in dental–velar nasal finals and medial clusters *thin–thing–think*, *ban–bang–bank*, *sin–sing–sink*, and *finger* can be used rather than *hanger*. *Sing* allows us to illustrate the behaviour of *sing–singing–singer*, *long*, *strong* and *young* show the adjectival pattern *stronger/strongest*. The following "voicing pairs" are listed: *advice–to advise*, *breath–to breathe*, *close–to close*, *use–to use*, *excuse–to excuse*, *house–(housing)*, *belief–to believe*, *relief–to relieve*, *use–to use*. Additionally, *anger* and *angry* can be used to demonstrate a more complex pattern, so there is no shortage of phonologically interesting cases.

To illustrate mimimal pairs, one could choose from these: *life–live–to live*, *tear–to tear*, *wind–to wind*, *a wound–he wound*, *heart–art*, *high–I/eye*, *boat–bought*, *coat–caught*, *law–low*, *raw–row*, *cold–called*, *ball–bowl*, *cord–code*. Once the principle is familiar, one can and should expand the vocabulary to further examples. The following homophone pairs, homographs or not, from those listed in Scheuer 2009, Tang 2005, Vince et al. 2007 are on the *Longman* list: *board–bored*, *bean–been*, *book–to book*, *buy–by–bye*, *file–to file*, *great–(grateful)*, *hour–our*, *know–no*, *higher–hire*, *light* (adj)–*light* (n), *mail–male*, *match*–to *match* ("sports event" or "piece of wood used to light a fire"?), *meat–meet*, *mine–mine*, *peace–piece*, *pole–poll*, *read–red*, *he saw–sore*, *scene–seen*, *so–sew*, *sole–soul*, *steal–steel*, *won–one*, *wood–would*, *right–write*, *root–route, sea–see, sight–site, son–sun, sort–sought, wait–weight, way–weigh, where–wear, whole–hole, war–wore*.

Disyllabic words are primarily interesting for their stress placement, especially across word classes. Stress alternation is attested only in *conduct–to conduct, an increase–to increase, object–to object, record–to record, a transfer–to transfer* but many of the examples listed in Tang 2005 and Vince et al. 2007, such as *concert–to concert* (not listed), *confines* (not listed)*–to confine, conflict–to conflict* (not listed), *frequent–to frequent* (not listed), *a survey–to survey* (not listed) do not appear because one term is absent from the *Longman* list. And many, such as *accent, addict, affix, compound, compress, conscript* or *rebel*, are not listed at all. Most notable among all-time favourites, the triplet *a desert–to desert–a dessert* is only represented by *a desert*. It has to noted though that the *Longman* list does not add further pairs to the lists in the workbooks.

Non-alternating disyllabic noun–verb pairs (both terms listed) are much more numerous: *account, address, adult, advance, advice–advise, answer, appeal, approach, attack, attempt, balance, benefit, challenge, comment, concern, contact, control, copy, correct, cover, damage, decline, defeat, delay, demand, design, direct, display, email, escape, offer, practice–practise, promise, purchase, regard, release, repair, reply, report, reserve, result, return, review, supply, support*. This information can be used to insist on as full a list as possible of the alternating pairs in the preceding paragraph, going thus beyond the *Longman* list.

3.3. Suffixes

Numerous suffixes are represented in the list, stress-neutral and stress-fixing suffixes (or endings), Germanic and Latinate alike. Grammatical suffixes like plural *-(e)s*, comparative and superlative *-er, -est* are only represented in lexicalized cases (*best, means*, etc). This is only to be expected since the predictable forms are subsumed under their lemmas. Past particple *-(e)d/-(e)n*, and present participle *-ing* forms appear much more frequently since they are often lexicalized as adjectives: *bored, boring; feeling*. Sometimes such lemmas are listed without the base verb being listed: *confused, confusing, exciting*. The same goes for adjectival *-ly* as in *hardly* (although many perfectly regular adverbs such as *frequently* also made it on the list). Stress-neutral derivational suffixes are amply represented. For French students of English, the suffix *-age* is of great importance and it is represented by 20 words including *advantage, garbage, manage, package, passage, percentage, storage* rather than *baggage, luggage, orphanage, pilgimage, pillage, salvage, savage*. Other stress-neutral suffixes are amply represented: *-er* (*buyer*), *-ful* (*beautiful*), *-ment* (*establishment*), *-ness* (*awareness*), *-ship* (*partnership*), *-y* (*lucky*), etc. Suffixes *-dom* (*freedom*), *-less* (*hopeless*), *-wise* (*otherwise*) figure once, and *-hood* only in *childhood* and *neighbourhood*. These suffixes are not very interesting from a phonological point of view, apart from their non-interference with stress placement and that some such suffixes do not have [ə], like *-hood* (*childhood*) and *-ship* (*championship*). (They can actually be used to demonstrate that [ə ɪ ʊ] are all examples of reduced vowels.) The suffix *-able* should be mentioned because the *Longman* list has many items that do not derive from a verb (ratio=4:6): *available, capable, inevitable, vulnerable* as opposed to *acceptable, considerable, enjoyable, remarkable, suitable, unbelievable*. Moreover, *-able* also figures in *comfortable, reasonable, valuable, vegetable* where the stem seems to be a noun rather than a verb. All in all, the particular instances of *-able* do not seem to support the implicit morphological argument that *-able* derives adjectives from verbs. The suggestion here is that the attested examples need to be explained and taught so that students can deal on their own with other words having this suffix.

What is conspicuous, however, is the fairly random presence of auto-stressed suffixes such as *-esque* among the 3000 most frequent words in English. Of this group only *-ee, -eer, -ine* and *-ique* are represented by some regular examples: *employee, guarantee, career, engineer, machine, magazine, routine, technique, unique* (together with exceptional *committee* and *coffee*). Isolated final-stressed *cartoon, cigarette, hotel, personnel, polite, precise, princess, regime, secure, severe, success* are also listed. Not all these words find their way to workbooks because they are not always easy to classify. Embarrassingly, only irregular *decade* represents its suffix. Long-time favourites such *balloon, cassette, kangaroo, mature, nominee, promenade, referee, shampoo, sincere* do not make it on the top 3000 list. This is an important factor to take into consideration in curriculum development since words like *campaign, canoe, catamaran, duress, fatigue, fricassee, sardine, trombone* (randomly gleaned from Giegerich 1992: 183 and Guierre 1981: 126) might be both new to students and their phonological pattern, even if they recognize the word, may be just an additional obstacle. Consequently, a more comprehensive coverage of words with a self-stressed suffix should be left for later. Notice that the phenomenon itself of auto-stressed suffixes can be addressed early on in the curriculum provided that *Longman* list words of *-ee*, *-eer*, *-ine*, *-ique* are used to illustrate it. In this way, a fair (and full) presentation of the phonology of English does not need to be compromised.

As for stress-fixing (non-neutral) suffixes, most of them are amply represented, which makes it possible to illustrate all stress patterns triggered by a non-neutral suffix. Again, the *Longman* list can be used to choose current, everyday lemmas to show the various patterns. The "-ion rule" (or "CiV Rule") can be illustrated on regular *-(t)ion* words (by far the most frequent suffix: 160 items) like *action*, *condition*, *impression*, *intention*, *legislation*, *vision* and irregular *television*. Other words illustrating the phenomenon: *actual*, *artificial*, *cereal*, *continuous*, *convenient*, *gorgeous*, *industrial*, *influence*, *media*, *radio*. Also *area* should be added to the standard stock of examples. *Anxiety*, *capacity*, *charity*, *community*, *facility*, *priority*, *security* and *university* amply illustrate the behaviour of *-ity*. Note in passing that *anxiety* can be profitably used to explain certain voicing patterns of <x> and *society* and *variety* can illustrate the spelling pattern <-ety>/<-ity>. But then *authenticity*, *eccentricity*, *heredity*, *notoriety*, *rarity*, *similarity*, *technicality* and *velocity* (from Lagadec 2008) are all absent from the list. The differing behaviour of *-ate* and *-ate* can be demonstrated on *associate*, *demonstrate*, *generate*, *illustrate*, *incorporate*, *locate*, *negotiate*, *participate*, *relate*, *translate* vs. *accurate*, *adequate*, *approximate*, *climate*, *immediate*, *private* and the contrasting pairs *to estimate–an estimate* and *to separate–separate*, rather than *dictate*, *emancipate*, *insinuate*, *perpetuate*, *radiate* vs. *inarticulate*, *inconsiderate*, *pirate*, *protectorate* (Lagadec 2008). "Greek suffixes" are only sporadically represented: *apology*, *psychology*, *technology*.

The *Longman* list is definitely limited in examples for stress alternations triggered by various suffixes attached to the same base. Nevertheless, examples such *associate–association*, *author–authority*, *determine–determination*, *drama–dramatic*, *economy–economic(s)*, *history–historical–historian*, *public–publicity–publication*, *science–scientific*, *operate–operator–operation*, *compete–competition–competitive*, *create–creative–creation–creature*, *specific–specify*, *resign–resignation*, *sign–signal–signiture–significant–significance*, *invest–investment*, and *investigate–investigation* can introduce students of English phonology to the phenomenon. The presence of *politics–political–politician* can be used to introduce how words irregular in their base form (here, *politics*) come to behave regularly under the influence of further suffixes.

3.4. Limits to the usefulness of the *Longman* list

Clearly there are areas of (morpho)phonology that cannot be studied based on the *Longman* list for the simple reason that most lemmas illustrating that particular phenomenon are missing from the list. First of all, since there are no proper nouns in the *Longman* list, suffixes related to names of nations like *Japanese*, *Chinese*, *Portuguese*, *Danish*, *Spanish*, *Finnish* cannot be demonstrated. Phonological classics like *Canada–Canadian* are absent. Notice that some of them follow stress placement rules which are accessible on the basis of the list: the stress pattern and the qualitiy of the stressed vowel in *Armenian* or *Hungarian* follows that of many "CiV" words in the list. Additionally, some suffixes are underrepresented (or absent like *-igible* as in *negligible*). The suffix *-(i)tude* of *exactitude*, *fortitude*, *longitude*, *magnitude*, *vicissitude* is only represented by *attitude*. Such phonological theoretical favourites as *atom–atomic–atomistic* are completely missing. Related to this is the absence of most of the exceptions to suffixes like *-ic* (only *politics* is in the list). However, a notable exception to the "CiV rule" is listed: *spiritual*. Furthermore, the list also leaves out compound words that are spelt as separate words (apart from *according to*, *credit card*, *ice cream*, *mobile phone*, *ought to*, *post office*, *used to*) or are hyphenated (apart from *long-term*). Therefore, compound stressing cannot be profitably studied.

A major drawback is that frequency indications sometimes do not make it staightforward which meaning of a given lemma is counted. For instance, *pitch* (noun) has S3 W3, but it is not clear whether "tar", "musical height" or "sports gound" is meant since these meanings, probably erroneously, are grouped under the same lemma in the dictionary, although they have different origins. Similarly, it is likely that *mummy* S1 is "diminutive of mother" rather than "dead body", although, unfortunately, they are under a common entry.

Furthermore, grammatical and phonological terminology is by and large missing, but needless to say they should figure in a curriculum: *affix–to affix*, *a compound–to compound*, *syllabify–syllabification*, *accentuate–accentuation*, *demonstrative*, *fricative*, *interrogative*, *phonology*, *predicative*, *terminology*, etc. Of course, *phonology* and *phonologist* will have to be taught, not because they made it on the *Longman* list, but because it is what we are teaching them! It has to be noted here that this expansion of the vocabulary can be adjusted to the professional specialization of students. This is all the more recommendable since the 2000-term *Longman Defining Vocabulary* (LDOCE, pp.2060-2066) itself also has items which are not in the *Longman* list.

It also has to be mentioned that curriculum developers might want or need to consider specific demands arising from the native language of the (majority) of the students targeted. No doubt, this is an important factor in France, where French has intricate and very complex linguistic–socio-cultural links with English. On the phonological level this is manifested in the pronunciation of English lexical items that bear a close relationship to a French form (semantic features of the items concerned are not discussed here). In this case, extreme care must be taken to draw attention to the more or less prominent differences in pronunciation (vowel reduction, different stress-placement, and so on). To put it differently, lexical items might also be incorporated in the curriculum to give more weight to such customized teaching goals. The relation of this method of customized item selection with the *Longman* list must be deferred to another discussion though.

4. CONCLUSIONS

The paper discussed some implications of the *Longman Communication 3000* word list for the teaching of English pronunciation. The most important observations that have emerged from this investigation are the following. Since the list enumerates all the grammatical words, vowel reduction and the basics of rhythm must continue to be included high on the agenda in any curriculum of English pronunciation. Another major area of pronunciation teaching, graphophonemics, can be effectively approached on the basis of this list, covering the full list of vowel sounds and the full range of their graphic representations as well as a fair number of vocalic and consonantal peculiarities. In the area of morphophonology, the *Longman* list is particularly useful since it can help us both to choose current, everyday lexical items for illustration in class and to assemble realistic, useful, everyday lists of exceptions instead of long lists of words that are rare anyway. However, minimal pairs (both segmental and stress-based), homographs, homonyms and homophones should be chosen more carefully for expository purposes since often one term of the pair (or both of them) is of so much lower frequency that it is not on the list. This does not mean, of course, that the relations between spelling and pronunciation do not need to be taught or their teaching is compromised. It can be established that workbooks generally cover the *Longman* vocabulary and often go beyond it, as they should. I have made a few suggestions as to what examples could replace or complement already existing vocabulary material. I have tried to show that approaching pronunciation teaching on the basis of corpora is relevant in providing teachers and students with up-to-date language material to work on.

5. REFERENCES

Giegerich, Heinz J. 1992. *English Phonology: An introduction*. Cambridge: Cambridge University Press.

Guierre, Lionel. 1987. *Règles et exercices de prononciation anglaise*. Paris : Armand Colin-Longman.

Lagadec, M-J. 2007. *Anglais oral A1 B11* [workbook for BA 1st year students]. Paris: Service des Publications, Université Paris III–La Sorbonne Nouvelle.

Lagadec, M-J. 2008. *Anglais oral A2 B11* [workbook for BA 1st year students]. Paris: Service des Publications, Université Paris III–La Sorbonne Nouvelle.

LDOCE = *Longman Dictionary of Contemporary English*. 2009. Harlow: Pearson Education Limited.

Leech, G., Rayson, P., Wilson, A. *Word Frequencies in Written and Spoken English* (*based on the British National Corpus*). Harlow: Pearson Education Limited.

Scheuer, S. 2009. *Workbook 2009-2010. Anglais oral A4 B11* [workbook for BA 2nd year students]. Paris: Service des Publications, Université Paris III–La Sorbonne Nouvelle.

Tang, Maria. 2005. *Langue orale. Laboratoire* [workbook for BA 3rd year students]. Rennes: Université Rennes 2.

Vince, J., Caillate, C., Wright, G. 2007a. *Anglais oral A4 B11* [workbook for BA 2nd year students]. Paris: Service des Publications, Université Paris III–La Sorbonne Nouvelle.

Vince, J., Kühnert, B., Caillate, C. 2007b. *Anglais oral A3 B11* [workbook for BA 2nd year students]. Paris: Service des Publications, Université Paris III–La Sorbonne Nouvelle.

Phonetic cues used by Swedish speaking learners in perception of Japanese quantity

Miyoko Inoue

Nagoya University
inoue.miyoko@nagoya-u.jp

ABSTRACT

Swedish and Japanese are known as languages that have quantity and both of them use duration as a primary acoustic cue for the long/short distinction. An identification task was conducted for Swedish learners of Japanese (SJ) and native Japanese speakers (NJ) with duration and pitch change as the variables. The results showed that SJ performed similarly to NJ with the variation of duration and pitch accent for long/short vowels, but not with pitch accent for long/short consonants. This suggested that the Swedes employed both duration and F0 change to identify the long/short vowels like the native speakers, but used only duration for consonants.

Keywords: Japanese quantity, perception, Swedish learners, duration, pitch accent

1. INTRODUCTION

This research deals with the perception of phonological length contrast, or quantity, in Japanese as a second language. Japanese quantity is known to be difficult for learners to acquire (Sugito 1989, Muraki and Nakaoka 1990, Toda 2003), but most of the previous research so far has dealt with learners of Japanese whose mother tongues (L1) are non-quantity languages. It is unknown whether it could be easier for a speaker of a quantity language to acquire Japanese quantity or if this common feature would instead disturb the acquisition. The current research, thus, examines the perception of Japanese quantity by learners who have also a quantity language as their L1 – Swedish.

1.1. SWEDISH AND JAPANESE QUANTITY

Swedish and Japanese have long/short contrast in both vowels and consonants (Sugito 1989 for Japanese, Elert 1964 for Swedish), and a primary acoustic cue for the perception of the contrast is known to be duration in both of the languages (Hirata 2004; Hirata and Whiton 2005 for Japanese; Behne et al. 1997, 1998a, 1998b, 1999; Thorén 2008 for Swedish). As can be suggested from such similarities and also the successful result in perception of Swedish quantity by Estonian speakers (McAllister et al. 2002), one could anticipate that Swedish learners of Japanese (SJ) would be rather successful in perceiving the long and short sounds in Japanese.

However, there are also noticeable differences in Swedish and Japanese regarding phonological and phonetic characteristics related to quantity. One of them is that Swedish has complementary quantity, which means that a long vowel is followed by a short consonant (V:C) and a short vowel is followed by a long consonant (VCC) (Elert 1964). At the same time, it has been reported that, in Swedish, the duration of consonant has less impact than that of vowel (Thorén 2008) or plays even no role (Behne et al. 1998b) in the perception of quantity. In Japanese, on the other hand, it seems to be unlikely that either vowel or consonant plays a more important role than the other in the perception of quantity since quantity of vowel or consonant does not depend on quantity of the respective contiguous counterpart. This might cause asymmetry between SJ's perception of vowel and consonant in Japanese quantity.

Further, previous reports on Japanese indicated that the F0 contour affected the perception of vowel quantity (Kinoshita et al. 2002; Nagano-Madsen 1992), and the difference of accent type influenced the perceptual boundary of consonant quantity (Hirata 1990; Ofuka 2003). It is unknown whether or how the perception of quantity can be influenced by different F0 patterns or pitch accent types in Swedish. However, F0 fall of pitch accent is timed with the onset of an accented vowel in Swedish (Nagano-Madsen 1992), and therefore pitch fall can signal that the syllable is stressed and thus has a long vowel or consonant in that syllable. But it may not indicate which of vowel or consonant should be long or short. It is thus conceivable that SJ would show some different characteristics in the perception of quantity from native Japanese speakers.

Accordingly, the following characteristics can be expected in SJ's perception of Japanese quantity: (1) SJ successfully distinguishes the quantity contrast in Japanese using duration, (2) the perception of consonant quantity is not as successful as for that of vowel quantity, and (3) the perception of quantity is not influenced by pitch accent or influenced in a different way from NJ. In order to investigate whether these expectations are true, an identification task was conducted with duration and pitch accent type as the parameters.

2. METHODOLOGY

2.1. Subjects

The subjects are 23 Swedish learners of Japanese (15 males and 8 females) aged 19-27 (M=21;11, SD=1;11). The subjects are limited to speakers of central Swedish, and especially, speakers of southern Swedish are excluded since this dialect has been suggested to have quantity contrast only for vowels (Malmberg 1944). A group of native Japanese listeners (NJ) also participated in the experiment for comparison. They were 9 males and 6 females from around Tokyo area[1] aged 19-35 (M=24;5, SD=5;4).

2.2. Stimuli

The measured data of the stimuli are shown in Table 1. The original sounds with phonologically short and long vowels /mamama, mama:ma/ and consonants /papapa, papappa/ were recorded by a female native Japanese speaker. Accented and Unaccented sets were recorded, where the 2nd mora had the accent nucleus for the accented set. The tokens with durational values closest to the average were picked from each condition (Vowel/Consonant x Accented/Unaccented). The stimuli were then prepared by editing the tokens with originally 'long' V2 (for Vowel stimuli) and C3 (for Consonant stimuli) using Praat (Boersma and Weenink 2004) so that the duration varies in 7 steps. To avoid possible influence from the F0 curve of originally 'long' sounds, the F0 peak, the location of F0 peak in V2 and the final F0 were fixed at the average value of the 'long' and 'short' tokens.

Table 1: The measurements of the stimuli for the long/short vowels and consonants. The unaccented and the accented stimuli are differentiated by the final F0. Vowel stimuli are adopted from Kanamura (2008: 30, 41) with permission.

Stimulus No.	Ratio	Vowel stimuli					Consonant stimuli				
		V2 Duration (ms)	Word Duration (ms)	F0 Peak (Hz)	Peak Location in V2	Final F0 (Hz)	C3 Duration (ms)	Word Duration (ms)	F0 Peak (Hz)	Peak Location in V2	Final F0 (Hz)
S1	0.25	78	582	330	48%	242 (unac.) 136 (acc.)	85	463	295	96%	231 (unac.) 116 (acc.)
S2	0.40	128	627				136	514			
S3	0.55	168	673				188	566			
S4	0.70	213	718				239	617			
S5	0.85	259	764				290	668			
S6	1.00	303	810				341	719			
S7	1.15	349	855				392	770			

2.3. Procedure

In the identification task, the subjects listened to the stimuli and clicked on a designated key depending on whether the stimuli were heard as words with a long sound or not. Four sessions (Vowel/Consonant x 2 accent types) were given for each subject. Each session consisted of 70 stimuli (7 steps x 10 times) and they were played in a random order for every subject.

2.4. Method of analysis

If the responses were plotted on a chart where the x-axis shows the stimuli numbers (S1 to S7) and the y-axis the ratio of the 'short' responses to the stimuli, one can expect that a native speaker shows such an S-shaped curve (a reversed S-shape in the current research) with a sharp drop at an interval between a pair of adjacent stimuli. Based on the responses obtained from the identification task, the categorical boundary between short and long and the consistency of categorization (or *steepness*) were estimated following the calculation procedure by Ylinen et al. (2005). The procedure can be formulated as follows.

Let n be the number of stimuli and s_i the duration of the i:th stimuli in milliseconds for each $i \in \{1, 2, \ldots, n\}$. Further, let R_i be the number of short responses to the i:th stimuli. Let m be the number of times each stimuli was tested. For each $i \in \{1, 2, \ldots, n\text{-}1\}$ set $D_i = R_i - R_{i+1}$. Then D_i/m measures the probability that the boundary point lies between s_i and s_{i+1}. We approximate the location of the boundary point by:

$$M = \sum_{i=1}^{n} \frac{D_i}{m} \frac{s_i + s_{i+1}}{2} \tag{1}$$

Then, in order to determine the steepness, a difference curve for each subject was drawn using the values of D_i. The more sharply the drop of curve mentioned above is, the more sharply-peaked the difference curve becomes. The value of standard deviation of difference curve can then be interpreted as an index of the steepness, i.e. consistency of categorization. A smaller value of steepness, thus, means a greater consistency.

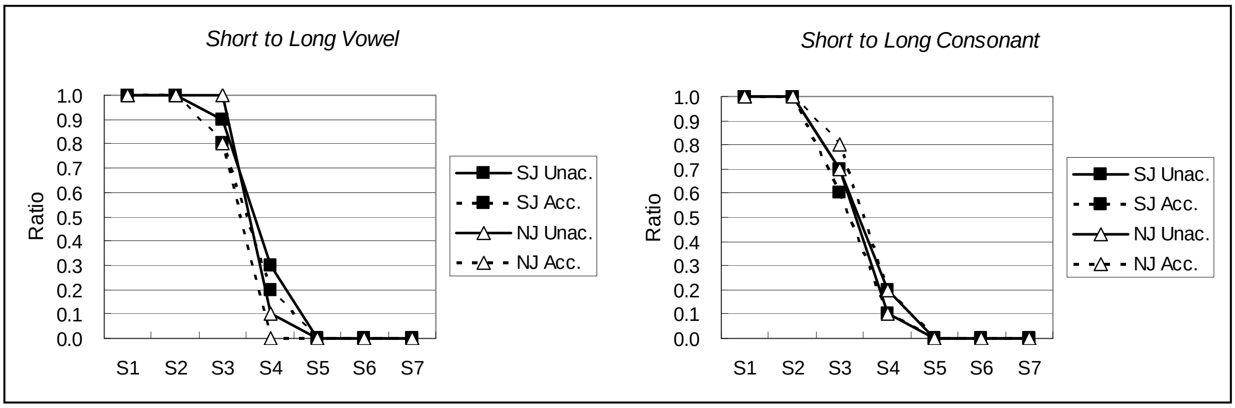
3. RESULTS

3.1. General tendency

The left chart of Figure 1 shows the mean ratio of 'short' response of all subjects for Vowel stimuli. Both of SJ and NJ display a curve with a sharp drop for both unaccented and accented words, and the sharpest fall appears between S3 and S4 for all four lines. This suggests that the categorical boundary exists somewhere between these two stimuli.

The right chart of Figure 1 shows the ratio of 'short' response for Consonant stimuli. Similar to the case of vowel stimuli, both SJ and NJ present a curve with a sharp drop for both unaccented and accented words. The most abrupt fall seems to appear between S3 and S4 similarly to the vowel case, but there is also a rather sharp drop between S2 and S3. This implies that the categorical boundary varied greater among subjects for consonants than for vowels. However, there might also be a possibility that the size of the intervals between the stimuli did not have the same size of significance for the vowels and the consonants.

Figure 1: The ratios of 'short' responses for the stimuli S1 to S7 in the Vowel series (left) and the Consonant series (right).



3.1.1. Categorical boundary

The left half of Table 2 indicates the mean values of categorical boundary and steepness for the Vowel series. A mixed two-way ANOVA was conducted for the factors Speaker Group (SJ, NJ) and Accent Type (Unaccented, Accented). As a result, no interaction between the factors was found, while the main effect of Accent Type was significant ($F(1,31)=15.010$, $p=.001$)[2]. Speaker Group, however, did not show any significant main effect. Both SJ and NJ seemed to need a longer V2 duration to respond 'long' for unaccented words than for accented words.

The right half of Table 2 presents the mean values of categorical boundary and steepness for Consonant series. A mixed two-way ANOVA for the same two factors indicated a significant interaction ($F(1,32)=5.129$, $p=.030$). Simple main effect for Speaker Group was not significant, but that for Accent Type at NJ tended to be significant ($F(1,32)=4.13$, $p=.10$). Thus, it seemed that SJ performed similarly to NJ as far as the durational variation is concerned but was not affected by Accent Type although NJ was.

Table 2: The values of categorical boundary (in millisecond) and steepness of Long/short Vowel and Consonant.

	Long/short Vowel				Long/short Consonant			
	Boundary (SD) in ms		Steepness (SD)		Boundary (SD) in ms		Steepness (SD)	
	SJ	NJ	SJ	NJ	SJ	NJ	SJ	NJ
Unaccented	199 (13)	196 (6)	28.3 (12.7)	16.1 (5.9)	216 (33)	204 (26)	52.3 (43.6)	22.1 (8.0)
Accented	192 (15)	183 (10)	29.7 (17.2)	19.1 (7.0)	205 (24)	222 (34)	40.9 (26.0)	38.7 (37.4)

3.1.2. Steepness

As mentioned above, a smaller value of steepness indicates more consistency in the categorization. The mean values of steepness by SJ were always greater than those by NJ, which demonstrates that SJ was generally less consistent in categorizing the long/short vowels than NJ was. A mixed two-way ANOVA with the same factors as above showed no interaction between the factors. No significant main effect of Accent Type was found, but there was a significant main effect of Speaker Group ($F(1,31)=9.435$, $p=.004$).

SJ's mean values of steepness for Consonant series were also always greater than those by NJ similarly to the result for Vowel series. The steepness value of SJ for Accented words was almost the same as that of NJ in comparison to other values. However, ANOVA indicated a tendency of significant main effect only for the factor Speaker Group ($F(1,32)=4.225$, $p=.048$).

3.2. Distributional pattern

3.2.1. Categorical boundary

Figure 2 displays the distribution of the categorical boundary values of each subject for Vowel series (Chart A) and Consonant series (Chart B). Chart (A) shows that the values generally range below the x=y line, which means that the boundary tend to be shorter for Accented words as in the above analysis. Most of the subjects in both SJ and NJ were found around the range of 180 to 210 ms for Unaccented words, but SJ shows a wider distribution than NJ for Accented words. The minimum values for Accented words are the same for both groups, but the maximum value of SJ exceeds that of NJ for 20 ms. SJ's having a wider range for Accented words suggests that the effect of pitch fall differed for different SJ. A little less than half of SJ (9 out of 21) fell within the same range as NJ's while the other half did not.

The distribution state for the long/short consonant (Chart B), on the other hand, shows that NJ tended to appear above the x=y line (which means an earlier boundary for Unaccented words) contrary to the long/short vowel. SJ has a greater distribution than NJ in the Vowel series. Sixteen subjects, which was more than two-thirds of all SJ appeared around the x=y line, while there were certain numbers of SJ which located either above or below the x=y line. Such a pattern of distribution might have contributed to the result in the general analysis above where there was no specific tendency of a later or earlier boundary for one of the accent types. Among such a wide range of SJ's boundary values, those came outside the range of NJ were the values visibly higher for Unaccented words. This was the opposite tendency to the Vowel series.

Figure 2: Scatter diagrams of the categorical boundary. The solid lines indicate approximate distribution ranges of SJ and the dotted lines those of NJ.

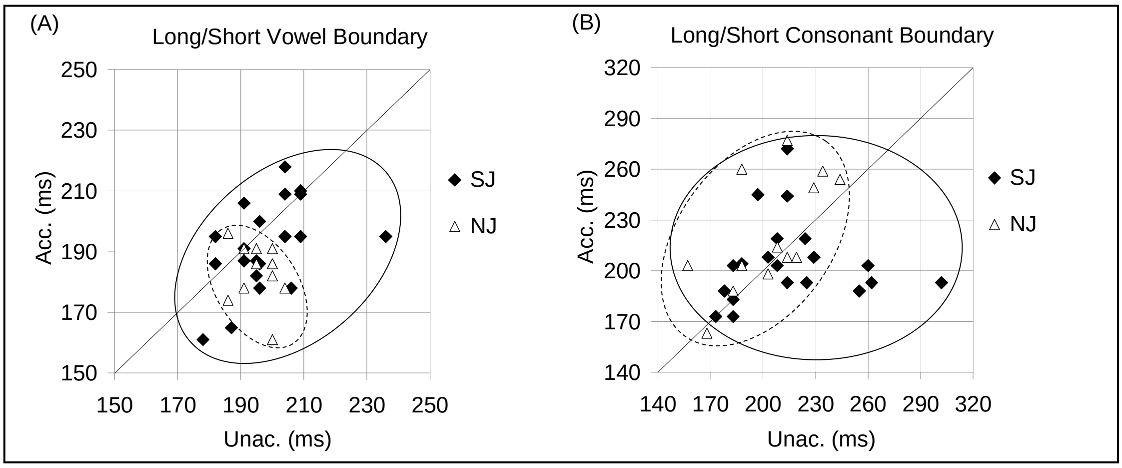
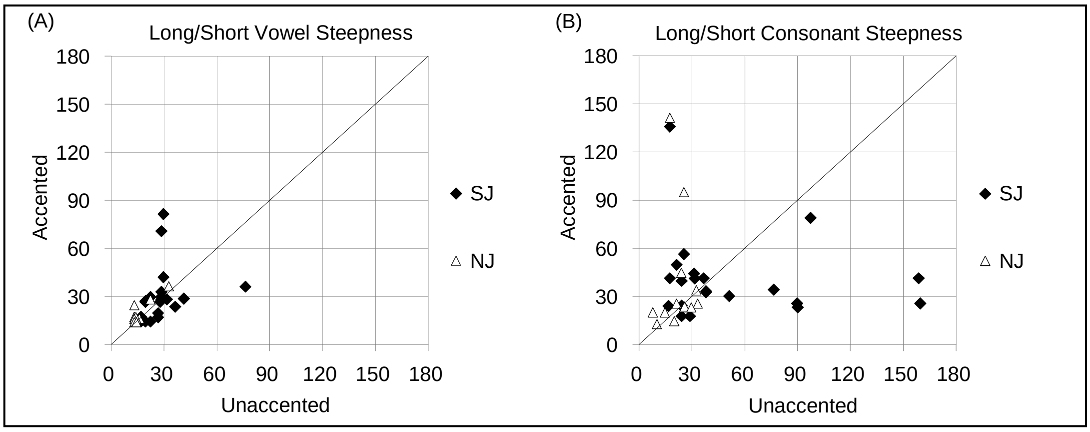


Figure 3: Scatter diagrams of the steepness.



3.2.2. Steepness

Figure 3 is the scatter charts of the steepness for Vowel series (Chart A) and Consonant series (Chart B). As for Vowel series, most of the subjects except for a few SJ appear around the range of 15.0 to 30.0 near the x=y line. Those SJ who present relatively high values were not necessarily the same subjects as those who were out of NJ's range of the categorical boundary. It thus seemed that the relatively less consistency of categorization and the deviation from NJ's categorical boundary were not directly related to each other.

The scatter chart of the steepness for Consonant (Chart B of Figure 3) also indicates that many of the subjects were concentrated in the area of 15.0 to 30.0 around the x=y line. However, both SJ and NJ showed a wider range of distribution than in the case of the long/short vowel. There were 7 subjects of SJ who visibly deviated from the mass and exceeded above 60.0. Most of them (5 subjects) appeared below the x=y line, which means that they exhibited less consistency for Unaccented words than for Accented words. Further, four of them were those who stood at the apparently higher values of the categorical boundary for Unaccented words. While most of the other SJ performed similarly to NJ in terms of both the categorical boundary and the steepness, these four SJ seemed to have difficulty in appropriately and consistently perceiving the contrast of long/short consonants when there was no pitch fall.

4. CONCLUSION

It was found out that SJ reacted similarly to NJ regarding the categorical boundary for the long/short vowels. Both of the speaker groups appeared to be affected by the accent types and presented an earlier boundary for Accented words than for Unaccented words. On the other hand, the result was different for the long/short consonants in some respects. SJ’s boundary values did not show any significant difference from NJ’s like in Vowel series, but generally, there was little effect from the accent types unlike the case of NJ.

It can be concluded that SJ used the durational change for the long/short distinction for both vowels and consonants, but used the pitch variation only for vowels and not for consonants. This result, therefore, conforms to the first prediction but only partly to the second and the third. However, individual characteristics shown by the scatter charts suggested different tendencies for different subjects. Further analyses on within-subject patterns should be conducted to obtain a clearer view of L2 quantity perception by quantity language speakers.

5. REFERENCES

Behne, D., Czigler, P., Sullivan, K. 1997. Swedish quantity and quality: A traditional issue revisited. *Phonum* 4. 81-83.

Behne, D., Czigler, P., Sullivan, K. 1998a. Perceived Swedish vowel quantity: Effects of postvocalic voicing. *Phonum* 6. 83-89.

Behne, D., Czigler, P., Sullivan, K. 1998b. Perceived Swedish vowel quantity: Effects of postvocalic consonant duration. *Phonum* 6. 91-97.

Behne, D., Arai, T., Cziglerer, P., Sullivan, K. 1999. Vowel duration and spectra as perceptual cues to vowel quantity: A comparison of Japanese and Swedish. *Proceedings of the 14th ICPhS, San Francisco*. 857-860.

Boersma, P., Weenink, D. 2004. Praat: Doing phonetics by computer (Version 4.2) [Computer program]. Retrieved March 4, 2004, from <http://www.praat.org>. Last access 4 Mar 2010.

Elert, C.-C. 1964. *Phonologic studies of quantity in Swedish*. Stockholm: Almqvist & Wiksell.

Hirata, Y. 1990. Perception of geminated stops in Japanese word and sentence levels. *The bulletin of the Phonetic Society of Japan* 194. 23-28. [In Japanese]

Hirata, Y. 2004. Effects of speaking rate on the vowel length distinction in Japanese. *Journal of Phonetics* 32(4). 565-589.

Hirata, Y., Whiton, J. 2005. Effects of speaking rate on the single/geminate stop distinction in Japanese. *The Journal of the Acousical Society of America* 118(3). 1647-1660.

Kanamura, K. 2008. Why the acquisition of long and short vowels in L2 Japanese is difficult?: The perception and recognition of Japanese long and short vowels by Chinese learners. Doctoral dissertation. Nagoya University. [In Japanese]

Kanamura, K. 2009. The L1 transition on the acquisition of L2 speech rhythm: The Japanese speech rhythm acquisition by Vietnamese and Thai native learners. Grants-in-aid for scientific research FY2007 annual research report.

Kinoshita, K., Behne, D., Arai, T. 2002. Duration and F0 as perceptual cues to Japanese vowel quantity. *Proceedings of ICSLP, Denver*. 757-760.

Malmberg, B. 1944. Die Quantität als phonetisch-phonologischer Begriff. *Lunds universitets årsskrift N.F.* Avd. I, 41(2).

McAllister, R., Flege, J., Piske, T. 2002. The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics* 30, 229-258.

Muraki, M., Nakaoka, N. 1990. Hatsuon to sokuon: Eigo-, Chugokugo-wasya no hatsuon. In: Sugito, M. (ed.) *Nihongo no onsei, on’in* [Japanese phonetics and phonology]. Tokyo: Meiji Shoin. 47-71. [In Japanese]

Nagano-Madsen, Y. 1992. Mora and prosodic coordination: A phonetic study of Japanese, Eskimo and Yoruba. *Travaux de l'Institut de Linguistique de Lund* 27. Lund: Lund University Press.

Ofuka, E. 2003. Perception of a Japanese geminate stop /tt/: The effect of pitch type and acoustic characteristics of preceding/following vowels. *Journal of the Phonetic Society of Japan* 7(1). 70-76.

Sugito, M. 1989. Onsetsu ka haku ka. In: Sugito, M. (ed.), *Nihongo no onsei on’in* [Japanese phonetics and phonology]. Tokyo: Meiji Shoin. 154-177. [In Japanese]

Thorén, B. 2008. The priority of temporal aspects in L2-Swedish prosody: Studies in perception and production. Doctoral thesis, Stockholm University.

Toda, T. 2003. Acquisition of special morae in Japanese as a second language. *Journal of the Phonetic Society of Japan* 7(2). 70-83. [In Japanese]

Ylinen, S., Shestakova, S., Alku, P., Huotilainen, M. 2005. The perception of phonological quantity based on durational cues by native speakers, second-language users and nonspeakers of Finnish. *Language and Speech* 48(3). 313-338.

NOTES

[1] The Japanese subjects are from Tokyo, Saitama, Chiba, Kanagawa and Gunma prefectures. However, as the result of the outlier examination described in the footnote 2, the subject from Gunma was excluded from the analyses.

[2] It was examined whether there were any extreme outliers to be excluded from the subsequent analysis. As a result, 2 subjects of SJ and 3 subjects of NJ were excluded from the analysis of the long/short vowels; 2 subjects each of SJ and NJ were excluded from the analysis of the long/short consonants. Thus, the number of subjects were SJ=21 and NJ=12 for the Vowel series, and SJ=21 and NJ=13 for the Consonant series.

Phonological representations and perception of L2 contrasts

Susan Jackson and John Archibald

University of Calgary
sejackso@ucalgary.ca, john.archibald@ucalgary.ca

ABSTRACT

This research investigates factors which underlie the perception of second language (L2) phonological contrasts by highlighting an asymmetry in the perception of the four-way laryngeal stop contrasts in Hindi by native speakers of English and French. We argue that this asymmetry is a result of the influence of differing underlying representations in the first language (L1). Based on a theory that English uses the feature [spread glottis] while French uses the feature [voice] to distinguish voiced and voiceless stops, it was predicted that native speakers of these two language groups would perceive the four-way Hindi contrasts differently.

Monolingual Canadian English (n=18) and monolingual Canadian French (n=18) speakers were tested on their perception of Hindi minimal pairs using an ABX discrimination task with a long interstimulus interval. Results supported the predictions; English speakers performed significantly better on contrasts involving the feature [spread glottis] and the French did significantly better on contrasts involving the feature [voice]. However, perception of all pairs of contrasts by both language groups was above chance, suggesting a role for both phonological interference and phonetic factors.

Keywords: non-native perception, laryngeal contrasts, Hindi

1. INTRODUCTION

1.1 Models of L2 perception

Successful acquisition of phonological contrasts presupposes their accurate perception. While children do this with remarkable ease (e.g. Eimas et al. 1971; Jusczyk 1997), adult learners of a second language are known to have difficulty discriminating between certain sounds that are not employed contrastively in their own language. For example, young children in diverse linguistic environments are able to perceive a contrast between /ɹ/ and /l/, which are contrastive in English but not in Japanese. However, once they begin to construct their native phonological representations, older English children maintain this contrast while Japanese children of the same age are no longer able to perceive it (Goto, 1971; Yamada, 1995).

There is variation in the *degree* of difficulty with which non-native sounds are perceived, which has led to questions about the relationship between the L1 and L2 grammars such as: what factors determine ease of perceptibility, and on what level does the L1 grammar influence the developing L2 grammar? Research in this area has resulted in several models of L2 speech perception. For example, it has been suggested that the degree of difficulty directly relates to the degree of perceived phonetic similarity or dissimilarity between L1 and L2 sounds (see Flege 1995), with more similar phones being the most difficult to perceive, thereby inhibiting the learner from setting up new phonetic categories. Along the same lines, it has been suggested that the perception of non-native contrasts is a process of categorical assimilation (e.g. Best 1995; Best and Tyler 2007) whereby two L2 phones are assimilated to L1 categories based on the perceived similarity between them. Different patterns of assimilation determine the difficulty a non-native listener will have in discriminating the contrast. For example, phones that assimilate to the same L1 category, e.g. dental and retroflex stops for a speaker of English, are the hardest to perceive, while those that assimilate to different categories, e.g. short- and long-lag stops, are easier.

Models in the generative framework argue that the presence or absence of phonological features in the L1 plays a role in the difficulty a learner may have perceiving certain second language speech contrasts (Brown 1998, 2000). This model predicts that if learners lack a particular feature in their L1 grammar that is used to distinguish an L2 contrast, they will be unable to perceive that contrast. However, if the feature in question is represented in learners' L1, regardless of how it is used, they will be able to perceive the L2 contrast employing this feature. For example, Brown (1998) found that despite the lack of phonemically contrastive liquids in Chinese, native Chinese-speaking subjects were able to perceive a contrast between English /l/ and /ɹ/. She argued that this result was due to the presence of the feature [coronal] in their L1, which is the contrastive feature in English liquids. In Chinese, however it is used to contrast fricatives. The native Japanese-speaking subjects, thought to lack this feature in their phonology, were unable to perceive the contrast.

1.2 Representations of laryngeal contrasts

In order to research L1 phonological interference in L2 speech perception, one must integrate it with a theory of phonological feature representation. Determining the correct distinctive feature in an L1 contrast, for example, is crucial to any argument about its influence on the perception of non-native contrasts.

Two types of laryngeal stop systems can be distinguished in the languages of the world: a first type that contrasts prevoiced with short-lag stops, e.g. [b/d/g] vs. [p/t/k], and a second that contrasts short-lag with long-lag stops, e.g. [b̥/d̥/g̊] vs. [pʰ/tʰ/kʰ]. Despite this difference, both these types have traditionally been thought to involve the same distinctive phonological feature, [voice], with the variation in their manifestation being a matter of language-specific phonetic implementation (see Keating 1984). Yet there is compelling evidence that this is not the case, and that the second type (short-lag and long-lag stops) uses the feature [spread glottis] contrastively and not the feature [voice] (e.g. Iverson and Salmons 1995, Kager et al. 2007). Aspiration that accompanies long-lag stops in certain positions is considered to be the phonetic result of a [spread glottis] feature that is part of the phonology. We adopt the position that English is such a language and that it contrasts with French; a true voicing language.

2. STUDY

To investigate the phonological characterization of English and French laryngeal contrasts and their influence on non-native perception, we tested the perception of Hindi laryngeal contrasts by native speakers of these two language groups. While English and French both have a two-way system of laryngeal contrasts, Hindi is a language with a four-way system that has been expressed phonologically as contrasting both [voice] and [spread glottis] (Davis 1995; Avery and Idsardi 2001). Not only does it have voiced, plain and aspirated stops, but it contrasts a fourth type which is both voiced and aspirated, also called 'breathy voice' (Ladefoged and Maddieson 1996), that occurs in neither English nor French.

2.1 Predictions

Assuming a model of L1 feature interference in L2 perception, and taking into account our assumptions about the active laryngeal features of English and French, English speakers were expected to perceive [spread glottis] contrasts well, and have difficulty with [voice] contrasts. Likewise, as French uses [voice] contrastively, French speakers were expected to perceive [voice] contrasts well and have difficulty with [spread glottis] contrasts. Pairs of Hindi stops that contrast by both [voice] and [spread glottis] were expected to be well perceived by participants of both language groups, due to the presence of the relevant laryngeal feature for each language. Predictions are indicated in Figure 1.

Figure 1: Predicted perception of all Hindi contrast types by English and French speakers. ✓ indicates good perception, x indicates poor perception.

	[voice]		[spread glottis]		both features	
Contrast	/b-p/	/bʰ-pʰ/	/pʰ-p/	/b-bʰ/	/bʰ-p/	/b-pʰ/
English	x	x	✓	✓	✓	✓
French	✓	✓	x	x	✓	✓

2.2 Method

An ABX perception task was used to test how well native-English and -French speaking subjects were able to perceive the Hindi voicing and aspiration contrasts. We used a long interstimulus interval of 1500 ms, designed to ensure phonemic judgements, as opposed to acoustic or phonetic (Werker and Logan 1985).

2.3 Participants

The participants were 18 native speakers of Canadian English and 18 native speakers of Canadian French. All were considered to be functional monolinguals, as described by Best and Tyler (2007), in that they were not actively using an L2 or in the process of learning an L2. As functional monolinguals, they were expected to have difficulty categorizing or discriminating phonetic contrasts of non-native languages that are not used to distinguish lexical items in their native language. The English speakers were all residents of Calgary, Alberta while the French speakers were residents of Montreal, Quebec. All participants reported having some experience with either English or French as a foreign language, mainly due to educational policies in Canada that require all children undergo a certain minimum of second language instruction in one of the two official languages. As well, in Montreal, exposure to ambient English through radio, television and in the public sphere is much higher than the same type of exposure to French in many other parts of Canada. In such an environment, it is difficult to establish true monolingualism. However, all participants self-reported as monolinguals in that they had not attained higher than an intermediate level of proficiency in French or English as an L2 in school and reported not to use it in any capacity in day to day life. None of the subjects reported any prior experience with Hindi.

2.4 Stimuli

Natural stimuli were developed with two female native speakers of Hindi who were recruited through the Southern Alberta Heritage Language Association (SAHLA). Hindi words and non-words whose initial stop consonants contrasted by voice, aspiration and place of articulation (dental, alveolar, retroflex and velar) were recorded digitally. All contrastive segments were word-initial and were followed by –/ʌʌl/, e.g. /pʌʌl, p^{h}ʌʌl, bʌʌl, b^{h}ʌʌl/, in order to create minimal pairs.

Each speaker recorded words singularly as well as in a carrier sentence, which was presented in Devanagari script. Stimuli for the experiment were chosen based on similar intonation, to avoid non-relevant variation influencing the participants' judgement of similarity. A total of 192 trials, counterbalanced in terms of order of words, targets and speakers, were randomized and presented in six blocks of approximately five minutes each.

2.5 Task

For each trial, participants listened to a series of three words and were told that the first two words (the minimal pair AB) differed by the first consonant, to avoid any confusion about other possible phonetic differences, such as variation in the length of the vowel. Participants were asked to judge if the third word (X) in the series was most like the first word they heard or the second. As an example, they were presented with a sequence such as (A) *paal*, (B) *baal*, and (X) *paal*. If they judged (X) to be most like the first (A) or most like the second (B), they were instructed to press number 1 or 2 respectively. A result of correct or incorrect and the speed at which they responded were displayed on screen after each trial.

3. RESULTS

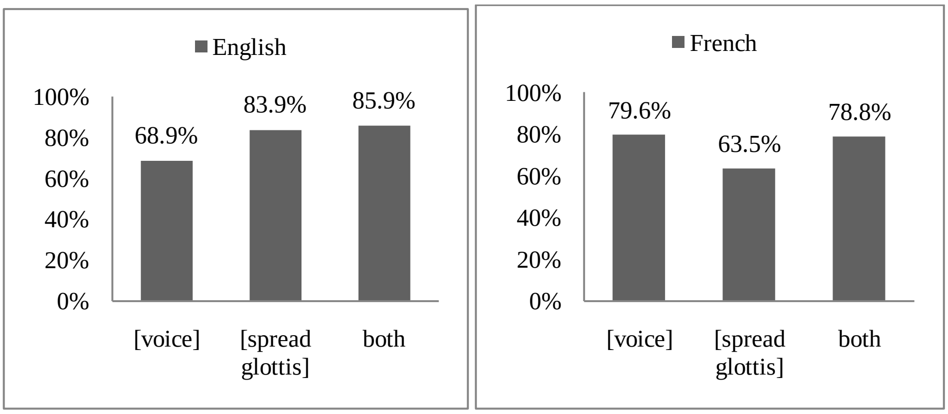
Data was analyzed using a multivariate analysis of variance (ANOVA) of the percentage of correct responses by language group. Comparisons were made between and within groups. The confidence interval was 95%.

On all pairs of stops that contrasted by [voice] only, the French subjects performed significantly better than the English subjects [$F(1, 34) = 20.05$, $p<0.05$], while on all pairs that contrasted by [spread glottis] only, the English did significantly better than the French [$F(1, 34) = 64.15$, $p< 0.05$].

The graphs in figure 1 below show that the English subjects performed similarly on pairs that contrasted by [spread glottis] and those that contrasted by both [spread glottis] and [voice], with 83.9% and 85.9%

correct responses respectively. The difference was not statistically significant (p=0.156). Therefore, it would appear their performance was unaffected by cues to contrastive [voice] in the Hindi stimuli. Likewise, regardless of the presence of aspiration, the French subjects performed quite similarly on the [voice] contrasts and those pairs that contrasted by both features with 79.6% and 78.8% correct responses respectively. Again, the difference was not statistically significant (p=0.768).

Figure 1: Average discrimination scores for pairs grouped by contrastive feature.

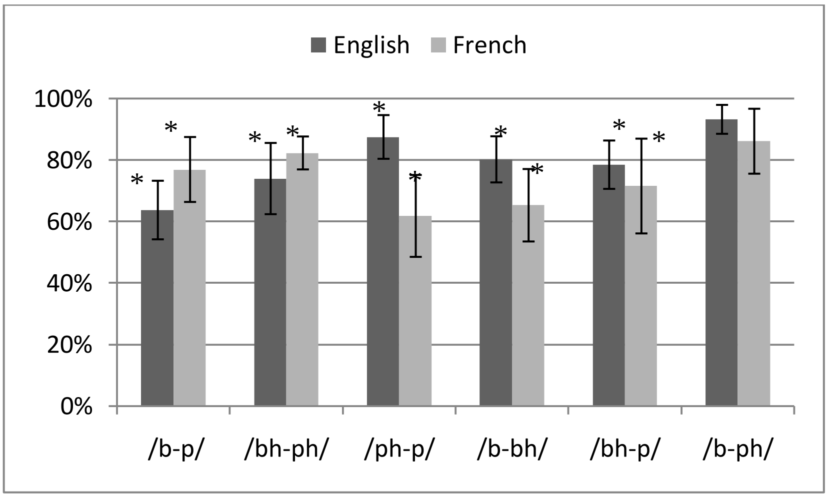


We predicted the same pattern to hold for individual pairs contrasting by these features, such that the French speakers were expected to distinguish the /b-p/ and the /bʰ-pʰ/ pairs more easily than the English speakers, as these pairs contrast by the feature [voice]. It should be kept in mind that /b/ refers to the Hindi prevoiced /b/. For /pʰ-p/ and /b-bʰ/ pairs, the opposite was predicted: the English speakers were expected to distinguish them better than the French, as these pairs contrast by the feature [spread glottis]. Performance on pairs that contrast by both features was expected to be good for both participant groups. The results from the experiment can be seen in Table 2, followed by a graph in Figure 2.

Table 2: English and French average discrimination scores for individual pairs. Numbers represent percentage of correct responses with standard deviations in parentheses. /b/= voiced, /bʰ/= voiced, aspirated, /p/= voiceless, unaspirated (or plain), and /pʰ/= voiceless, aspirated

	[voice]		[spread glottis]		both features	
Contrast	/b-p/	/bʰ-pʰ/	/pʰ-p/	/b-bʰ/	/bʰ-p/	/b-pʰ/
English	63.7 (9.6)	74.0 (11.6)	87.5 (7.1)	80.2 (7.5)	78.5 (7.9)	93.2 (4.7)
French	76.9 (10.6)	82.3 (5.4)	61.8 (13.3)	65.3 (11.8)	71.5 (15.4)	86.1 (10.6)

Figure 2: English and French average discrimination scores for individual pairs. Error bars represent standard deviation.



For all pairs that contrast by a single feature, there was a significant difference in performance between the subject groups, confirming the main predictions. However, looking at those pairs that contrast by both features, there was a significant difference in performance between groups on results for the /b-p^h/ pair, while there was no significant difference for the /b^h-p/ pair, [$F(1, 34) = 2.90$, $p<0.05$].

4. DISCUSSION AND CONCLUSION

The present study investigated the effect of differing underlying laryngeal representations on English and French speakers' perception of the Hindi four-way laryngeal contrasts. Participants' native language had a significant effect on their ability to distinguish between these contrasts. A clear asymmetric pattern of perception between language groups emerged: English speakers were able to perceive contrasts that employ the [spread glottis] feature more readily than contrasts that employ only [voice], while the reverse was found for French speakers.

However, upon examination of the results for individual contrastive pairs, additional complexities were found. For example, there was no significant difference between the results for the English speakers on the pairs /b^h-p^h/ (74.0%) and /b^h-p/ (78.5%), ($p = 0.12$), yet the former contrasts the [voice] feature only. Also, there was no significant difference between the results for the French speakers on the pairs /b-b^h/ (65.3%) and /b^h-p/ (71.5%), ($p=.08$), yet the former contrasts only the [spread glottis] feature. What is notable here is that deviations from the predicted pattern centre on voiced aspirated stops; the segment that is novel to both languages yet employs distinctive features from each.

It is possible that the presence of certain cues in the voiced aspirated stops may aid English listeners to better distinguish between a voiced aspirated and a voiceless aspirated stop despite the lack of a [spread glottis] contrast. Archibald (2005) proposes that second language learners are able to acquire phonological features that are absent from their L1 when the acoustic cues are perceptually robust. As such, their strong auditory saliency and resistance to masking (Wright, 2004) are able to override L1 phonological filters.

The individual French contrasts conformed to predictions with the exception of /b^h-p/. Interestingly, the presence of aspiration appeared to lower the ability to perceive this contrast (71.5%) for this group despite the different in voicing, although it is possible poor discrimination was based on other factors.

Finally, while Brown's Feature model of perception predicts that the absence of a feature in the L1 should prohibit discrimination of certain contrasts, all subjects performed above chance on all contrasts. This suggests that a strong interpretation of this model cannot be supported. A more likely interpretation is one in which multiple factors – phonetic and phonological – are involved in the discrimination of non-native contrasts.

5. REFERENCES

Archibald, J. 2005. Second language phonology as redeployment of L1 phonological knowledge. *Canadian Journal of Linguistics* 50: 285 -314.

Avery, P. and Idsardi, W. J. 2001. Laryngeal dimensions completion and enhancement. In: Hall T. (ed.), *Distinctive feature theory*. Berlin: Mouton de Gruyter. 41-70.

Best, C. T., and Tyler, M. D. 2007. Non-native and second-language speech perception: Commonalities and complementarities. In: O.-S. Bohn and M. Munro (eds.), *Language experience in second language speech learning: In honor of James Emil Flege*. 12-34. Amsterdam: John Benjamins.

Best, C.. T. 1995. A direct realist view of cross-language speech perception In: Strange, W. (ed.), *Speech perception and linguistic experience: issues in cross-language research*. York: Timonium. 171-204.

Brown, C. 2000. The interrelation between speech perception and phonological acquisition from infant to adult. In: John Archibald (ed.), *Second language acquisition and linguistic theory*. Oxford: Blackwell. 4-63.

Brown, C. 1998. The role of the L1 grammar in the acquisition of L2 segmental structure. *Second Language Research* 14: 136-193.

Davis, K.. 1995. Phonetic and phonological contrasts in the acquisition of voicing: Voice Onset Time production in Hindi and English. *Journal of Child Language* 22: 275-305.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., Vigorito, J. 1971. Speech perception in infants. *Science*. 171: 303–306.

Flege, J. 1995. Second language speech learning: Theory, findings, and problems. In: Winifred Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Baltimore: York Press. 233-73.

Goto, H. 1971. Auditory perception by normal Japanese adults of the sounds "L" and "R". *Neuropsychologia* 9: 317-23.

Iverson, G. and Salmons, J. 1995. Aspiration and laryngeal representation in Germanic. *Phonology* 12: 369–96.

Jusczyk, P. W. 1997. *The discovery of spoken language*. Cambridge, MA: MIT Press.

Kager, R., van der Feest, S., Fikkert, P., Kerkhoff, A., Zamuner, T. 2007. Representations of [voice]: Evidence from acquisition. In: van de Wiejer, J. M. And van der Torre, E. J. (eds.), *Voicing in Dutch*. Amsterdam: John Benjamins, 41-80.

Werker, J. and Logan, J. 1985. Cross-Language Evidence for Three Factors in Speech Perception. *Perception and Psychophysics*. 37: 35-44.

Wright, R. 2004. A review of perceptual cues and cue robustness. In: Hayes, B., Kirchner, R., Steriade, D. (eds.), *The phonetic bases of markedness*. Cambridge: Cambridge University Press. 34-57.

Yamada, R. A. 1995. Age and acquisition of second language speech sounds: perception of American English /g/ and /l/ by native speakers of Japanese. In: Strange, W. (ed.), *Speech perception and linguistic experience: issues in cross-language research*. York: Timonium. 305-320.

The Relation between Perception and Production of L2 English Lexical Stress

Jirapat Jangjamras and Ratree Wayland

Program in Linguistics, University of Florida
jirapat@ufl.edu, ratree@ufl.edu

ABSTRACT

This study investigated the perception and production of English lexical stress by native Thai speakers (NT). Thai speakers and American English (AE) speakers were first asked to produce English disyllabic non-words with initial and final stress. The syllabic structures of these disyllabic non-words were specifically designed to test the effect Thai tonal distribution rules on lexical stress assignment. They were then asked to identify stress location in the same set of disyllabic non-words produced by a trained phonetician. The results suggested that both groups were equally accurate in their production of stress, both showing a higher accuracy rate on word-initial stress. Results also showed that a syllable with a short vowel followed by an obstruent was the most difficult structure on which to accurately produce stress. NT speakers relied mainly on duration and intensity in their implementation of word-initial stress, while AE speakers employed multiple cues. Both groups produced word-final stress using duration and average F0 cues. NT were also as accurate as AE in their identification of stress location. Reaction time analysis revealed that NT spent more time on final stress identification. Finally, a moderate correlation between the perception and production of lexical stress was observed among NT.

Keywords: lexical stress, tone, second language acquisition, speech production, speech perception

1. INTRODUCTION

This study investigated prosodic transfer in both perception and production of English lexical stress by native Thai speakers who are adult learners of English. Thai and English were selected due to their vastly different prosodic systems. Thai is a tone language with phrase final stress, the dominant production cue to which is duration (Potisuk, 1995). English lacks tone and has variable stressed patterns employing dynamic cues (Cutler, 2005; Guion et al., 2003). At present, little is known about how a first language (L1) tone system may influence the perception and production of second language (L2) lexical stress.

Previous studies have revealed that an L1 stress system has an effect on L2 stress production and perception. Dupoux (2001), for example, found that native speakers of French, a fixed-stress language, had difficulty discriminating CVCV non-words that differed only in stress position, while native speakers of Spanish, a contrastive stress language, did not. Archibald (1997) investigated stress production and perception in speakers whose L1 did not have stress by using disyllabic and multisyllabic English real words as paired stimuli. Native speakers of a tone language, Chinese, were found to be better at producing stress in a read-aloud task than identifying stress location, and native speakers of a pitch accent language, Japanese, outperformed the Chinese in the stress identification task. Altman (2006) reported that native speakers of fixed-stress languages (Arabic, Turkish, and French) experienced difficulty perceiving stress location but excelled in a stress production task. In contrast, speakers from a non-lexical stress background (Chinese, Japanese, Korean) earned near-perfect stress location identification scores but produced stress that differed from AE's stress assignment patterns. However, it was not clear if the poorer performance among the speakers of non-lexical stress languages was due to their inaccurate stress assignment patterns or their non-native implementation of stress. Among the few acoustic correlate studies on L2 stress production, Adams & Munro (1978) reported that L2 stress production by eight native speakers of various Asian languages in a connected speech context was characterized by invariable patterns of F0 (rise only or fall only instead of rise/fall or fall/rise F0 contours). In addition, falls in amplitude at the boundaries of stressed syllables resulted in staccato rhythm and longer duration of unstressed syllables than those of Australian English native speakers. Chen at al. (2001) analyzed acoustic correlates of English sentential stress produced by 40 Chinese speakers with extensive length of residence in the US. Their results showed that Mandarin speakers produced stressed words with higher F0 and shorter duration; they also produced unstressed words with higher F0 and greater intensity than AE group. Zhang and Alexander (2006) also reported a higher F0 average on stressed syllables produced by Mandarin speakers of L2 English than the AE control group.

The majority of previous work on second language stress has not directly investigated what was transferred from the L1 prosodic system. Nor has it presented a comprehensive view of the relation between L2 stress production and perception. To fill this research gap, the present study focuses on the transfer of L1 tone assignment rules, stress patterns and acoustic correlates to L2 stress production and perception. Our research questions include

- Do Thai tone assignment rules based on syllabic structures affect Thai speakers' production and perception of English stress?
- Does the stress pattern in Thai influence Thai speakers' perception and production of lexical stress?
- What is the relation between L2 lexical stress perception and production?

All five Thai tones can be assigned to CVV syllables, but only the low and falling tones can occur on CVVO syllables. It was therefore predicted that Thai speakers would have difficulty producing stress on CVVO syllables. Secondly, it was predicted that Thai participants would produce final stress more accurately than initial stress due to a transfer from their L1 fixed final stress pattern. Furthermore, since Thai is a tone language with a phonemic vowel length distinction, it was predicted that Thai speakers would not have difficulty perceiving English lexical stress because they could rely on pitch and vowel duration as main perceptual cues. The combined results from the production and perception tasks were expected to yield insight into the relation between L2 stress production and perception.

2. METHODS

2.1. Participants

15 Thai and 15 AE speakers were recruited from the University of Florida. Thai participants, (age: 20-38; length of residence: 2.5 months – 5 yrs), paid for their time, included 12 graduate students, 2 undergraduate students and a postdoc. AE participants (age: 18-27) included 13 undergraduate students participating for course credit; 2 were paid for their time. All passed a bilateral hearing screen (750 – 8000 Hz at 25 dB HL).

2.2. Materials

All materials used in the speech production and perception task were drawn from a specially designed English non-word corpus. To test the influence of Thai tone constraints, only syllabic structures conforming to Thai phonotactics were chosen to create disyllabic non-word stimuli. The five syllabic structures, CVV, CVS, CVVS, CVVO, and CVO (S = sonorant, O = obstruent), were paired to generate 25 *word types* such as CVV.CVS and used in the speech perception task. Only 12 word types were used in the production task. See table 1 for a sample wordlist. Note that stress production stimuli are in parentheses.

Table 1: Selected word types and sample test words used in stress perception and production task.

3. CVV.CVVO	**4. CVV.CVO**	**7. CVVS.CVVO**	**10. CVVS.CVS**	**12. CVVO.CVO**
da:.feɪt (t^hu:.t^ha:p)	ba:.bɛt (t^hi:.t^hip)	næ:n.feɪt (t^him:.t^ha:p)	tæ:n.bɪm (t^ha:m.t^hip)	keɪt.bɛt (t^hi:p.t^hup)
13. CVVO.CVV	**14. CVVO.CVVS**	**17. CVO.CVS**	**19. CVO.CVVS**	**20. CVO.CVVO**
peɪt.ba: (t^hip:.t^hu:)	keɪt.bæ:n (t^hi:p.t^hi:m)	sɛt.bɪm (t^hʊp.t^hɪp)	sɛt.bæ:n (t^hɪp:.t^hi:m)	nɛt.feɪt (t^hup.t^hi:p)

2.3. Procedure I : Speech Production

The speech production task was administered first to avoid familiarity effects from exposure to stressed stimuli in the perception task. In the speech production task, participants were asked to concatenate two monosyllabic non-words into one single non-word with a specified stress location with the aid of a visual prompt. In one trial, for example, participants heard two nonsense syllables, [t^hip] and [t^hu:], with a 1000 ms pause in between; saw "X _"; and then produced a new disyllabic non-word with initial stress [THIP.t^hu:]. After rehearsing the new word or when ready, they produced the target non-word first in isolation and then in a sentence frame, "They said __ twice," three times. There were 72 trials in total. Stimulus presentation was controlled by E-Prime version 1.1. The speech production was recorded in a sound booth using a Marantz PMD 660 digital recorder and a Shure SM10 A head mounted microphone.

One out of three tokens of each target non-word was extracted from the sentence frame and perceptually evaluated for its stress position by AE judges and submitted to a detailed acoustic analysis. In the perceptual evaluation, a total of 2160 stimuli (72 tokens x 30 talkers) were randomly divided into two listening blocks, and each block was evaluated by 21 AE judges. The judges identified which syllable they heard as stressed by pressing a computer key: 1 for initial stress and 2 for final stress. Accuracy scores of each token were collected. Acoustic analysis was then performed using Praat to investigate acoustic properties of Thai talkers' production of stress and how they deviated from the AE talkers. The acoustic parameters measured included average F0 (Hz), F0 range (F0 max – F0 min), average intensity (dB), and duration (ms) of stressed and unstressed vowels. Note that measurements were taken from the vowel portion only, not the entire rhyme node. Obtained values were normalized by calculating the ratio of the values for stressed and unstressed vowels for all parameters except for intensity, where the difference in dB instead of the ratio was calculated.

2.4. Procedure II : Speech perception

NT and AE participants listened to pre-recorded test materials through a headset in a quiet room. They were asked to identify the stress position of each disyllable by pressing computer keys labelled '1' for initial stress or '2' for final stress as quickly as possible. The stimuli for this task were produced by a phonetician and consisted of 250 test words, half of which were produced with initial stress and the other half final stress. The stimulus presentation was controlled by E-PRIME. Accuracy scores and reaction times were collected.

3. RESULTS

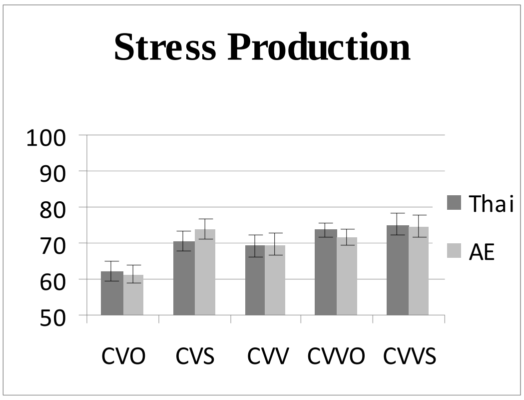
The statistical design for accuracy and reaction time analyses in this section involved a 2 x 2 x 5 ANOVA with language as a between-subjects factor (Thai and English) and Position (initial and final stress) and Syllabic Structure (CVO, CVS, CVV, CVVO, CVVS) as within-subjects factors. The Bonferroni post-hoc test (adjusted $p < .05$) was selected for pair-wise means comparisons.. When Mauchly's sphericity test result was significant, the degrees of freedom corrected by Huynh-Feldt Epsilon were reported.

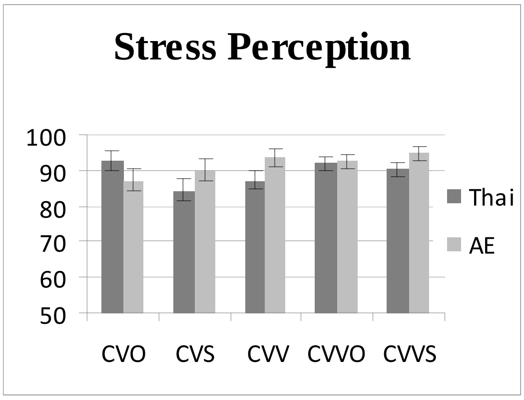
3.1. Speech production

3.1.1. Accuracy

Stress location indentification accuracy scores obtained from 21 AE judges for the 56 disyllabic non-words produced by each of the 30 speakers were used for this analysis. However, 21 judgements were excluded due to a sound file extraction error. Figure 1 shows mean percentage accuracy of stress production for both NT and AE for each of the five syllable structure types. A three-way repeated measures ANOVA analysis revealed no evidence of overall difference between NT and AE's accuracy in their production of stress [main effects of Group, $F(1,28) = .001$, $p = .978$]. The analysis also showed that their stress production accuracy varied significantly across the five syllable structures [main effects of Structure, $F(3.28,92.06) = 17.24$, $p = .000$]. Post-hoc, pair-wised comparisons indicated that stress production was less accurate for CVO (61.75%) than for the four other structures, and that CVV (69.3%) was produced significantly less accurately than CVVS (74.9%). These results indicate that CVO was the most dificult structure for stress production, CVV was not as difficult, and CVVS was the easiest. Finally, the analysis also revealed that both groups produced initial stress more accurately than final stress (73.37% vs. 67.02%) [main effect of Position, $F(1,28) = 5.47$, $p = .027$]. No significant interaction between any factors was found.

Figure 1: Mean percentage and standard errors of stress production and perception accuracy by syllabic structures by Thai and AE speakers.





3.1.2. Acoustic analysis

This analysis examined which acoustic properties AE judges relied on when identifying stress location produced by the two language groups. Four separate stepwise regressions were conducted: one for each stress position (initial and final) and one for each group. Judges' accuracy score was the dependent variable, and four acoustic parameters, duration ratio (ms), F0 range ratio (Hz), average F0 ratio (Hz) and average intensity difference (dB), were independent variables. A total of 1220 cases were included in this analysis after removing 4.5 % of acoustic measurement data considered outliers (4 SDs below or above the means).

AE initial stress regression results, based on 285 cases, showed that AE judges used all four acoustic measures when identifying stress, $F(4, 284) = 32.929$, $p<.01$, with a good model fit, $R^2 = .32$. This full model accounted for 32% of the variance. The strongest predictor of perceived stress was vowel duration (16% of the variance), followed by average F0 (additional 10%), average intensity (4%) and F0 range (3%).

Thai initial stress results, based on 287 cases, showed that a model with three predictors, with the exclusion of F0 range, was statistically reliable, $F(3, 286) = 80.6$, $p<.01$, with a very good model fit, $R^2 = .46$. The strongest predictor was vowel duration (accounted for 26% of the variance), followed by average intensity (14%) and average F0 (6%).

AE final stress results, based on 324 cases, showed that a model with three predictors was statistically reliable, $F(3, 323) = 123.323$, $p<.01$, with a very good model fit, $R^2 = .536$. This full model accounted for 54% of the variance. The strongest predictor was vowel duration (24%), followed by average F0 (22%) and then average intensity (18%).

For the Thai final stress results, based on 324 cases, a model with three predictors was statistically reliable, $F(3, 323) = 85.694$, $p<.01$, with a good model fit, $R^2 = .445$. This full model accounted for 44% of the variance. The strongest predictor was duration (27%), followed by average F0 (11%) and average intensity (6%). Standardized coefficients show that, for Thai, duration has the strongest magnitude of change in the mean of judges' rating (0.46), controlling for other variables in the model, followed by average F0 (0.31) and average intensity (0.25). In contrast, for AE, the three predictors accounted for a more nearly equal magnitude of change in ratings: duration (0.43), average f0 (0.41), and intensity (0.30).

3.2. Speech perception

3.2.1. Accuracy

7,500 stress location responses from both NT and AE were used in this analysis. Mean percentage perception accuracy by syllabic structure types is shown in Figure 1. Results of a repeated measures ANOVA indicated no significant difference between NT and AE accuracy in their identification of stress location [main effect of Group, $F(1,28) = .521$, $p = .476$]. Participants were significantly more accurate in identifying stress in initial position (93.65%) than in final position (87.3%) [main effect of Position, $F(1,28)$

= 16.69, p = .000] and their performance varied significantly across the 5 syllable structures [main effect of Structure, $F(4,112) = 4.99$, p = .001]. Post-hoc, pair-wised comparisons showed that they were less accurate on CVS (87%) than on CVVS (92.6%) and CVVO (92.2%). A significant interaction between Structure and Group [$F(4,112) = 6.18$, p = .000] was also found. Follow-up tests revealed that this was due mainly to the fact that AE, but not NT, listeners identified stress postion significantly less accurately on CVO (87.33%) than CVVS (94.80%) sylllables, CVS (90.00%) than CVVS (94.8%) syllables, and CVS than CVVO (92.53%) syllables. No other significant interaction was found.

3.2.2. Reaction time

6,438 accurate responses with RT below 5000 ms, measured from stimulus onset, were used in this analysis. RT means of each syllabic structure in each stress position were the dependent variables. A three-way repeated measures ANOVA analysis revealed no overall significant difference in response times for NT and AE [main effect of Group, $F(1,28) = .744$, p = .396]; and overall more time was spent on final stress identification (3038 ms) than initial stress (2901 ms) [main effect of Position, $F(1,28) = 31.429$, p = .000]. The analysis also suggested that RT varied significantly across the five syllabic structures [main effect of Structure, $F(3.45, 96,73) = 5.68$, p = .001]. Post-hoc, pair-wised comparisons showed that RT mean for the CVS structure (3035 ms) was significantly longer than for CVV (2930 ms), CVVO (2952 ms), and CVVS (2929 ms). Significant two-way interactions between Position and Group [$F(1,28) = 11.266$, p = .002], and Position and Structure [$F(4,112) = 3.32$, p = .013] were found. Follow-up tests revealed that the position effect was due to the fact that NT, but not AE, listeners spent more time judging word-final stress (3131 ms) than initial stress (2912 ms). The second follow-up tests also revealed that both groups spent more time on final-stress identification of CVS (3149 ms) than on CVV (2970 ms), CVVO (3033 ms), and CVVS (2964 ms).

3.3. Relation between speech perception and production

The relation between lexical stress perception and production of English non-words was investigated in two ways. First, the overall correlation between the perception and production accuracy response analyses was examined. A Pearson correlation showed a positive correlation between stress perception and production in the Thai group, although this was not statistically significant at $p<.05$ [r = .488, p = .065]; no relation was observed in AE group [r = .216, p = .439]. Further, the relationship between perception and production for specific syllabic structures was investigated. All responses from the production task and only perception responses drawn from matched 12 wordtypes were reported across the two stress positions and five syllabic structures and treated as dependent variables. Significant correlation was only found for the NT. In word-initial position, Spearman’s rho revealed a significant correlation for CVV structure [r = .549, p = .034]. In word-final position, CVV was also found to have a significant correlation between perception and production [(Pearson correlation) r = .584, p = .022; (Spearman’s rho) r = .706, p = .003 ($p< 0.01$)].

4. DISCUSSION

The present study investigated the transfer of L1 tonal rules, stress patterns and acoustic correlates onto L2 stress production and perception. First, we hypothesized that CVVO structure, occurring only with low and falling tones in Thai, would be the most difficult structure for stress production for Thai speakers due to a restriction on allowable pitch contour or pitch level on this syllabic structure. Instead, the stress production results showed that the CVO structure was the most difficult for both language groups. These results suggested that native Thai speakers’ production of lexical stress was not adversely affected by this Thai lexical tone distributional rule. Instead, similar to AE speakers, their stress production accuracy appeared to be affected by vowel duration: it was more difficult to accurately produce stress on a short vowel, suggesting that vowel lengthening was main acoustic correlate to stress implementation by Thai speakers. This is consistent with the results of the stress perception experiment. Specifically, it was found that both NT and AE listeners relied mostly on vowel duration when asked to identify stress in both initial and final positions. This supports a previous finding (Potisuk, 1995) that vowel duration lengthening is the main acoustic correlate of Thai phrasal stress production among Thai speakers. Interestingly, stress position identification accuracy among the Thai did not vary as a function of syllabic structure, and they relied less on F0 than AE listeners did when identifying stress. This result is inconsistent with previous findings (e.g. Chen at al., 2001; Zhang & Alexander, 2006) that F0 was the dominant cue to stress production among Chinese speakers. Further investigation into this contradictory finding is needed. Thai speakers have also been reported to employ less F0 contrast and F0 range than AE speakers in an oral reading and free speech task that tested intonational contrasts in their discourse (Wennerstorm, 1997). The present finding provides evidence that the restricted F0 contrast employed by Thai speakers may originate from a lower prosodic unit, lexical stress.

Based on previous reports that Thai exhibits a fixed final stress in polysyllabic words (Luksaneeyanawin, 1983), we predicted that Thai speakers would perceive and produce final stress better than initial stress. However, the results obtained did not support this hypothesis and suggested, instead, that NT produced and perceived initial stress significantly more accurately. The finding that final stress required more reaction time further confirmed Thai speakers' difficulty with final stress. Thus, it is possible that instead of word final stress, Thai actually exhibits a phrase final stress. The fact that Thai speakers experienced difficulty with final stress suggests that stress patterns at the phrasal level did not necessary facilitate L2 production and perception of stress at the word level. Further research is needed to shed light on L1-L2 transfer of stress at different prosodic levels.

Finally, the relationship between speech perception and production of English non-words was examined, and differences between the two groups were observed. A moderate positive correlation between stress production and perception based on the overall accuracy scores was found for the Thai group. A paired correlation also showed a significant positive correlation between perception and production for the CVV structure in both stress positions, but not for other syllable structures. This shows that when the phonetic environment allows longer vowel duration, Thai could rely on duration cues and showed parallel performance in both tasks. The lack of any patterns or correlation among AE might have been due to the fact that, as native speakers, they were able to rely on multiple cues and different cues, possibly not tested here, in stress perception and production.

5. CONCLUSION

In general, Thai adult learners of English can perceive and produce English lexical stress as well as native speakers of American English. We found little evidence of transfer from Thai tone assignment rules or from the Thai fixed final stress pattern on their L2 stress production and perception. However, there is evidence of L1 transfer of an acoustic correlate of stress: Thai speakers lengthened vowel duration in their implementation of stress in both Thai and English. Their reliance on vowel duration may have also been responsible for the observed positive correlation between their production and perception of long, stressed vowels.

6. REFERENCES

Adams, C., & Munro, R. R. (1978). In search of the acoustic correlates of stress: fundamental frequency, amplitude, and duration in the connected utterance of some native and non-native speakers of English. *Phonetica, 35*, 125-156.

Altmann, H. (2006). The perception and production of second language stress: a cross-linguistic experimental study. University of Delaware.

Archibald, J. (1997). The acquisition of English stress by speakers of nonaccentual languages: lexical storage versus computation of stress. *Linguistics, 35*, 167-181.

Chen, Y., Robb, M., Gilbert, H., & Lerman, J. (2001). A study of sentence stress production in Mandarin speakers of American English. *The Journal of the Acoustical Society of America, 109*(4), 1681-1690.

Cutler, A. (2005). Lexical Stress. In D. B. Pisoni & R. E. Remez (Eds.), *The Handbook of Speech Perception* (pp. 264-289). Malden: Blackwell Publishing.

Dupoux, E. (2001). A robust method to study stress "deafness". *The Journal of the Acoustical Society of America, 110*(3), 1606-1618.

Luksaneeyanawin, S. (1983). *Intonation in Thai.* University of Edinburgh.

Potisuk, S. (1995). *Prosodic disambiguation in automatic speech understanding of Thai.* Purdue University, West Lafayette.

Wennerstrom, A. (1994). Intonation meaning in English discourse: a study of non-native speakers. *Applied Linguistics*, 15(4), 399-420.

Zhang, Y. a. F., Alexander L. (2006). Acoustic correlates of English lexical stress produced by native speakers of Mandarin Chinese. *The Journal of the Acoustical Society of America, 120*(5), 3175.

Effect of L2 Phonetic Learning on the Production of L1 Vowels: a Study of Mandarin-English Bilinguals in Canada

Haisheng Jiang

University of Aberdeen
h.jiang@abdn.ac.uk

ABSTRACT

To test the hypothesis that the L1 phonetic categories established in childhood may undergo modification when similar L1 and L2 sounds interact in the process of L2 learning (Flege, 1995), a perception experiment was administered. Mandarin monolingual listeners evaluated the goodness of the Mandarin vowel production by Mandarin-English bilinguals. Results show that, when compared with Mandarin monolinguals, Mandarin-English bilinguals received a significantly lower rating for Mandarin vowel /y/, a vowel non-existent in English. Besides, some speakers were accented in the production of /a/, /aj/, /aʊ/, /e/, /i/, /oʊ/, and /u/. There is no evidence indicating that Mandarin-English bilinguals of low L1 use outnumbered those of high L1 use in being judged as accented. An acoustic analysis revealed that the acoustic dimensions that possibly contributed to Mandarin-English bilinguals' accent in L1 vowel production include lower F1, larger F2 movement and tone deviation. This study provides further evidence for the claim that the L1 phonetic system established in childhood is susceptible to change.

Keywords: L2 influence; L1 Mandarin vowel; accentedness; L1 use

1. INTRODUCTION

According to Flege (1995: 241), the interaction between the first language (L1) phonetic system and the second language (L2) phonetic system coexisting in a bilingual's mind is bidirectional in nature. That is, the L1 phonetic system influences the L2 phonetic system, and vice versa. With regard to the latter, Flege's *Speech Learning Model* (1995, 2003) makes a specific prediction that the L1 phonetic categories established in childhood do not remain static; instead, they may undergo modification when similar L1 and L2 sounds interact in the process of L2 learning. A variety of factors, such as the acquisition of L2 vowel or consonant categories, the age of L2 learning, the length of residence (LOR), the amount or extent of L1 use, and the pronunciation proficiency in an L2, are related to L2 influence on L1 at the phonetic level. Previous studies have shown that a bilingual's existing L1 phonetic categories may be reorganized in the process of acquiring the corresponding L2 phonetic categories (Flege, 1987; Guion, 2003). Both early learners (Baker & Trofimovich, 2005; Yeni-Komshian et al., 2000; Harada, 2003) and adult learners (Flege and Hillenbrand, 1984; Peng, 1993) may modify their L1 as a result of L2 learning. With regard to LOR, early bilinguals of longer LOR, but not those of shorter LOR, tend to modify their L1 phonetic categories (Baker & Trofimovich, 2005). However, the role that amount of L1 use plays in L2 influence on L1 is not apparent (McRobbie, 2003; Guion et al. 2000). This gives rise to the need to examine further whether the amount of L1 use is an important factor when examining L2 influence on an L1.

A perception test was administered to examine the influence of English learning on Mandarin vowel production among a group of Taiwanese Mandarin-English bilinguals living in Vancouver, Canada. The research questions to be addressed are "Do the Mandarin-English bilinguals have an accent in their L1 vowel production? If so, what acoustic properties are associated with this accent? Are Mandarin-English bilinguals of high L1 use and those of low L1 use equally judged as accented?"

2. Methodology

2.1. Speakers

Thirteen monolinguals of Taiwanese Mandarin (MonoM) and 33 Taiwanese Mandarin-English bilinguals differing in amount of L1 use (BiMH and BiML) participated in this study. The characteristics of MonoM, BiMH and BiML are summarized in Table 1.

Table 1: Characteristics of Mandarin monolinguals (MonoM), Mandarin-English bilinguals of high L1 use (BiMH) and Mandarin-English bilinguals of low L1 use (BiML).

		% use of Mandarin	Age (year)	AOA (year)	LOR (year)	Years of English Study	% of English TV	% of English Movie	% of English Radio
MonoM	M	97	24	24	.1	9	-----	-----	-----
	SD	(3)	(4)	(4)	(.04)	(2.0)	-----	-----	-----
	Range	90-100	17-32	17-32	.08-.16	6-12	-----	-----	-----
BiMH	M	65	22	11.6	9.9	12	66	80	65
	SD	(8)	(2)	(1.2)	(2.2)	(2.5)	(29)	(20)	(28)
	Range	60-80	19-25	10-13	6-14	6-15	10-100	20-100	10-100
BiML	M	30	21	10.9	10.4	12.9	80	91	85
	SD	(9)	(2)	(1.6)	(2.0)	(2.9)	(27)	(14)	(24)
	Range	20-40	18-24	9-13	6-14	8-17	0-100	50-100	30-100

The 13 Mandarin monolinguals (5 males and 8 females) were from Taipei, Taiwan. They were either visitors or ESL students in Vancouver. Their LOR and ages ranged from 1 to 2 months (mean=1.5 months) and 17 to 32 years (mean =24 years), respectively. The majority rated themselves “not fluent” in speaking English. Despite having some knowledge of Taiwanese, most claimed “not fluent” in speaking it. No one reported having knowledge of Hakka, a dialect spoken in Taiwan. The self-reported mean percentage of their Mandarin use at home, in school, at work and with friends was as high as 97%. This indicates that Mandarin was the main language in their daily communication.

The 33 Mandarin-English bilinguals were undergraduates at Simon Fraser University, Canada. All of them had immigrated to Vancouver from Taipei, Taiwan, with their parents when they were children. This group had an age of arrival (AOA) ranging from 9 to 13 years (mean=11.3 years). At the time of the study, their LOR and ages ranged from 6 to 14 years (mean=10.2 years) and 18 to 25 years (mean=22 years), respectively. The majority of these speakers spoke Mandarin and English only. Only 9 out of the 33 speakers reported having some knowledge of Taiwanese in addition to Mandarin and English. However, they rated themselves “not fluent” in speaking Taiwanese and claimed having very few chances to speak it in Vancouver. No one reported speaking Hakka.

Based on the amount of L1 Mandarin use, the 33 Mandarin-English bilinguals were further divided into a group of high Mandarin use (BiMH) (n=16) and a group of low Mandarin use (BiML) (n=17). Participants were asked to answer questions regarding the language most used at home, at school, at work, with friends and in day-to-day affairs. Three answers were possible. They were “Mandarin”, “English” or “Both about the same”. Following the practice of Guion et al. (2000), participants were given one point for each “Mandarin” answer, half a point for each “Both about the same” answer, and zero for each “English” answer. The possible maximum total score for a participant was five. Since the question regarding the most used language at work might not apply to every participant, it was possible some participants had answers for only four questions and thus a possible maximum total score of four. To calculate the percentage of L1 Mandarin use, a participant’s total score was divided by the number of questions answered, and then multiplied by 100.

Those who used Mandarin 40% and less were categorized as “low L1 use” and those who used Mandarin 40% and above were put into the category of “high L1 use”. The BiMH’s use of Mandarin was significantly more than that of the BiML [t (31) = 11.71, p < .001]. However, the two groups were comparable in age, AOA, LOR, years of English study, overall percentage of watching English TV programs, overall percentage of watching English movies and overall percentage of listening to English radio programs. A one-way ANOVA revealed a main effect of age between the MonoM, MonoE, BiMH, and BiML [F (1, 31) = 6.9, p<.001]. A Tukey test showed a significant difference only between MonoE and the two Mandarin bilingual groups (for both groups, p<.01). This suggests that MonoM, BiMH and BiML were comparable in age and MonoE were older than BiMH and BiML

2.2. Listeners

All 16 Mandarin listeners (7 males and 9 females) were from Taipei, Taiwan. At the time of the study, they were either adult ESL students studying at Simon Fraser University, recent immigrants or short-term visitors. Their ages ranged from 20-39 years (mean=29). Their LOR ranged from half a month to 3 months (mean=2 months). In general, their English speaking proficiency was low. They all reported having some knowledge of Taiwanese. However, their spoken Taiwanese was generally “not fluent”. No one reported having knowledge of Hakka.

2.3. Stimuli

The Mandarin vowels examined in this study were /i, y, u, a, aj, aʊ, e, oʊ/. The Chinese words containing the Mandarin vowels were “僻[pi] (out-of-the-way), 绿[ly] (green), 铺[pu] (store)，怕[pa] (fear)，派[paj] (send), 炮[paʊ] (canon), 配[pe] (match), 豆[doʊ] (bean)”. They were inserted in the sentence frame “*zhe ge zi shi*________.” (This character is ________.) to elicit production data. The production data collected in recording were used for the perception test. In total, there were 104 Mandarin tokens by Mandarin monolinguals and 264 Mandarin tokens by Mandarin-English bilinguals.

2.4. Procedure

2.4.1. Recording and acoustic analysis

Speakers were recorded in a sound-treated recording booth in a phonetics lab using a digital recorder (PMD 670 Marantz) and a microphone (SHURE KSM109). The recorder was set at a sampling rate of 44KHz and a resolution of 16-bit. The duration, F1, F2, F3 and F0 of each target vowel were measured using *Praat* (Version 4) (Boersma & Weenink, 2005).

2.4.2. Perception test

Stimuli for the perception test were divided into two blocks and were presented to listeners for goodness rating via E-Prime 1.0 on a laptop computer. Listeners were instructed to rate the goodness of the word they heard on a 7-point scale, with “1” being the worst and “7” the best exemplar of the Mandarin target word.

3. Results

3.1. Inter-rater reliability

Inter-rater reliability estimates by vowel were computed by using Cronbach’s α. The results are provided in Table 2. The values are all above.70, a cut-off point for acceptable reliability (Nunnaly, 1978).

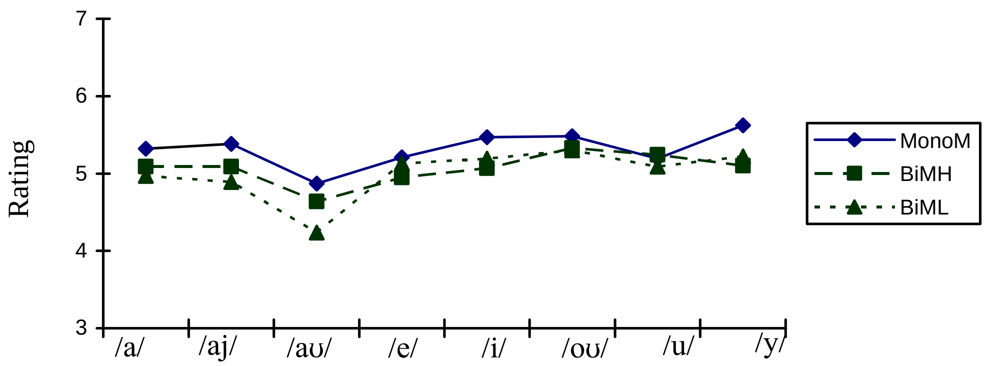
Table 2: Mean inter-rater reliability scores.

Vowel	/a/	/aj/	/aʊ/	/e/	/i/	/oʊ/	/u/	/y/
Cronbach’s α	.87	.86	.86	.85	.75	.80	.79	.76

3.2. Group differences

The mean ratings for each vowel assigned by listeners to each speaker group (pooled across listeners) are given in Figure 1. A repeated measures ANOVA was conducted on the rating scores assigned to each of the three speaker groups for /a/, /aʊ/, /e/, /i/, /oʊ/ with speaker group as a between-subjects factor (MonoM, BiMH, BiML). The analysis revealed neither a significant main effect of speaker group [F (2, 43) = 1.10, p =.34] nor significant vowel by speaker group interaction [F (2, 43) = .77, p =.630], indicating a non-significant difference between MonoM, BiMH, and BiML in the rating scores assigned to /a/, /aʊ/, /e/, /i/, and /oʊ/. For the rating scores assigned to /aj/, Mann-Whitney independent samples test (U) revealed a non-significant difference between MonoM and BiMH and between MonoM and BiML. For the rating scores assigned to /u/, two separate analyses were conducted. Mann-Whitney U test was conducted on the rating scores of MonoM and BiML due to the significant deviation from normality in the rating scores of BiML. Since the rating scores of BiMH were normally distributed, an independent samples *t*-test (two-tailed) was conducted on the rating scores of MonoM and BiMH. Neither the Mann-Whitney U test nor the *t*-test revealed a significant between group effect, indicating no significant group difference between the Mandarin-English bilinguals and the Mandarin monolinguals in the rating scores for /u/. For the same reason, two parallel tests were conducted on the rating scores assigned to vowel /y/. Both the Mann-Whitney U test [Z = -2.42, p < .05] and the *t*-test [t (28) = 2.46, p < .05] showed a significant between group effect.

Figure 1: Mean ratings assigned by listeners (n=16)



3.3. Individual differences

In Flege, Munro & MacKay's study (1995) of Italian learners of English, those who obtained a mean rating falling two standard deviations below the mean rating assigned to native English speakers were considered to have accented English pronunciation. The same accentedness criterion was adopted in the present analysis. That is, if a Mandarin bilingual's rating score is two standard deviations below the Mandarin monolinguals' mean, it is considered accented. As many as 13 Mandarin-English bilinguals (6 BiMH, 7 BiML) had rating scores falling two standard deviations below the Mandarin monolinguals' mean in /y/. This is consistent with the finding in group differences that both BiMH and BiML had significantly lower ratings than MonoM. For /oʊ/, 5 Mandarin-English bilinguals (2 BiMH, 3 BiML) were judged accented; for each of the vowels /i/ and /a/, 4 Mandarin-English bilinguals (2 BiMH, 2 BiML each) were judged accented; for /e/, 2 Mandarin-English bilinguals (1 BiMH, 1 BiML) were judged accented; for each of the vowels /aj/, /aʊ/ and /u/, 1 Mandarin-English bilingual was judged accented. As the numbers in the brackets indicate, BiML did not outnumber BiMH in being judged as accented.

3.4. Individual differences and acoustic data

The acoustic properties that possibly contribute to individual Mandarin-English bilinguals' accent are summarized in Table 3.

Table 3: Possible acoustic properties attributing to Mandarin-English bilinguals' accentedness

Vowel	Number of speakers judged as accented	Possible acoustic properties
/y/	13 (6 H, 7 L)	Lower F1
/oʊ/	5 (2 H, 3 L)	tone deviation, exaggerated duration
/a/	4 (2 H, 2 L)	tone deviation
/i/	4 (2 H, 2 L)	tone deviation
/e/	2 (2, H)	tone deviation, short duration
/aj/	1	larger upward F2 movement
/aʊ/	1	tone deviation
/u/	1	larger downward F1 movement

Note: H=high L1 use, L=low L1 use

The 13 Mandarin-English bilinguals who were judged to have accented /y/ had significantly lower F1 than MonoM, indicating that, compared with MonoM, they produced a significantly higher /y/. Among the five speakers who were judged accented in /oʊ/, one speaker produced Tone 4 instead of Tone 1. Another speaker had an exaggerated duration of 457 ms. An examination of the remaining three speakers' acoustic data did not reveal anything unusual. The acoustic parameters contributing to these speakers' accent were not identified. For vowel /a/, two speakers were found to have produced Tone 4 /pa/ like Tone 1 /pa/. It is not known why the remaining two speakers were assigned a low rating score. For vowel /i/, one speaker's production of Tone 4 /pi/ was Tone 1 /pi/. An examination of the other three speakers' acoustic data did not identify the acoustic properties that caused them to be perceived as accented. For vowel /e/, one speaker mispronounced Tone 4 /pe/ as Tone 1 /pe/. The duration of her /e/ was as short as 158 ms. The other speaker was also found to have a very short vowel duration. For /aj/, /aʊ/, /u/, each had one speaker being judged as accented. For /aj/, the speaker's accentedness was probably due to his large upward F2 movement (=3.2 bark, the second largest movement among all 33 Mandarin-English bilinguals). For /aʊ/, the speaker's production of Tone 4 /paʊ/ was actually Tone 1 /paʊ/. The difference between her F0 peak (196 Hz) and F0 valley (188 Hz) was 8 Hz only. For /u/, an examination of the speaker's acoustic data showed that his F1 movement of /u/ was -1.21 bark, the second largest among all Mandarin-English bilinguals.

4. Discussion and Conclusion

A clear pattern of accentedness was observed for many Mandarin-English bilinguals in the production of /y/. This result is surprising, given that Mandarin /y/ does not have an English counterpart. According to Flege's principle of *equivalence classification* (1987, 1992, 1995), only an L1 sound that is similar to its L2 counterpart is predicted to undergo reorganization. A possible interpretation of this finding is that the crowded vowel space in the vicinity of /y/ (Mandarin /i/, /y/ and English /i/, /ɪ/ occupy this space) triggered the raising of /y/ to allow for sufficient contrast. It is possible that the bilinguals raised their /y/ to keep it perceptually distinct from its surrounding vowels. The Mandarin-English bilinguals' raising of L1 /y/ suggests that an L1 sound that does not have an L2 counterpart and is therefore not "similar" to an L2 sound may also be adjusted to maintain perceptual contrast in the shared L1 and L2 vowel space. One speaker's larger upward F2 movement in Mandarin /aj/ may suggest the influence of English /aj/ (The F2 movement of English /aj/ is larger than that of Mandarin /aj/. See Jiang, 2008). Assuming that Flege's principle of *equivalence classification* was at work, these Mandarin-English bilinguals assimilated L2 English /aj/ with their L1 Mandarin /aj/, a result of which was that some characteristics of L2 /aj/, such as larger upward F2 movement, was carried over to the corresponding L1 vowel that, consequently, was modified. Since the acoustic properties that possibly attributed to the accented /oʊ/, /a/, /i/ and /aʊ/ were either tone deviation or duration, there is no evidence indicating a reorganization of these vowels in the bilinguals' minds. For /u/, a conclusion as to whether it has been modified is not possible due to the small number of accented Mandarin-English speakers (1 accented) and the finding that English /u/ and Mandarin /u/ do not differ in F1 movement (Jiang, 2008).

Neither the group differences nor the individual differences showed an effect of the amount of L1 use on the Mandarin-English bilinguals' accent. For example, both BiMH and BiML were significantly lower than MonoM in the rating of Mandarin /y/. A similar scenario was revealed in the analysis of individual differences, where, in most cases, there was a balanced number of BiMH and BiML whose ratings fell two standard deviations below the Mandarin monolinguals' mean. A possible interpretation of the lack of L1 use effect on L1 vowel modification is that an L2 will exert influence on an L1 if a bilingual regularly uses and is exposed to L2.

The findings in the present study further confirm the claim that the L1 system established in childhood does not remain static; instead, it may undergo modification when the L1 phonetic system and the L2 phonetic system interact in a common phonological space (Flege, 1995). However, it must be noted that an L1 segment that does not have an obvious L2 counterpart may also be reorganized as a result of L2 learning (e.g. Mandarin /y/ in this study). Therefore, phonetic segments (e.g., vowels) must be examined in the whole system in which all L1 segments reside (e.g., all vowels in a vowel space).

5. ACKNOWLEDGEMENTS

I would like to thank Zita McRobbie, Murray Munro and Yue Wang for their invaluable suggestions and Alexei Kochetov for his financial help. I also thank all those who participated in the experiments.

6. REFERENCES

Baker W., & Trofimovich P. 2005. Interaction of native and second language vowel system(s) in early and late bilinguals. *Language and Speech, 48*, 1–27.

Boersma, P., & Weenink, D. 2005. Praat: doing phonetics by computer (Version 4.3). <http://www.praat.org/>. Last access 26 May 2005.

Cronbach, L. J. 1951. Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334.

Flege, J. E. 1987. The production of "new" and "similar" phonemes in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics, 15*, 47-65.

Flege, J. E. 1992. Speech learning in a second language. In: Ferguson C., Menn L., & Stoel-Gammon C. (eds.), *Phonological development: Models, research, and application*. Timonium, MD: York Press. 565-604.

Flege, J. E. 1995. Second language speech learning: Theory, findings and problems. In: Strange W. (eds.), *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press. 233-277.

Flege, J. E., Munro, M. J. & MacKay, I. R. A. 1995. Factors affecting the strength of perceived foreign accent in a second language. *Journal of Acoustical Society of America*, *97*, 3125-3134.

Flege, J. E., Schirru, C. & MacKay, I. 2003 . Interaction between the native and second language phonetic subsystems. *Speech Communication*, *40*, 467-491.

Guion, S. G., Flege, J. E., & Loftin, J.D. 2000. The effect of L1 use on pronunciation in Quichua-Spanish bilinguals. *Journal of Phonetics*. *28*, 27-42

Guion, S. G. 2003. The vowel systems of Quichua-Spanish bilinguals: Age of acquisition effects on the mutual influence of the first and second language. *Phonetica*, *60*, 98-128.

Harada, T. 2003. L2 influence on L1 speech in the production of VOT. In: Solé M., Recasens D., & Romero J. (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS)*. Barcelona: Universitat Autónoma de Barcelona. 1085-1088

Jiang, H. 2008. Effect of L2 Phonetic Learning on L1 Vowels. PhD dissertation. Simon Fraser University.

McRobbie, Z. 2003. Sound change in progress: An acoustic study of vowel variation in the speech of Hungarians in Vancouver, British Columbia. In: McRobbie Z. & So C. (eds.), *Proceedings of the 13th Conference of the Finno-Ugric Studies Association of Canada.* Burnaby: Simon Fraser University. 83-91.

Nunnaly, J. 1978. *Psychometric theory*. New York: McGraw-Hill.

Peng, S. H. 1993. Cross-language influence on the production of Mandarin /f/ and /x/ and Taiwanese /h/ by native speakers of Taiwanese Amoy. *Phonetica*, *50*, 245-260.

Yeni-Komshian, G. H., Flege, J. E. & Liu, S. 2000. Pronunciation proficiency in the first and second languages of Korean-English bilinguals. *Bilingualism: Language and Cognition*. *3*, 131-49.

Investigating the concept of talent in phonetic performance

Matthias Jilka[1], Natalie Lewandowska[2], Giuseppina Rota[2]

[1]Dept. of English Linguistics, University of Stuttgart,

[2]Institute for Natural Language Processing, University of Stuttgart

jilka@ifla.uni-stuttgart.de, Natalie.Lewandowski@ims.uni-stuttgart.de, giuseppina.rota@ims.uni-stuttgart.de

ABSTRACT

Research in second language acquisition typically focuses on factors describing external circumstances of acquisition such as age of learning, length of residence or amount of L1 and L2 use. These parameters refer to a speaker's proficiency in the L2, but not to specific qualities inherent to the speaker, i.e., talent.

This study introduces the conceptual basis of a research project attempting to identify, evaluate and measure phonetic talent. To this purpose the notions of talent and proficiency are defined. The problem of ensuring that the administered tests are appropriate and do not just measure proficiency in a foreign language is discussed. A brief overview of the research of factors significant in second language acquisition is given and a wide variety of influences including neurophysiological, cognitive, sociopsychological or simply language use-related aspects and their possible interactions is covered. Control of as many of these external factors as possible is essential in order to know their contribution and get to the core of the underlying talent.

Keywords: concept - phonetic talent – proficiency - testing

1. INTRODUCTION

Adult individuals vary greatly in quality and speed when acquiring a second language. This is especially true for the acquisition of the L2 sound system. Phonetically oriented studies in second language acquisition (SLA) typically investigate the significance of external factors such as age of learning, age of arrival, length of residence or amount of L1 and L2 use. With respect to these factors it should thus be possible to create optimal conditions for the acquisition of a second language. The assumption of factors inherent to the speaker, on the other hand, is connected to the idea that some aspects of language learning ability are immune to external influences. This is what would generally be termed "talent" or, more weakly, a disposition toward good performance in language-related activities. Possible accounts for such inherent differences reach from special genetic equipment to particular evolved brain networks, differences in declarative (i.e. learning and use of fact and event knowledge) and procedural (i.e. acquisition and expression of motor and cognitive skills) memory (Ullman 2007), intelligence and personality factors such as motivation, extraversion or even empathy.

2. CONCEPTS OF TALENT

A distinction has to be drawn between proficiency, i.e. the overtly observable performance of a particular skill, and talent per se. Factors such as motivation, practice and experience contribute to the degree of the proficiency but are not part of the talent. This view is in agreement with popular concepts that consider talent to be an innate, somewhat mysterious ability that a person either has or does not have.

The idea that a certain talent is innate and therefore reflected in a person's biological make-up is relatively straightforward with respect to purely physical talent. Relating non-physical abilities to the brain seems to be a logical extension of this line of reasoning, but is not as widespread. While there are some neurologically-oriented studies of talent, or rather genius, as typified by such efforts as the examination of sections of Einstein’s brain (Diamond et al. 1985), the study of this field is not broadly established. However, investigations of the neurological substrates of language talent have been undertaken for example by

Geschwind and Galaburda (1985) in their description of pathological language talent as being related to the increased growth of particular brain areas (accompanied by the delayed growth in others).

2.1. Language talent

Influential neuropsychological models of the source and structure of talent are based on a model of distinct faculties where special abilities are adjacent to each other, and linguistic talent is comparable to musical, logical, spatial talent etc. (Gardner, 1983). In the majority of cases, the measurement of language talent simply consists of tests of general ability. However, in individual cases of exceptional, sometimes pathological, language talents (i.e., extremely fast and successful L2 learners) extensive tests have been applied in the attempt to diagnose the exact nature of the skill. Novoa et al. (1988), for example, used a test battery that examined a speaker's abilities with respect to IQ, vocabulary skill, verbal fluency, verbal memory, apprehension of abstract patterns, and learning of code systems. Just as language talent is considered one of many discrete talents within the model of distinct faculties, it is also generally assumed that language talent consists of different independent linguistic skills. These roughly reflect Chomskyian concepts of linguistic modularity and are identified both in different types of language disorders and processes of L2 acquisition. In Long's model of maturation, for example, the loss of competence in acquiring an L2 is cumulative, i.e. increases with age. The deterioration supposedly begins at 6 years of age with phonology being affected relatively early (Long, 1990).

2.2. The special status of phonetic skills

The special position of phonetic skills as opposed to other linguistic abilities is widely acknowledged. Typically a fundamental distinction is drawn between two substrates of linguistic ability, described as talent for grammar vs. talent for accent (Schneiderman and Desmarais, 1988). In fact, this generally assumed special difficulty of pronunciation acquisition in contrast to other linguistic features is virtually proverbial, as exemplified by the common use of the term "Joseph Conrad Phenomenon" (e.g., Bongaerts, 1999, Abu-Rabia and Kehat, 2004), which refers to the Polish-born novelist's native-like abilities in English grammar (syntax, morphology), vocabulary and style being accompanied by his strongly accented pronunciation.

In neurophysiological concepts of second language acquisition (e.g. Schneiderman and Desmarais 1988) the neurological substrates of "grammar" and "accent" are assumed to face different challenges in the acquisition of an L2: while both must display neurocognitive flexibility in order to bypass the system established for L1, in the acquisition of pronunciation there is the additional need to bypass established motor pathways in order to control articulatory movements. This additional effort is claimed to account for both the greater difficulties in acquiring the phonetic aspects of language and the differences between children and adults.

Sociopsychological concepts of language acquisition also attempt to explain the special position of pronunciation. Guiora's approach, for example, sets pronunciation apart from other linguistic features as speech constitutes a higher manifestation of self-representation (Guiora 1990). Contrary to other aspects of language, pronunciation ability and empathy are influenced by the permeability of ego boundaries (i.e., the enhanced flexibility of psychological processes), constituting the so-called "language ego". The theory predicts that if the ego boundaries are weakened, pronunciation of non-native speech sounds will improve. Studies provoking an enhancement of ego-permeability by means of hypnosis (Schumann et al. 1978), alcohol (Guiora et al. 1972) or valium (Guiora et al. 1980) claim to confirm this notion.

Non-phonetic aspects of competence in the acquisition of a second language are generally described to a smaller extent. Most such studies analyze L2 speakers' abilities in the interface of morphology, syntax and semantics. Birdsong (1992) for example tested non-native speakers' grammaticality judgments in the L2 and found not only that a large number of individuals showed native-like competence, but also that there was no difference between features of the L2 that are marked or unmarked in universal grammar.

The phonetic subsystem is generally thought to be more difficult to acquire, as it is assumed to rely mostly on hard-wired biological processes that cannot easily be influenced by conscious learning efforts. Accordingly, virtually everyone who acquires an L2 after a certain critical period (Lenneberg 1967) will exhibit a foreign accent. There is, however, no agreement regarding the cut-off point, i.e. the age at which accent-free mastery of the L2 on both the segmental and suprasegmental level should still be possible. The age limits proposed within the literature range from young infancy (Kuhl et al., 1992) over childhood (5 years, Krashen 1973) to puberty or adolescence (Johnson and Newport, 1989).

Language acquisition within this critical age period should thus always be successful and lead to native speaker status. True second language acquisition would only occur after this period and as a consequence should not allow the so-called "late learner" to attain a native-like pronunciation. Indeed the Fundamental Difference Hypothesis (Bley-Vroman, 1989) postulates an elementary distinction between child language development (L1) and foreign language learning (L2) which is caused by differences in the internal cognitive state of adults vs. children and a resulting change in the language faculty that denies adults access to Universal Grammar. The great majority of phonetics-oriented studies in SLA research, however, reject a strict interpretation of the Critical Period Hypothesis (CPH) and the implication that at best an extremely small number of late L2 learners can acquire phonetic skills that are indistinguishable from native speakers.

The general criticism of the CPH is summarized in Flege (1987), who argues that there is no discontinuity in neural development that coincides with a change in speech-learning abilities. No abrupt differences in L2 learning success can be found in studies among speakers of gradually rising age of learning (e.g., Flege 1995) as would be expected if there were a clear-cut critical period. Instead, Flege suggests the concept of a less strictly defined “sensitive period” that accounts for children’s greater abilities in the acquisition of non-native speech. He theorizes that children learn in an auditory rather than phonetic mode as they have less firmly established L1 categories. This would allow them to form more accurate perceptual targets.

Work by Hensch (2005, 2008) on the mechanisms underlying critical periods of brain development may reconcile the differing views. He states that critical periods are essential in the consolidation of neural systems by shaping cortical areas crucial for an organism’s development and survival. The primary cortex (including the auditory cortex and the motor cortex) is therefore subject to effects of a critical period, and, as a consequence, language-related abilities are affected. Once a particular function is established within the critical period it continues to organize connections to other parts of the brain on the one hand, and, on the other hand, also inhibits the establishment of competing functions. However, it is not evident that critical periods are also as important with respect to the association cortex (e.g, the connection between motor and sensory areas) which are more directly associated with language. This distinction would explain the situation encountered in second language acquisition, i.e., a clear age-dependent influence that is not equally strong with respect to the different linguistic levels (e.g. phonetics vs. syntax vs. lexicon).

3. TESTING LANGUAGE SKILLS

3.1. First and second language skills

Individual differences in language-related performance and especially acquisition ability are typically perceived with respect to proficiency in a non-native language, although general rhetorical abilities, involving such factors as choice of words (verbal intelligence), syntactic constructions and appropriateness in the pragmatic context can also be seen as expressions of proficiency and talent in the native language.

However, general abilities in speech production and perception as needed in normal communicative situations are by definition considered equal in native speakers not suffering from any language or cognitive disorders. Similarly, all native speakers should also exhibit equal levels of pronunciation skills within their native language or dialect. Exceptional control of pronunciation in the L1 may show itself in the mimcry/imitation of dialects, foreign accents or specific, characteristic voices. While a few studies see such abilities as an expression of a universal pronunciation skill and relate them to L2 pronunciation talent, most research focuses on measuring skills in the L2 alone. A notable exception is Markham (1997) who investigated talented imitators of L1 dialects and of other people, and was able to show that a considerable number of talented speakers were in fact able to also reproduce L2 speech in a native-like fashion. Another interesting approach is presented by Flege and Hammond (1982), who asked native speakers of English to read English with a Spanish accent in order to test these speakers' awareness of non-categorical, non-phonological features (such as differences in voice onset time and syllable-final lengthening) of Spanish-accented English and, therefore, by extension, of Spanish itself. The study’s focus was, however, on determining the differences in the speakers’ perception of categorical vs. non-categorical features. It did not directly investigate the speakers’ accent mimicry skills. Nevertheless, Flege and Hammond's results imply a connection between an L1-based awareness of phonetic features and L2 pronunciation skills.

Similarly, the concept of the Linguistic Coding Differences Hypothesis (Sparks et al. 1998) maintains that L1 skills are the foundation of successful L2 learning, i.e., overall L1 skill reflects overall L2 skill.

Neuropsychological accounts of L2 performance (e.g., Schneiderman and Desmarais 1988) assume a stronger separation between L1 and L2, postulating that L2 performance will be better if the cognitive pathways established for the L1 are avoided such that a direct interaction of the brain’s inherent language-processing skills with the L2’s properties is preferred.

The predominant models of phonetic second language acquisition like the Speech Learning Model (Flege 1995), the Perceptual Assimilation Model (Best 1995) or the Native Language Magnet Theory of speech perception and production (Kuhl 1991) share this basic idea and consider the representations established for the L1 to be crucial for the manifestation of foreign accent in the L2, as interference is predicted especially in cases of similar phoneme categories. In contrast to this, completely new categories are claimed to be acquired with greater accuracy. Similar effects can also be shown for prosodic phenomena (e.g., Ladd and Morton 1997, Jilka 2000).

3.2. Distinguishing proficiency and talent

A speaker's performance is of course testable in many different ways. Abilities can be examined, measured and evaluated in various task types such as reading, comprehension, speaking, grammar etc. Such tests are typically geared toward the foreign learners of a particular language, not toward native speakers.

The distinction between proficiency and talent implies of course the complication that proficiency as evaluated in a straightforward performance test consists of both inherent (i.e. talent-like), and external factors (such as, for example, amount of L1 and L2 use). It is not a trivial task to separate inherent talent from these other factors by means of experimental design and determine what it contributes to overall proficiency.

Accordingly, individual test tasks should be defined and constructed in such a way that the targeted abilities are indeed investigated. A general idea of the major influences on performance is necessary. These influences can be summarized and classified as involving first of all developmental issues such as age of learning, secondly, psychological factors like motivation and attitude towards the L2, and finally, the important question of practice and experience, i.e. amount of language use. The factor of experience can be extended to the more abstract notion of general linguistic expertise, i.e. a greater familiarity with the wide variety of possible linguistic structures (e.g. unusual types of sounds, syntactic structures etc.) even without concrete knowledge of the specific language in question.

To get at the core of "talent", these factors would have to be controlled or - even better - completely excluded. In the case of experience this could only be achieved by using tasks where experience is by definition equally high for all test subjects because it involves the L1 or a completely unknown language, where none of the test subjects have any prior experience.

This would, however, not guarantee an absolute neutralization of the factors of experience and practice as the aforementioned problem of expert knowledge still remains. Also, it may only be an idealized notion that all native speakers are equal in the amount of experience they have with their L1. In the selection of subjects for such an experiment it is therefore preferable to have a large homogenous group of the same age and "learning career", i.e. identical time and circumstances of the acquisition of the L2. Nevertheless, it is essential to collect detailed information with respect to all these issues in a questionnaire administered to each of the test subjects such that possible correspondences with performance do not remain undetected.

A similar procedure needs to be applied for the examination of the psychological factors influencing performance. A number of cognitive and sociopsychological tests and questionnaires should be carried out in order to assess various aspects of the test subjects' cognitive abilities (working memory, intelligence etc.) and personality traits (extraversion vs introversion etc.) including motivation and interest in language acquisition. Clearly, the factor of motivation has a strong influence on performance quality during a test. Motivation and ambition, thus, have to be taken into account on several levels, not only in terms of a general personality trait that helps or hinders the overall long-term success in language acquisition. There is no obvious way of preventing this factor from having an influence, but it might also be argued that psychological features, like motivation, are part of the personality, and thus part of the talent as well.

3.3. Individual factors influencing degree of performance

In the study of second language acquisition, most investigations concentrate on one individual factor and show it to have a significant influence on performance in a non-native language. It is, however, to be expected that one factor alone like, for example, age of learning onset cannot be the sole determining factor of L2 ability. Birdsong (2006, 2008) discusses the common effect of several age-related effects that include progressive L1 entrenchment, neuro-cognitive development and other biological factors. It is quite apparent that within groups of learners who acquire a certain L2 at roughly the same age, there will be some who perform better than others.

This is especially clear in the context of formal learning, e.g. in the classroom (e.g. Sparks et al. 1998), where factors such as age or amount of L1 and L2 use can be controlled to a considerable degree. The influential concept of the "critical" period is weakened by the different competence levels of learners, as well as the lack of abrupt differences in L2 learning success in speakers of gradually rising age of learning. In addition to this, it has been shown that there are natural learners (i.e., immersed in the L2 culture) who nevertheless do not achieve native-like competence despite having started acquisition well within the critical period (Flege 1987) and, on the other hand, that it is possible for late (adult) learners to attain native-like pronunciation ability (e.g., Bongaerts et al. 1995, Amunts et al. 2004). These latter studies do take factors such as talent and other favorable prerequisites (high motivation, specific pronunciation instruction) into consideration, but restrict themselves to the examination of pronunciation performance, often in tasks with a relatively low cognitive load.

As stated earlier the majority of prevalent SLA studies have been content to demonstrate the significance of the one factor they examined. However, some correlation studies intended to identify the interaction of factors that are significantly correlated with high achievement have been carried out as well, suggesting a complex mix of confounding factors. Analyses performed by Bongaerts and colleagues (Bongaerts 1999, Bongaerts et al. 1995, 2000), for example, find that on the basis of innate talent, specific pronunciation training, high motivation, substantial L2 input, and typological proximity of the L1 are very likely to lead to native-like performance.

4. APPLICATION

The challenge posed by the above reflections and insights with respect to the nature of talent and how to identify it, is of course to implement them in an actual experimental set-up. We attempt to achieve this in a large-scale research project which uses 102 native speakers of German as test subjects. Among them a core group of 50 university students of English share a number of key variables such as age, age of onset of L2 English learning and type of experience/training. The tests focus on English, as this increases the likelihood of finding individuals with native-like pronunciation skills, and because comparative (English vs. German) descriptions, for both segmental and prosodic characteristics are available. Tasks employing German (native experience) and Hindi (no experience) are also used. A large variety of phonetic skills is tested with respect to production and perception. Early results indicate an particular significance of prosody perception and interpretation abilities.

5. REFERENCES

Abu-Rabia, S., Kehat S. 2004. The Critical Period in Second Language Pronunciation: Is there such a Thing? Ten Case Studies. *Educational Psychology, 24(1)*, 77-98.

Amunts, K., Schleicher. A. & Zilles, K. 2004. Outstanding language competence and cytoarchitecture in Broca's speech region. *Brain and Language, 89*, 346–353.

Best, C. 1995. A Direct-Realist View of Cross-Language Speech Perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*. Timonium, MD. York Press, 171-204.

Birdsong, D. 1992. Ultimate Attainment in Second Language Acquisition. *Language, 68(4)*, 706-755.

Birdsong, D. 2006. Age and second language acquisition and processing: A selective overview. *Language Learning, 56*, 9-49.

Birdsong, D. 2008. Timing and Geometry of Age Effects in Second Language Acquisition. Invited Lecture at the International Symposium on Language, Mind and Brain: Mechanisms of brain plasticity and critical periods. McGill University, Montreal

Bley-Vroman, R. 1989. What is the logical problem of foreign language learning? In Gass, S., Schachter, J. (eds.), *Linguistic perspectives on second language acquisition.* New York, Cambridge University Press, 41-68.

Bongaerts, T., Planken, B., Schils E. 1995. Can late starters attain a native accent in foreign language? A test of the critical period hypothesis. In D. Singleton and Z. Lengyel (Eds.), *The age factor in second language acquisition*. Clevedon. Multilingual Matters, 30-50.

Bongaerts, T. 1999. Ultimate Attainment in L2 Pronunciation: The Case of Very Advanced Late L2 Learners. In Birdsong, D. (ed.): *Second Language Acquisition and the Critical Period Hypothesis*. Mahwah. Lawrence Erbaum Assoc.

Bongaerts, T., Mennen, S., van der Slik, F. 2000. Authenticity of Pronunciation in Naturalistic Second Language Acquisition. Studia Linguistica, 54(2), 298-308.

Diamond, M.C., Scheibel, A.B., Murphy, G.M., Harvey, T. 1985. On the Brain of a Scientist: Albert Einstein. *Experimental Neurology, 88(1),* 188-204.

Flege, J., Hammond, R., 1982. Mimicry of Non-Distinctive Phonetic Differences between Language Varieties. *Studies in Second Language Acquisition, 5(1),* 1-17.

Flege, J. 1987. A Critical Period For Learning to Pronounce Foreign Languages? *Applied Linguistics, 8,* 162-177.

Flege, J. 1995. Second Language Speech Learning: Theory, Findings, and Problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*. Timonium, MD. York Press, 233-277.

Gardner, H. 1983. *Frames of Mind: The Theory of Multiple Intelligences.* New York. Basic Books

Geschwind, N., Galaburda, A. M. 1985. Cerebral lateralization: Biological mechanisms, associations, and pathology: I. A hypothesis and a program for research. *Archives of Neurology, 42*, 428-459.

Guiora, A., Beit-Hallahmi, B., Brannon, R., Dull, C., Scovel, T. 1972. The Effects of Experimentally Induced Changes in Ego States on Pronunciation Ability in a Second Language: An Exploratory Study. *Comprehensive Psychiatry, 13(5),* 421-428.

Guiora, A., Acton, W., Erard, R., Strickland, F. 1980. The Effects of Benzodiazepine (Valium) on Permeability of Language Ego Boundaries. *Language Learning, 30(2),* 351-363.

Guiora, A. 1990. A Psychological Theory of Second Language Pronunciation. *Toegepaste Taalwetenschap in Artikelen, 37,* 15-23.

Hensch, T. 2005. Critical period period plasticity in local cortical circuits. *Nature Reviews Neuroscience, 6,* 877-888.

Hensch, T. 2008. Neural mechanisms of critical periods and their control. Invited Lecture at the International Symposium on Language, Mind and Brain: Mechanisms of brain plasticity and critical periods. McGill University, Montreal

Jilka, M. 2000. *The Contribution of Intonation to the Perception of Foreign Accent*. PhD Dissertation. Arbeiten des Instituts für Maschinelle Sprachverarbeitung 6(3), University of Stuttgart.

Johnson, J., Newport, E. 1989. Critical Period Effects in Second Language Learning: The Influence of Maturational State on the Acquisition of English as a Second Language. *Cognitive Psychology, 21*, 60-99.

Krashen, S. D. 1973. Lateralization, Language Learning, and the Critical Period: Some New Evidence. *Language Learning, 23*, 63-74.

Kuhl, P. 1991. Human Adults and Human Infants Show a Perceptual Magnet Effect for the Prototypes of Speech Categories, Monkeys do not. *Perception and Psychophysics, 50(2),* 93-107.

Kuhl, P., Wiliams, K., Lacerda, F. 1992. Linguistic Experience alters phonetic perception in infants at 6 months of age. *Science, 255,* 606-608.

Ladd, D. R., Morton, R. 1997. The Perception of Intonational Emphasis: Continuous or Categorical? *Journal of Phonetics, 25*, 313-342.

Lenneberg, E., 1967, *Biological Foundations of Language*. New York. Wiley

Long, M. 1990. Maturational Constraints on Language Development. *Studies in Second Language Acquisition, 12 (3),* 251-281.

Markham D. 1997. *Phonetic Imitation, Accent, and the Learner*. PhD Dissertation. Travaux de l'Institut de Linguistique de Lund 33. Lund University Press

Novoa, L., Fein, D., Obler, L. 1988. Talent in Foreign Languages: A Case Study. In Obler, L., Fein, D. (eds.), *The Exceptional Brain: The Neuropsychology of Talent and Special Abilities.* New York. Guildford, 294-302.

Schneiderman, E., Desmarais, C. 1988. A Neuropsychological Substrate for Talent in Second-Language Acquisition. In Obler, L., Fein, D. (eds.), *The Exceptional Brain: The Neuropsychology of Talent and Special Abilities,* New York. Guildford , 103-126

Schumann, J., Holroyd, J., Campbell, N., Ward, F. 1978. Improvement of Foreign Language Pronunciation under Hypnosis: A Preliminary Study. *Language Learning, 28(1).*

Sparks, R., Artzer, M., Ganschow, L., Siebenhar, D. , Plageman, M. 1998. Differences in Native Language Skills, Foreign-Language Aptitude, and Foreign-Language Grades. *Language Testing, 15(2),* 181-216.

Ullman, M. 2007. The Neurocognition of Second Language: New Evidence from Training and Aptitude. *Invited Lecture at Language Processing in First and Second Language Learners.* Max Planck Institute for Psycholinguistics, Nijmegen

Serbian affricates contrasts in foreign language perception investigated by means of a neurophysiological experiment

Nuria Kaufmann, Martin Meyer*, Stephan Schmid°*

Institute of Psychology (Division of Neuropsychology), University of Zurich*
Phonetics Laboratory, University of Zurich, Switzerland°
nuria_kaufmann@gmx.ch, m.meyer@psychologie.uzh.ch, schmidst@pholab.uzh.ch

ABSTRACT

This study addresses the question to which extent phonetic contrasts of a foreign language are perceived more easily by speakers of a native language that shares similar phonetic categories. The focus lies on two post-alveolar and two palatal affricates of Serbian: [tʃ] (post-alveolar, voiceless), [tɕ] (alveolo-palatal, voiceless), [ʤ] (post-alveolar, voiced) and [ʥ] (alveolo-palatal, voiced). Swiss-German dialects have the post-alveolar voiceless affricate [tʃ] only, while the Rhaeto-Romance variety of Sursilvan has three different affricates, e.g. [tʃ], [tɕ] and [ʥ].

In an EEG experiment using a MMN paradigm, 15 Swiss-German and 15 Rhaeto-Romance speakers between 20-30 years were instructed to focus on reading a random text while not paying attention to the auditory stimuli. The hypothesis was a significant difference in processing between the two groups: Swiss-Germans would not be able to reliably distinguish the four Serbian affricates. Rhaeto-Romance speakers, on the other hand, are expected to be able to distinguish all four affricates.

MMN curves revealed that both groups were able to perceive all phonetic contrasts. Swiss German speakers showed significantly higher amplitude peaks for four out of the twelve affricate contrasts. A significant group-effect was found to corroborate that Rhaeto-Romance speakers process the Serbian affricates differently than Swiss-German speakers do.

Keywords: Phonetic Contrasts, Serbian Affricates, Swiss-German, Rhaeto-Romance, MMN

1. INTRODUCTION

There is a diversified discussion on how and when we best learn a foreign language (L2). Some advocate that foreign-language learning is no longer possible without any accent after a 'Critical Period' (e.g., Lenneberg 1967; Kuhl 2004). Others plead in favor of a continuous mode of foreign-language learning which does not differ significantly between children and adults (e.g. Friederici 2005). This would conform to the assumption that language competence in the L2 affects processing patterns more significantly than age of acquisition (e.g. Winkler et al. 1999). Various models interpret the influence of the L1 on foreign language learning differently: the NLNC (Native Language Neural Commitment) and NLM (Native Language Magnet) (Kuhl 2004) propose that neural networks that are dedicated to the L1 are less sensitive towards non-native speech sounds the more established they become. According to the SLM (Speech Learning Model) (Flege 1993), new but not similar sound contrasts of the L2 are perceivable by the language learner. According to this hypothesis, only dissimilar phonetic contrasts would elicit a MMN and higher use of the L1 would influence the successful acquisition of L2 phonetic categories negatively. The PAM (Perceptual Assimilation Model) (Best et al. 2001), on the other hand, predicts that (similar) L2-sounds can be assimilated to the listener's L1 phonetic category and would elicit a MMN. As opposed to the SLM, NLNC and NLM, PAM views a well established L1 to be supportive for L2 learning.

Mismatch negativity paradigms have shown that fluent non-native speakers develop a cortical auditory memory for foreign language phonemes (Näätänen et al. 1997; Winkler et al. 1999). Even in a well-learned second language, however, phoneme representations of the native language were found to exert a strong influence on contrast detection (Nenonen et al. 2005). Consequently, different mother tongues (L1s) could out-fit one differently to learn a certain foreign language. The relevant question for our study is whether phonetic information that seems irrelevant to the acquired L1-specific representations is neglected or filtered out once L1-networks have been established, or whether listeners are still able to differentiate phonetic contrasts of a foreign language as the auditory cortex should not be considered an indispensible part of the language network.

MMN experiments have been conducted on various syllable types (compare Näätänen et al. 1997; Lipski 2006), but the considerable variety of affricate categories across the languages of the world (Ladefoged and Maddieson 1996; Gordon et al. 2002) calls for advanced research on this specific topic. Let us therefore briefly illustrate the affricate subsystems of the three languages involved in the present study: Serbian, Rhaeto-Romance and Swiss-German. Serbian differentiates four categories that are used in our experiment, namely [ʧ] (postalveolar, voiceless), [ʨ] (alveolo-palatal, voiceless), [ʤ] (postalveolar, voiced), and [ʥ] (alveolo-palatal, voiced); the fifth affricate [ts] (alveolar, voiceless) is not part of the experiment. For Rhaeto-Romance, we chose the most spoken dialect Sursilvan. It shares three of the four affricates with Serbian, namely [ʧ], [ʨ] and [ʥ] (Haiman and Beninca 1992). Sursilvan [ʧ] seems to share the typical lip rounding of Serbian (Morén 2006), but it lacks the voiced postalveolar affricate [ʤ] that exists in Serbian. For Swiss-German, we referred to the Zurich dialect, which contains four voiceless affricates, namely labial [pf], alveolar [ts], postalveolar [ʧ], and velar (sometimes uvular) [kx] (Fleischer and Schmid 2006). Thus, out of the manner of articulation we are interested in, Swiss-German only has [ʧ]. Considering that the four affricates in our study differ in voicing and place of articulation, it must be pointed out that Swiss-German speakers do not differentiate contrasts of the affricate category in either dimension.

2. METHODS

2.1. Subjects

Fifteen Swiss-German and fifteen Rhaeto-Romance speaking subjects participated. All subjects reported undisturbed speech and hearing capacities were right-handed and on average 23 years old. None of the participants had prior knowledge of any Slavic language nor were they professional musicians. Rhaeto-Romance speakers were generally bilingual; on average they learnt Swiss-German at the age of seven years – speaking it either regularly in school or with one of their parents. As the second language in the bilingual cases did not include an extra category of the phoneme category under investigation, the possible advantage is not relevant for the present investigation. Participants were asked to fill in a questionnaire to file their details and their language background. Influence of second language knowledge has been found on word recognition as well as on phoneme distinction abilities (compare Ventura et al. 2007; Pattamadilok et al. 2007).

2.2. Stimuli

The four Serbian syllables [ʨa], [ʥa], [ʤa], and [ʧa] served as experimental stimuli. Naturally spoken stimuli were chosen because it has been shown that natural material is processed more robustly than synthetic stimuli which are less immune against manipulations (Lacerda 2001). The speaker had to be a female due to more consistent inter-speaker variation (Titova and Näätänen 2001). For the recording we asked her to produce the phrases as naturally as possible to avoid over articulation. The recording was done twice. The usage of CV (consonant-vowel) syllables was motivated by the fact that isolated affricates, especially voiceless ones, resemble non-speech noise. Because the vowel [a] is universally unmarked, we decided to apply this vowel. In contrast to [u] and [o], [a] does not lead to anticipatory lip rounding during the production of the affricate and there is no co-articulatory influence of a palatal glide for [a].

All stimuli were digitally recorded in an anechoic chamber. A sampling rate of 44100 Hz and 16 bit quantization was used. Our native speaker of Serbian read the four syllables aloud in twelve variations each: three times each in a CV sequence, in a VCV sequence and in an existing Serbian word (*ćaskati* "to chat", *čarapa* "sock", *đavol* "devil", *džaba* "frog"). The syllables were pronounced inside a carrier phrase

(*Prvo ća, drugo ća, treće ća* - “first ća, second ća, third ća”). Acoustic analysis included manual measurement of duration, closure, release phase and duration of the syllables in all alternations on the basis of an inspection of wave forms and spectrograms provided by *Praat* (Boersma and Weenink 2009). The ‘Centre of Gravity’ (CoG) was calculated using the apposite function in *Praat*. The CoG represents the power-spectrum from the release burst to the voicing onset of the following vowel (compare Forrest et al. 1988; Gordon et al. 2002). Voicing clearly affected duration and spectral characteristics: the two voiceless affricates are longer and have higher values for the CoG than the voiced ones. Regarding the place of articulation, it results that the two alveolo-palatal affricates display a higher CoG, relatively shorter closure phase and a longer release phase than the two postalveolar affricates.

The four stimuli used in the experiment were selected according to the following criteria: Duration for affricate and vowel about 150 ms, even, constant fundamental frequency (F0) trend. Editing included stylizing the pitch using *Praat* 5045 (Boersma and Weenink 2009) and setting the overall intensity to 70 dB. Normalization was done using *Audition*[1]. This did not change the intensity relation between affricates and vowels in the individual syllables. A Butterworth filter was applied as low-pass filter (5000 Hz) to cut background- and click-sounds using *Audition.* At the onset and at the end of the syllables a smooth rising/falling ramp with duration of 10 ms was added (Gaussian filter). F0 was set to a constant value throughout the vowel with respect to initial F0 value. Duration was normalized by clipping the affricate onset and vowel offset so that each syllable had duration of between 120-185 ms. Finally, the vowel of the syllable [ʧa] was stabilized at a length of 92 ms and was used for all four syllables. The last step was done in full awareness of the loss of information that is provided by the specific transition of the affricate to the following vowel (compare Recasens and Espinosa 2007). After the final editing, three Serbian and three Rhaeto-Romance speakers were asked to judge the syllables for their ‘naturalness’ and their discriminability (e.g. Nenonen et al. 2005). Serbian speakers could reliably ascribe each syllable. Rhaeto-Romance speakers encountered increased difficulties, yet they clearly made out “three or more” different syllables.

2.3. Procedure

During the EEG experiment, subjects were seated in an electrically shielded and acoustically attenuated chamber. The data were recorded using a Biosemi active-two amplifier system. 64 active electrodes were installed according to the 10/20 electrode system (Jaspers 1958). The sampling rate was 512Hz. Impedance was kept below 40kΩ.[2] For off-line re-referencing, an electrode was attached to the tip of the nose. Vertical and horizontal eye movements were recorded by two bipolar channel pairs. For head and body movements, participants were monitored through a close-circuit camera system. The paradigm follows the idea of Näätänen et al. (2004)’s Optimal Paradigm. Three deviants were presented alongside the standard and not compared individually against the standard as in the classic oddball paradigm. We used an oddball paradigm with 50 percent standard (e.g. [ʧa]) and 50 percent deviant ([ʨa], [ʥa] and [ʤa]) proportion. This paradigm was chosen according to the Optimal-1-Paradigm of Näätänen (2004; compare Pakarinen et al. 2007). Furthermore, we used a Multiple-Deviant Paradigm which means that every deviant once acted as the standard. The Inter-stimulus Interval (ISI) was set to 750 ms. An additional Stimulus Onset Asynchrony (SOA) of 400 ms was jittered. A rapid and unpredictable rate of stimulus presentation was chosen to divert possible attention processes (compare Sinkkonen and Tervaniemi 2000; Muller-Gass et al. 2006). The first two minutes were recorded for closed and open eye-movements (resting EEG). Thereafter, eight passive listening blocks followed. Block sequences were randomized between subjects. Participants were asked to read an unrelated text and not pay attention to the syllables they heard. At the beginning of each block there were 15 repetitions of the standard to attune the subjects’ ears to the respective standard. Therefore a stimulus block included 150 deviants and 165 standards. In total 1200 deviant repetitions and 1320 standard repetitions were used, whereof each of the four stimuli appeared 330 times as a standard and 300 times as a deviant. On average, participants wished to recess for three minutes between blocks.

2.4. Data Analysis

The data was analyzed using *BrainVision Analyzer 1.05.0005*[3] and *eegLab 6.01*[4] (Matlab). EEGs were offline treated with a 24 dB zero-phase bandpass-filter from 0.1 to 30 Hz. To correct for artefacts, a manual and an automatic *Raw DataInspector* was applied, whereby changes exceeding 150µV at any channel were marked and neglected for further analysis. Unfortunately, we could not use the nose as reference, as the coordinates of this electrode are unknown to the *eegLab system.* Common average Reference (CAR) was therefore applied as reference. Eye blinks and horizontal movements were corrected by means of independent component analysis (ICA) (Stone, 2002) implemented in *eegLAB.* Due to technical problems while recording, six subjects (3 Rhaeto-Romance and 3 Swiss-German) had to be discarded. EEG recordings were segmented into 600-ms epochs (100 ms pre- and 500 ms post-stimulus) and averaged for each stimulus type separately with 100 ms pre-stimulus as a baseline. ERPs for all stimuli (each stimulus type as a standard and as a deviant) were averaged for each subject and across subjects. MMN difference waves were computed by subtracting ERPs (event related potentials) to the standard from ERPs to the deviant of a chosen stimulus and averaged. Being able to directly compare the response to a certain stimulus acting both as a standard and as a deviant is one of the main advantages of the MMN paradigm we applied (compare also Grimm et al. 2008). Peak-detection was carried out over a time-window of 180 ms (120-300 ms after stimulus onset). The presence of the MMN was statistically verified using analysis of variance, one-sample and independent *t*-tests with *SPSS*[5] at a significance level of 0.05. Analysis involved evaluation of factors group, stimuli, peaks and latencies. To verify the existence of a true MMN component, activations at Fz were compared with supra-temporal electrodes (TP7 and TP8; compare Näätänen et al. 2007).

3. RESULTS

Deviant-related MMN potentials were measured by subtracting ERPs elicited by the stimulus operating as a standard sound from ERPs elicited by the same stimulus operating as a deviant sound. This allowed a direct comparison of the physically identical stimulus differing only in its probability of occurrence. The Acoustic Event Related Potential (AERP) for both groups showed orderly N1 and P2 components at central Cz electrode. All MMN curves displayed a typical fronto-central maximum (Fz) with a polarity inversion at the mastoid leads (TP7 and TP8). With respect to the statistical evaluation of the MMN, a repeated-measures ANOVA was performed for peaks and latencies separately. Normal distribution was assured with a *Kolmogorov Smirnov* test. The ANOVA included the between-subject factor "Group" (Rhaeto-Romance vs. Swiss-German) and the within-subject factors "Stimulus" ([ʧa], [ʨa], [ʥa] and [ʤa]) and "Peak" or "Latency" (three peak or latency values per stimulus, representing the three deviant conditions). All main effects or interactions with two or more degrees of freedom in the numerator were adjusted with the procedure suggested by Huynh and Feldt (1970) and revealed a main effect of Group ($p < 0.05$, Greenhouse-Geisser $p = 0.010$) for the MMN amplitude. Our main hypothesis was therefore successfully confirmed. As expected, the comparison Group * Stimulus * Latency revealed no main effect Group. Due to slightly different stimulus length (up to 65 ms difference), a systematic latency effect was anticipated. Tests of Within-Subjects Effects indicated a significant main effect Stimulus (Greenhouse-Geisser $p = 0.013$). The one-sample *t*-Tests in both groups for the comparison of both peaks and latencies were all significant at the $p < 0.001$ level. Independent Samples *t*-Tests provide evidence that Rhaeto-Romance speakers process phonological contrasts significantly differently. Surprisingly, the stimulus [ʨa] elicited no significant group difference. Stimulus [ʥa] was processed significantly different if it served as a deviant beside the standard [tʃa] ($p = 0.01$) and if [ʤa] served as standard ($p = 0.03$) compared to acting itself as a standard. Stimulus [ʤa] also displayed significant differences between Rhaeto-Romance and Swiss-German speakers when serving as a deviant in standard blocks [ʨa] and [tʃa]. As expected, no differences in processing were found for the stimulus [tʃa]. This stimulus is common to speakers of both language groups and should therefore not evoke a significant difference in processing. MMN reliability was inspected by comparing the amplitudes of the MMN component at the frontal Fz electrode with the zero level. For both language groups in all deviant conditions, negative peaks were observed in the deviant-minus-standard difference waves.

4. DISCUSSION

The overall goals of the MMN experiment were two-fold: first and foremost, to examine the implications of the different language-backgrounds of the two groups, and second to test whether place of articulation or voicing had a stronger influence on the perception of a foreign language phonological contrast. A significant main effect Group confirmed a general difference as a function of language processing. The direction of the effect, however, was unexpected. There are two possible explanations for this finding. On the one hand, Swiss-German speakers might previously have attained cortical representations of the foreign sound category that enabled them to perform in a comparable way to the Rhaeto-Romance speakers. This would confirm the assumption that linguistic experience affects the critical time window for speech acquisition (compare Gandour et al., 2007). On the other hand, overlearning could have yielded smaller responses to the phonetic contrast in the Rhaeto-Romance group. The higher amplitudes in the Swiss-German subjects could be interpreted as increased neural activity reflecting the processing of unknown information (compare Tervaniemi et al. 2000; van Zuijen et al. 2005; Kujala et al. 2007).

The notion of establishing memory traces is cuncures with the idea of neural commitment proposed by Kuhl (2004): one assumes a reference between the input signal and stored regularities that have been established during language acquisition. The SLM by Flege (1993) and the PAM by Best (2001) both accentuate the automatic assimilation through which non-native sounds are mapped onto the nearest native speech sound representation. Here it seemed that having an equivalent phoneme in one's native language caused a smaller MMN amplitude for the processing of affricate sounds than having no equivalent phoneme in the L1. However, in both cases the ability to discriminate and categorize was not eliminated. Our results support the notion that phonetic information that seems irrelevant to the acquired L1-specific representations is not completely neglected or filtered out (compare e.g. Zhang et al. 2005). This strongly speaks in favour of the continuous ability to learn foreign language phonemes that are similar or dissimilar to the L1 phonetic category in adulthood.

This study contributed to research in language learning with the observation that language experience in auditory processing is also involved in the processing of affricates. A possible ambiguity of the mode of perception of these sounds might be in line with Lipski's findings (2006) who detected a less categorical perception for fricative sounds. Given the great variety of affricate categories across languages (Ladefoged and Maddieson 1996; Gordon et al. 2002) further research on this topic is needed. For instance, only coronal affricates were investigated in this study; the way dorsal affricates are processed and distinguished could be a future project. Moreover, most studies investigate the 'products' of language learning, whereas in order to make statements about the 'process' of language acquisition, longitudinal studies would be fruitful.

5. REFERENCES

Best, C., McRoberts, G. W., Goodell, E. 2001. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener´s native phonological system. *Acoustical Society of America* 109(2). 775-794.

Boersma, P., Weenink, D. 2009. Praat: doing phonetics by computer. < http://praat.org>

Corbett, G. 1987. Serbo-Croat. In: Comrie, B. (ed.). *The world's major languages*. London/New York: Routledge. 391-409.

Flege, J.E. 1993. Production and perception of a novel, second-language phonetic contrast. *Journal of the Acoustic Society of America* 93. 1589-1608.

Fleischer, J., Schmid, S. 2006. Zurich German. *Journal of the International Phonetic Association* 6. 243-253.

Forrest, K., Weismer, G., Milenkovic, P., Dougall, R. N. 1988. Statistical analysis of word-initial obstruents: preliminary data. *Journal of the Acoustical Society of America* 84. 115-123.

Friederici, A.D. 2005. Neurophysiological markers of early language acquisition: from syllables to sentences. *Trends in Cognitive Sciences* 9(10). 481-485.

Gandour, J., Tong, Y., Talavage, T., Wong, D., Dzemidzic, M., Xu, Y. & Lowe, M. (2007). Neural Basis of First and Second Language Processing of Sentence-Level Linguistic Prosody. *Human Brain Mapping* 28. 94-108.

Gordon, M., Barthmaier, P., Sands, K. 2002. A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association* 32(2). 141-174.

Grimm, S., Schröger, E., Bendixen, A., Bäss, P., Roye, A., Deouell, L. Y. 2008. Optimizing the auditory distraction paradigm: Behavioral and event-related potential effects in a lateralized multi-deviant approach. *Clinical Neurophysiology* 119. 934-947.
Haiman, J., Beninca, P. 1992. *The Rhaetoromance Languages*. Routledge, New York.
Huynh, H., Feldt, L.S. 1970. Conditions under which the mean square ratios in repeated measurement designs have exact F distributions. *Journal of the American Statistical Association* 65. 1582-1589.
Jaspers, K. 1958. [*The physician in the technical age*]. Klin.Wochenschrift 36. 1037-1043.
Kuhl, P. 2004. Early Language Acquisition: Cracking the Speech Code. *Nature Reviews* 5. 831-843.
Kujala, T., Tervaniemi, M., Schröger, E. 2007. The mismatch negativity in cognitive and clinical neuroscience: Theoretical and methodological considerations. *Biological Psychology* 74(1). 1-19.
Lacerda, F. 2001. Re-assessing the perceptual consequences of CV-transition speed. *Lund Linguistic Working Papers* 49. 98-101.
Ladefoged, P., Maddieson, I. 1996. *The Sound of the World's Languages*. Blackwell Publishers.
Lenneberg, E. 1967. *Biological foundations of language*. New York: Wiley.
Lipski, S. C. 2006. Neural correlates of fricative contrasts across language boundaries. *Institut für Maschinelle Sprachverarbeitung der Universität Stuttgart.*
Morén, B. 2006. Consonant-vowel interactions in Serbian: Features, representations and constraint interactions. *Lingua* 116. 1198-1244.
Muller-Gass, A., Stelmack, R. M., Campbell, K. B. 2006. The effect of visual task difficulty and attentional direction on the detecton of acoustic change as indexed by the Mismatch Negativity. *Brain Research* 1078. 112-130.
Näätänen, R. 1992. *Attention and brain function.* Hillsdale, NJ: Lawrence Erlbaum.
Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J., Luuk, A., Allik, J., Sinkkonen, J., Alho, K. 1997. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385. 432-434.
Näätänen, R., Winkler, I. 1999. The concept of auditory stimulus representation in Cognitive Neuroscience. *Psychological Bulletin* 125. 826-859.
Näätänen, R. 2001. The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology* 38. 1-21.
Näätänen, R., Pakarinen, S., Rinne, T., Takegata, R. 2004. The mismatch negativity (MMN): towards the optimal paradigm. *Clinical Neurophysiology* 115. 140-144.
Näätänen, R., Paavilainen, P., Rinne, T., Alho, K., 2007. The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology* 118. 2544-2590.
Nenonen, S., Shestakova, A., Huotilainen, M., Näätänen, R. 2005. Speech-sound duration processing in a second language is specific to phonetic categories. *Brain and Language* 92. 26-32.
Pattamadilok, C., Kolinsky, R., Ventura, P., Radeau, M., Morais, J. 2007. Orthographic Representations in Spoken Word Priming: No Early Automatic Activation. *Language and Speech* 50(4). 505-531.
Recasens, D., Espinosa, A. 2007. An electropalatographic and acoustic study of affricates and fricatives in two Catalan dialects. *Journal of the International Phonetic Association* 37(2). 143-172.
Sinkkonen, J., Tervaniemi, M. 2000. Towards Optimal Recording and Analysis of the Mismatch Negativity. *Audiol Neurootol* 5. 235-246.
Stone, J. V. 2002. Independent component analysis: an introduction. *Trends in Cognitive Sciences* 6(2). 59-64.
Tervaniemi, M. T. Ilvonen, J. Sinkkonen, A. Kujala, K. Alho, M. Huotilainen, R. Näätänen 2000. Harmonic partials facilitate pitch discrimination in humans: electrophysiological and behavioral evidence. *Neuroscience Letters* 279. 29-32.
Titova, N., Näätänen, R. 2001. Preattentive voice discrimination by the human brain as indexed by the mismatch negativity. *Neuroscience Letters* 308. 63-65.
Ventura, P, Kolinsky, R., Fernandes, S., Querido, L., Morais, J. 2007. Lexical restructuring in the absence of literacy. *Cognition* 105(2). 334-361.
Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., Czigler, I., Scepe, V., Ilmoniemi, R. J., Näätänen, R. 1999. Brain responses reveal the learning of foreign language phonemes. *Psychophysiology* 36. 638-642.
Zhang, Y., Kuhl, P., Imada, T., Kotani, M., Tohkura, Y. 2005. Effects of language experience: Neural commitment to language-specific auditory patterns. *NeuroImage* 26. 703-720.
van Zuijen, T. L., Sussman, E., Winkler, I., Näätänen, R., Tervaniemi, M. 2005. Auditory organization of sound sequences by a temporal or numerical regularity—a mismatch negativity study comparing musicians and non-musicians. *Cognitive Brain Research* 23(2-3). 270-276.

NOTES

[1] http://www.adobe.com/products/audition/

[2] http://www.biosemi.com/faq/shielding%20vs%20active%20electrodes.htm

[3] http://www.brainproducts.com/downloads.php?kid=1

[4] http://sccn.ucsd.edu/eeglab/

[5] http://www.spss.com/downloads/Papers.cfm?ProductID=00035&Name=SPSS_Base&DLType=Demo

Preferences in L2-Vowel Perception

Nadja Kerschhofer-Puhalo

University of Vienna, Austria
nadja.kerschhofer@gmx.at

ABSTRACT

Numerous studies have shown asymmetries in vowel perception, in which some contrasts seem to be better perceived than others. Some of these asymmetries seem to be language-universal. Concepts like phonetic salience, perceptual weakness, and universal markedness have often been used to account for them. A cross-language categorization test was carried out with L2 learners of German from 10 different native languages. The results show that, cross-linguistically, some vowel qualities are more often affected by perceptual assimilations than others. Front rounded and long mid vowels seem to be more prone for misperception and substitution processes than other vowels, while peripheral vowels are perceived more accurately and are used more often as perceptual targets and substitutes. The behaviour of front rounded vowels in perception will be discussed in detail.

Keywords: perception – (front rounded) vowels – implicational hierarchies – markedness

1. INTRODUCTION

In functional approaches to phonology, there are two main forces shaping phonology: the minimization of effort in articulation and the maximization of phonological contrast, i.e. minimization of confusion in perception. Second language speech models consider perception as crucial for L2 speech acquisition (Flege 1995; Kuhl/Iverson 1995; Kuhl 1998; Best 1995). The perception of sound categories is not universal but is influenced by the phonological system of a listener's L1. The variance in L2 performance is generally predicted or explained by the individual´s language background (L1) and his or her exposure to L2 speech input. A comparative approach would describe phonemic and phonetic characteristics of L1 and L2 to predict or explain problems in L2 perception. However, a contrastive analysis cannot account for all difficulties in L2. In particular, the reason why some L2 sounds seem to be more difficult to perceive than others cannot solely be explained by a contrastive account.

Second language speech models usually posit some kind of assimilation of non-native sounds to existing native categories, mostly referring to some notion of phonetic or phonological "similarity" of L1 and L2. But is this perceptual assimilation only based on the distribution of categories in L1 and L2 or are there other, more basic notions to be considered? This paper will describe learners´ difficulties and preferences in L2 vowel perception and will consider possible explanations for learners´ strategies, which may be influenced by L1 as well as by *language-universal* perceptual preferences.

2. ASYMMETRIES IN VOWEL PERCEPTION

Many studies on speech perception show a certain sensitivity to phonetic differences within a category, i.e. the ability to discern „good" from „less good" exemplars of a category (cf. Kuhl 1998). This indicates that perceptual discrimination ability includes phonological categories as well as phonetic differences (Best/Hallé/Bohn/Faber 2003). Best and Tyler (2007) suppose that non-native speech perception is not confined to differences relevant to native phonological contrasts. Some aspects of sensitivity to phonetic variation seem to be related to similarities between non-native stimuli and native speech patterns, while others may reflect language-universal tendencies of perception.

In fact, some asymmetries in perception have been observed in native and non-native listeners, infants and adults, suggesting the influence of *universal* rather than experience-based factors. Several studies suggest that perceptual asymmetries might exist because of differences in *perceptual salience* (Kaun 2004; Mielke 2002).

Polka and Bohn´s (2003) Peripherality Hypothesis postulates language-universal biases towards more peripheral referent vowels in contrast pairs. In their study with infants, peripheral vowels were better perceived than less peripheral ones. The higher perceptual salience and stability of more peripheral vowels may facilitate the formation of language specific vowel categories. They seemed to serve as reference vowels, exerting a kind of "magnet effect". The less peripheral member of a contrast is "assimilated" to the reference vowel, i.e. perceived as "a less good exemplar" of the more peripheral vowel. Polka and Bohn (2003) assume that language-specific tuning may emerge developmentally from the periphery of the vowel space to the interior. In this sense, (language-specific) phonology exerts an influence on perception and perception is influenced by language-universal biases, which in turn explains perceptual asymmetries. These assumptions may also hold true for vowel perception in L2.

3. DIFFICULTIES AND PREFERENCES

Single speech sounds are not difficult to perceive as such. While articulatory difficulty can be regarded as a property of an individual sound (in a particular context), perceptual difficulty is only definable in terms of the contrasts a sound enters into. The difficulty lies in the correct categorization onto the set of contrasting categories of a given language. The perceptual difficulty of sounds is postulated to be equivalent with *perceptual markedness*, which depends on the salience and stability of contrasts the sounds are involved in (Lindblom 1986; Flemming 2004). Less confusable and more stable unmarked contrasts are preferred over more confusable, marked ones. The concept of perceptual markedness also explains why some L2 vowels are more difficult to categorize, especially if there is no equivalent category in L1.

Perceptibility scales and *implicational hierarchies* can be deduced from this notion of perceptual markedness. Along an implicational hierarchy, the phonetic salience of sounds involved decreases with consequences on phonemic contrastivity, the preference and frequency of cross-language occurrence decreases and the perceptual difficulty increases, especially if the marked element is not present in L2.

In the UPSID vowel inventories (Maddieson 1984), there is a universal preference for peripheral vowels. *Peripheral vowels* are the front unrounded, back rounded, and low vowels which lie at the margins of the available vowel space such as /a, i, u, ɛ, e, o, ɔ/. The acoustic quality of vowels in these regions is more stable. These vowels are considered to be *unmarked*, whereas vowels in acoustically instable regions are more marked. Across languages, unmarked vowel qualities occur more frequently than marked qualities.

Universally, the three corner vowels /i, u, a/ are the least marked and most frequent vowel qualities occurring in most of the languages (91,5% of the systems use /i/, in 88% /a/, 83,9%/u/). As a set, they constitute a "typical", "complete" vowel inventory. Acoustically, they are maximally distinct, the easiest to perceive and the most resistant to misperception. In terms of their production, they all show quantal effects (Stevens 1989), i.e. changes in articulation do not produce correspondingly large acoustic changes. Their acoustic quality is more or less consistent along a wide range of articulations. Though they occupy areas of acoustic stability, they are surrounded by areas of instability. For all these reasons /i, u, a/ can be considered to be universally preferred perceptually unmarked vowel qualities (Maddieson 1984; Lindblom 1986).

Another example of relative perceptual markedness is the *round-nonround contrast in front vowels*. Front nonround vowels are peripheral, being at the margins of the vowel space, whereas front rounded vowels are considered to be non-peripheral. Cross-linguistically, there is a preference for back rounded (93,5%) and front unrounded (94%) vowels (Maddieson 1984). The perceptual explanation for this pattern is that the co-varying of backness and rounding maximises the difference in F2 frequency between front and back vowels. The contrast between front unrounded /i/ and front rounded /y/ is less distinct due to lower F2 because of lip-rounding. The contrast /i/-/y/ is therefore less preferred; /y/ is considered to be perceptually relatively marked.

Contrasts within the class of front rounded vowels like /y/ vs. /ø/ are perceptually even less salient due to the decrease of labiality and lower F2 values for lower rounded vowels (Terbeek 1977; Kaun 2004). Usually, higher vowels are articulatorily more rounded than lower vowels (Linker 1982, though there are some exceptions, see Ladefoged/Maddieson 1990: 99f.). This seems to be consistent with the acoustic patterning associated with rounded vowels. Lip rounding in high vowels has acoustically more dramatic consequences than in mid vowels (Stevens 1998: 293f.). Consequently, these articulatory and acoustic differences give rise to perceptual differences among rounded vowels as well. Terbeek's (1977) results on perceptual distances indicate that nonhigh vowels and front vowels are perceived to be relatively less rounded along the round-nonround continuum. Back round vowels such as /u/ and /o/ lie on the higher end of the scale relative to front vowels, higher vowels being relatively nearer to the higher end of the scale than non-high vowels; the front non-high [ø]-type vowel is perceived as least rounded (u > o >.y > ø). In this sense, /ø/ and /œ/ are considered to be perceptually weaker and more marked than higher front rounded vowels like /y/ and /ʏ/. Front unrounded, peripheral vowels like /i/ and /e/ are considered to be perceptually the most salient and the least marked front qualities. The decrease in perceptual salience from higher to lower front rounded vowels is also consistent with frequency and distribution of vowels in the UPSID inventories, where /y/ is the most frequent of the front rounded vowels.

4. PREFERENCES IN VOWEL INVENTORIES

One might assume that smaller vowel inventories contain more common sounds while larger inventories include less frequent sounds. In fact, some contrasts only occur if the number of vowels exceeds a certain number. E.g. the probability of using a secondary feature like length increases with the number of vowel quality contrasts. 53,8% of languages with 10 or more vowel qualities use length distinctions (Ladefoged/Maddieson 1990). A single unified hierarchy of frequency for segment types in inventories cannot be sustained, but there are some *implicational hierarchies* between particular types of segments (Maddieson 1984: 10ff):

- Mid vowels do not occur unless high and low vowels occur (2 exceptions in UPSID).
- Rounded front vowels only occur if unrounded front vowels of the same height occur (2 exceptions).
- /ø/ and or /œ/ only occur if /y/ also occurs (2-3 exceptions).

Note, that these implications are consistent with the facts mentioned above.

German, the target language in the study reported here, has a relatively large vowel inventory. In UPSID, German and Norwegian show the largest number of contrasting qualities, each having 15 vowels. Apart from the more preferred vowels /iː, ɪ, eː, ɛ, (ɛː), a, ɑː, oː, ɔ, uː, ʊ/, German has a series of (universally marked) front rounded vowels /y, ʏ, ø, œ/. In addition to quality, German uses quantity contrasts, i.e. length. The vowels in a short-long-contrast differ in quality, e.g. long /iː/ vs. short /ɪ/ or /eː/ vs. /ɛ/. This is consistent with Maddieson´s observation that height seems to be associated with length and that higher mid vowels are more likely to be long than lower mid vowels. Mid vowels seem to be raised when lengthened and/or lowered when shortened (Maddieson 1984: 129f).

To summarize:

1. Peripheral vowels are universally preferred. There is a special preference for the corner vowels /i, u, a/.
2. Back rounded vowels are preferred over front rounded vowels.
3. Front unrounded vowels are more preferred than front rounded vowels.
4. Front rounded vowels are more frequent than back unrounded.
5. Among front rounded vowels, higher vowels are more preferred than non-high vowels, lower front rounded vowels are least preferred.
6. Long higher-mid vowels seem to be preferred over low long and high long vowels.

So what happens when L2 learners with a comparatively smaller L1 vowel inventory acquire a second language with a larger inventory? Are universally marked L2 sounds discriminated and acquired less easily than more common sounds? And are universally unmarked sounds preferred as candidates for perceptual substitution or assimilation in L2 vowel categorization? To answer these questions, a cross-language perception test was carried out.

5. THE STUDY

The test consisted of a perceptual identification task. A native German speaking male (Austrian Standard German) was recorded producing multiple tokens of the 15 German vowel phonemes (/a, aː, ɛ, eː, ɪ, iː, ɔ, oː, ʊ, uː, œ, øː, ʏ, yː, ɛː/), each in varying consonantal contexts. The structure of the produced nonsense stimuli words were /pVC/ and /CVtə/ embedded in carrier sentences. The material contained no diphthongs. Every vowel occurred 15 times in varying pre- and postvocalic context. In total, the test material consisted of 270 test items. An acoustic analysis of the input data was carried out (F1-F4 and vowel duration).

173 adults participated in the test as unpaid volunteers. The sample consisted of learners at different levels and with differing length of residence in a German speaking country. All of them had normal hearing, speech, language, and reading abilities. Most of the subjects were undergraduate students. The mean age of the subjects was 24,7 years. Their first languages were Polish, Hungarian, Turkish, Arabic, Albanian, Romanian, English, Farsi, Mandarin, and SerBoCroatian. The sample described here consists of 46.710 responses (270 stimuli x 173 participants). Additionally, a control group of 18 L1 German adults was tested. The participants were told to listen to the nonsense-words embedded in a constant carrier-sentence and to identify the vowel contained in each of the nonsense-words. They were then asked to identify this vowel and to mark it on a form containing a table with all German vowel phonemes and 3 diphthongs.

1. RESULTS AND DISCUSSION

The present paper will only discuss a subsample of the study, namely the behaviour of front rounded vowels and their categorizations by Albanian, Polish, Romanian, SerboCroatian, and Turkish subjects. The languages presented here use vowel systems with 5 to 8 vowels: SerBoCroatian (5) /i, ɛ, a, ɔ , u/; Albanian (5-6): /i, y, e, (ə), a, o, u/; Polish (6): /i, ɨ, ɛ, a, ɔ, u/; Romanian (7) /i, e, ɨ, ə, a, o, u/; Turkish (8): /i, y, e, œ, ɯ, a, o, u/. For a complete discussion based on a contrastive analysis of the vowel systems involved, see Kerschhofer-Puhalo (in preparation).

Figure 1: Percentage of correct and wrong categorization (173 subjects, 10 L1s)

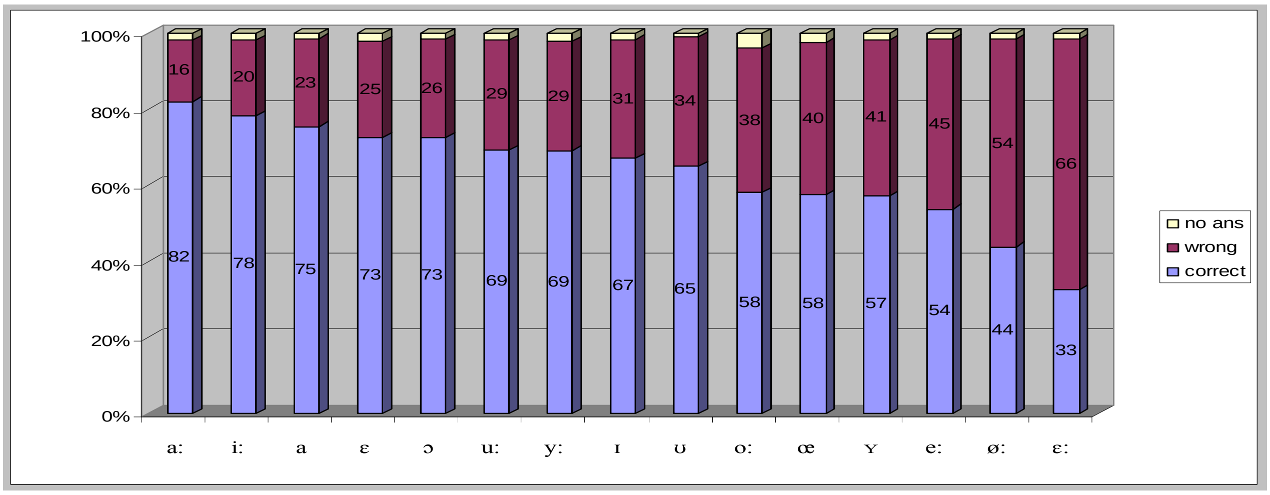


Figure 1 shows the total percentage of correct vs. wrong categorization by all 173 subjects. /aː, iː, a, ɛ, ɛ, uː, yː/ are the vowel categories, which are perceived the most correctly. It can be observed that the corner vowels /i/ and /a/ are discriminated best, while /uː, ʊ/ is not perceived to the same extent. The correct identification of front rounded vowels and of high-mid long vowels /eː, oː/ seems to be more difficult. /ɛː/ is the category causing most problems, with non-native as with native subjects, due to its marginal status. Among the front rounded vowels, /yː/ is most often identified correctly (69,1%), followed by /œ/ (57,6%) and /ʏ/ (57,3%). Long /øː/ is the least correctly perceived quality (43,8%). Although the large number of correct identifications for /yː/ is consistent with our predictions, the number of correct answers for short /œ/ is surprisingly low (see contradictory results in Kerschhofer 1998). The data show considerable confusion among the non-high front rounded qualities /ʏ, ø, œ/. The most preferred answering option among front rounded vowels is /y:/ followed by /ʏ/ and /œ/, while nonhigh long /ø:/ is the least common option.

Table 1: Confusion matrix for German vowel categories in percent by 13 Albanian, 31 Polish, 12 Romanian, 33 SerBoCroation, and 24 Turkish subjects

y:	***a***	***a:***	***ä:***	***e***	***e:***	***i***	***i:***	***o***	***o:***	***u***	***u:***	***ö***	***ö:***	***ü***	***ü:***	***ei***	***eu***	***au***	***wrong***
Alb							0,5			0,5	0,5	0,9	4,2	16,7	<u>70,8</u>				23,1
Pol			0,4	0,4	1,1	2,0	7,5		0,4	3,8	5,6	3,0	5,9	18,8	<u>47,3</u>	0,2	0,9		49,8
Rom										0,5	0,5		1,4	19,0	<u>78,7</u>				21,3
SBC			0,5	0,3			0,7			1,5	2,4	1,2	2,7	17,7	<u>72,2</u>		0,3		27,3
Turk		0,2			0,2						1,2		1,6	6,9	<u>85,0</u>		0,2		10,4
ʏ	***a***	***a:***	***ä:***	***e***	***e:***	***i***	***i:***	***o***	***o:***	***u***	***u:***	***ö***	***ö:***	***ü***	***ü:***	***ei***	***eu***	***au***	***wrong***
Alb					0,5	0,5	0,5		0,9	0,5		17,6	7,9	<u>44,4</u>	20,8	0,9			50,0
Pol		0,2	0,2		2,0	3,0	2,3	0,2	0,4	10,8	6,1	9,9	5,7	<u>38,2</u>	18,3	0,4	0,5	0,4	60,2
Rom				0,5						3,7	1,9	10,2	2,8	<u>68,5</u>	12,0		0,5		31,5
SBC					0,2	0,8		0,3		10,8	1,9	6,2	1,0	<u>65,3</u>	12,5		0,2		33,8
Turk						0,2				0,7	0,7	5,8	4,6	<u>59,3</u>	25,7				37,7
ø:	***a***	***a:***	***ä:***	***e***	***e:***	***i***	***i:***	***o***	***o:***	***u***	***u:***	***ö***	***ö:***	***ü***	***ü:***	***ei***	***eu***	***au***	***wrong***
Alb			0,9					0,9	2,8		0,5	13,4	<u>55,1</u>	4,6	16,2				39,4
Pol			0,2		1,4	0,5	1,1		0,9	3,9	13,1	6,5	<u>15,9</u>	14,5	40,9		0,2	0,2	83,3
Rom							0,5			1,4	2,3	9,3	<u>54,6</u>	7,4	23,1		0,9	0,5	45,4
SBC			0,2					0,5	1,2	0,8	5,2	5,9	<u>31,6</u>	12,0	42,1				67,8
Turk										0,2	3,2	1,4	<u>53,4</u>	4,6	30,9		1,9	0,5	42,7
œ	***a***	***a:***	***ä:***	***e***	***e:***	***i***	***i:***	***o***	***o:***	***u***	***u:***	***ö***	***ö:***	***ü***	***ü:***	***ei***	***eu***	***au***	***wrong***
Alb			3,7		0,5			1,9	0,9		1,4	<u>57,9</u>	25,9	1,4	1,4				37,0
Pol		0,5	7,9	20,0	12,2	0,5	1,3	1,6	1,3	1,4	0,7	<u>20,7</u>	8,3	12,6	5,9	0,9	2,5		77,7
Rom	0,5		3,2	1,4			0,5	0,5	0,5			<u>78,2</u>	11,6				2,3		20,4
SBC			0,7	2,4	1,7		0,2	4,5	1,5	1,5	0,5	<u>64,5</u>	11,1	7,4	2,4				33,8
Turk				0,2				0,2				<u>63,2</u>	30,1	0,2			1,6	0,2	32,6

For long /y:/, all the languages selected show a lack of quantitative discrimination, resulting in identification with /ʏ/. Only Polish shows a delabialization effect with a 7,5% identification of /y:/ as /i:/. Turkish shows the highest correct identification rate for /y:/.

While /ʏ/ is confused with the members of the whole class of front rounded vowels by Albanian, Polish, and Romanian subjects, Turkish subjects show this behaviour to a lesser degree. Their main problem is in the discrimination of the length contrast between /y:/ and /ʏ/. Additionally, SerBoCroatian and Polish subjects substitute /ʏ/ with short /ʊ/; Polish also uses front qualities like /e:, ɪ, i:/ as substitutes.

For /ø:/, there is considerable confusion among the class of front rounded vowels for all languages described. A "perceptual magnet" or raising effect of /ø:/ towards long /y:/ is especially frequent with Polish, SerBoCroatian and Turkish subjects and to a lesser degree with Albanian and Romanian subjects. Especially interesting is this confusion among Turkish subjects, since Turkish has both /y/ and /ø/ qualities. (For perceptual weakness of nonhigh front rounded vowels in Turkish see Kaun 2004). Surprisingly, with Polish subjects, we do not find a significant assimilation rate of long /ø:/ with front qualities. Rather, length seems to be the decisive feature for identification. This may be the reason why Polish subjects also use long /u:/ as a substitute.

It seems that it is rather easy for all subjects to discriminate short low /œ/ from higher /ʏ/ and /y/-qualities. Only Polish and SerBoCroatian subjects substitute /œ/ with higher vowels. Insufficient distinction of /ø:/ and /œ/ is the most common problem with /œ/ across all languages. Additionally, Polish subjects also use mid front unround qualities /ɛ, e:, ɛ:/ as substitutes. The identification of /œ/ with /ɛ:/ = <ä:> by all subjects apart from the Turkish might be due to the orthographical and perceptual similarity of the categories involved. In addition, as was the case for short /ʏ/, SerBoCroatian subjects substitute /œ/ with /o/-qualities, and to a lesser extent with /ɛ/.

To summarize, the data confirms most of the predictions made above. Among front rounded vowels, nonhigh front rounded vowels are universally disfavoured, L2 category formation for these vowels seems to be more difficult. Raising to higher and more-peripheral vowels is a common strategy in perception.

6. SUMMARY

According to the articulatory, acoustic, and perceptual data collected, perceptual asymmetry effects are caused by universal perceptual markedness. In a cross-language categorization experiment with German vowels, wrong categorizations occurred more often with universally disfavoured vowels. The test results indicate that, cross-linguistically, the corner vowels /a/, /i/, and /u/ cause fewer problems in perception than other vowels. The confusion rate among front rounded vowels and long mid vowels was considerably high. Within the class of front rounded vowels, German /y:/ is the perceptually clearest category, while other nonhigh rounded vowels showed a higher confusion rate. This fact can be attributed to the greater salience of higher rounded vowels. This facilitates category formation in L2. Although there is considerable variance across languages, some common strategies such as the raising of lower qualities towards more peripheral and longer ones were observed. The choice of substitutes for perceptually more marked vowel qualities appears to be conditioned by the listeners´ L1 as well as by language-universal preferences.

Acknowledgements
This study was sponsored by a grant of the University of Vienna, Austria (F-98-G). The author wishes to thank Kristin Dill, Wolfgang U. Dressler, Katarzyna Dziubalska-Kolaczyk, Sylvia Moosmüller, Friedrich Neubarth, and all the participants of the study.

7. REFERENCES

Best, C. T. 1995. "A direct realist perspective on cross-language speech perception". In: Strange, W. (ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press.. 171-204.

Best, C. T., Tyler M. D. 2007. "Nonnative and second-language speech perception: Commonalities and complementarities". In: Bohn, O.-S., Munro, M.J. (eds.). *Language experience in second language speech learning: In honor of James Emil Flege*. Philadelphia: John Benjamins Publishing Company. 13-34.

Disner, S. F. 1983. "Vowel quality: the relation between universal and language specific factors". *UCLA WPP* 58. 1-158.

Flege, J.E. 1995. "Second language speech learning: Theory, findings, and problems", In: Strange, W. (ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press. 233-277.

Flemming, E. 2004. "Contrast and perceptual distinctiveness", In: Hayes, B., Kirchner, R., Steriade, D. (eds.), *Phonetically Based Phonology*. Cambridge: Cambridge University Press. 232-276.

Hume, E., Johnson, K. 2001b. "A Model of the Interplay of Speech Perception and Phonology". In: Hume, E., Johnson K. (eds.), *Studies on the Interplay of Speech Perception and Phonology*. Working Papers in Linguistics 55. 1-22.

Kaun, A. 2004. "The typology of rounding harmony". In: Hayes, B., Kirchner, R., Steriade, D. (eds.). *Phonetically Based Phonology*. Cambridge: Cambridge University Press. 87-116.

Kerschhofer-Puhalo, N. 1998. "The acquisition of German vowels by Polish learners", In: James, A., Leather, J. (eds.). *New Sounds 97*. 167-175.

Kerschhofer-Puhalo, N. (in preparation). *Vowel perception in second language acquisition*. PhD dissertation, University of Vienna.

Kohler, K. 1990. "Segmental reduction in connected speech: Phonological facts and phonetic explanations". In: Hardcastle, W.J., Marchal, A. (eds.), *Speech Production and Speech Modelling*. Dordrecht: Kluwer Academic Publishers, 69-92.

Kuhl, P. 1998. "Effects of language experience on speech perception". *Journal of the Acoustical Society of America* 103. 2931.

Kuhl, P., Iverson, P. 1995. "Linguistic experience and the ´perceptual magnet effect´". In: Strange, W. (ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press. 121-154.

Ladefoged, P., Maddieson, I. 1990. "Vowels of the world’s languages". *JPhon* 18. 93–122.

Lindblom, B. 1986. "Phonetic universals in vowel systems". In: Ohala, J., Jaeger, J. (eds). *Experimental phonology*. Orlando: Academic Press. 13-44.

Linker, W. 1982. *Articulatory and acoustic correlates of labial activity in vowels: A cross-linguistic study*. PhD dissertation. University of California, Los Angeles.

Maddieson, I. 1984. *Patterns of sounds*. Cambridge: Cambridge University Press.

Mielke, J. 2002. "Turkish /h/ deletion: evidence for the interplay of speech perception and phonology". *ZAS Papers in Linguistics* 28. 55-72.

Ohala, J. 1993. "The phonetics of sound change". In Jones, Ch. (ed.), *Historical Linguistics: Problems and Perspectives*, London: Longman. 237-278.

Schwartz, J.-L., L. Boe, N. Vallée and C. Abry. 1997b. "The Dispersion-Focalization Theory of vowel systems". *JPhon* 25. 255-86.

Stevens, K. 1989. "On the quantal nature of speech". *JPhon* 17. 3-45.

Terbeek, D. 1977. *A cross-language multi-dimensional scaling study of vowel perception*. PhD dissertation. University of California, Los Angeles.

Individual variability in the perceptual learning of L2 speech sounds and its cognitive correlates

Yoon Hyun Kim and Valerie Hazan

Speech, Hearing and Phonetic Sciences, UCL, London, UK
toithaka@gmail.com, v.hazan@ucl.ac.uk

ABSTRACT

This study explored which cognitive processes are related to individual variability in the learning of novel phonemic contrasts in a second language. 25 English participants were trained to perceive a Korean stop voicing contrast which is novel for English speakers. They were also presented with a large battery of tests which investigated different aspects of their perceptual and cognitive abilities, as well as pre- and post-training tests of their ability to discriminate this novel consonant contrast. The battery included: adaptive psychoacoustic tasks to determine frequency limens, a paired-association task looking at the ability to memorise the pairing of two items, a backward digit span task measuring working memory span, a sentence perception in noise task that quantifies the effect of top-down information as well as signal detection ability, a sorting task investigating the attentional filtering of the key acoustic features. The general measures that were the most often correlated with the ability to learn the novel phonetic contrast were measures of attentional switching (i.e. the ability to reallocate attention), the ability to sort stimuli according to a particular dimension, which is also somewhat linked to allocation of attention, frequency acuity and the ability to associate two unrelated events.

Keywords: Perceptual training, L2 perception, individual variability.

1. INTRODUCTION

Many second-language learners have problems in perceiving sound distinctions in the foreign language that do not occur in their native language (e.g., Iverson et al., 2003; Miyawaki et al., 1975). Studies have now shown conclusively that even short periods of intensive phonetic training can improve the perception of novel consonant or vowel contrasts (e.g. Logan et al., 1991; Iverson & Evans, 2009). Perceptual training can transfer to improvements in the pronunciation of the novel sounds (e.g., Hazan et al., 2005) and the effects of perceptual training can be retained over a long time span (e.g. Bradlow et al. 1999). Studies that have provided data for individual learners have noted that both the ability to discriminate novel contrasts before training and the effect of perceptual training itself can vary quite dramatically across learners. For example Bradlow et al. (1999) found that pre-test English /r/-/l/ identification by Japanese speakers ranged from 52% to 86% for /r/ and 56% to 99% for /l/, and that individual gains after training ranged from 6% to 25% for /r/ identification and from -0.57% to +17% for /l/ identification.

Many potential factors could explain this individual variability. First, a number of factors relating to L2 experience have been shown to affect the learning of the phonetic aspects of a second language. These include age of L2 learning (e.g., Flege et al., 1995), duration of L2 exposure (e.g., Jia et al., 2006), degree of ongoing use of L1 (e.g., Flege et al., 1997). However, even when these factors are carefully controlled, individual variability in the effects of training remain. It has been suggested that individual variability in L2 training could be related to speech processing abilities in the L1 or to more general auditory abilities such as frequency and temporal discrimination (Wong and Perrachione, 2007). Finally, it has been suggested that perceptual learning may be related to more general cognitive abilities such as short-term memory, attention or the learning of associations between two unrelated items (e.g., Goldstone, 1998).

To our knowledge, no study of L2 phonetic training has included a broad battery of cognitive, psychoacoustic and phonetic tests that could help explore the correlates of individual variability in training. In our study, the phonetic contrast chosen for training was the Korean contrast between lenis and aspirated stops. Korean has a three-way stop voicing contrast; the contrast between the lenis and aspirated Korean stop is difficult for English listeners as they are both assimilated to the English voiceless category (Kang & Guion, 2006). For Korean native speakers, F0 at vowel onset and VOT are important cues to this contrast (e.g., Kim, 2007). Our study included a set of training sessions aimed at improving English listeners' ability to hear this novel distinction as well as a battery of pre/post tests. Our aim was to investigate whether the degree of learning and/or ultimate attainment in the perception of this novel contrast could be related to L1 speech processing abilities, auditory acuity or cognitive abilities.

2. METHODOLOGY

2.1. Participants

Participants were 25 British English speakers (18 females, 7 males), all students at UCL in London. They were aged between 18 and 29 years (median: 21) and were screened for normal hearing thresholds. Although some of the participants spoke other languages, none had studied Korean or another language with lenis/aspirated stop contrasts.

2.2. Training task and materials

The computer-based training followed the high-variability phonetic training approach (Logan et al.,1991). Participants were trained to identify the Korean alveolar lenis, /t/ and aspirated, /t^h/ stop. Participants heard a token and had to decide which phoneme they heard by clicking on a button labeled 't' or 'th'. If they made an error, both the sound and the correct label were repeated. Each phoneme was presented in a CV syllable with the vowels, /a/, /i/, or /u/, produced by six speakers (3 females, 3 males). 36 syllables were each repeated four times per training session; there were four training sessions in total.

2.3. Test Battery

2.3.1. Measuring individual variability in the perceptual learning of L2 speech sounds

2.3.1.1. Phoneme identification within syllables

Pre- and post-training tests were carried out to look at the impact of the training. In these tests, participants did the same task as in the training but received no feedback. The test stimuli were CV syllables with the lenis and aspirated stops combined with the vowels /i/, /a/, or /u/. The speakers recorded for the pre/post tests were different to those in the training sessions. Test materials consisted of 6 CV stimuli recorded by 2 men and 2 women, each presented 5 times (total: 120 trials).

2.3.1.2. Phoneme identification within word

Korean words with /t/ or / t^h/ in initial position were used to test the generalization of the training. There were 64 words (2 phonemes * 8 words * 4 speakers), each presented twice (total: 128 trials). The speakers were different from those of the syllable test and training sessions.

2.3.1.3. Acoustic-cue weighting

This test examined which acoustic cues English participants attended to before and after training. A natural syllable, /ta/, was manipulated with Praat to synthesize stimuli with a range of VOT (20 to 80 ms in 10 ms steps) and F0 (170 to 320 mels in 10 mel steps) values. Thus there were 112 stimuli (7 * 16). Participants had to judge whether the stimuli were /t/ or / t^h/.

2.3.2. Measuring individual variability in auditory acuity, L1 speech perception and cognitive tasks

2.3.2.1. Frequency discrimination

An adaptive procedure was used to determine difference limens for single-formant tokens varying in frequency in the F2 range (F2 test) and fundamental frequency range (F0 test). Participants were asked to choose the odd one out among three nonspeech sounds.

2.3.2.2. Attentional filtering of VOT and F0

In this sorting task (Garner, 1974), participants were asked to sort four stimuli into two categories by either F0 or VOT. In correlated-dimension sorting tasks, they classified stimuli into two groups that differed in both dimensions. In orthogonal-dimension sorting tasks, participants sorted stimuli according to one dimension while ignoring differences in the other. The same stimuli, selected from the stimulus set of the acoustic-cue weighting task, were used in the two orthogonal tasks. Their VOT was 20 ms or 80 ms and F0 was 180 mel or 230 mel.

2.3.2.3. Categorical perception of L1 voicing distinction

An adaptive identification test using a synthetic speech continuum was used to test the consistency of labeling (steepness of the identification function) for an English voicing contrast between the syllables ‘pea’ and ‘bee’ (see Hazan et al., 2009 for details).

2.3.2.4. L1 Speech perception in noise

This test assessed participants’ speech perceptual abilities in their native language. Materials were derived from the speech in noise (SPIN) sentences of Bradlow and Alexandler (2007) in which the final word varies in predictability from the context. As in that study, the test design included keywords (15) presented in ‘right context (RC)’ sentences (e.g., ‘the meat from a pig is called pork’) and ‘neutral context (NC)’ sentences (e.g., ‘he talked about the pork’); in addition, a further 15 keywords were presented in neutral context and ‘wrong context (WC)’ (e.g., ‘the meat from a pig is called dinner’).

2.3.2.5. Attention

A number of subtests of the Test of Everyday Attention (TEA, Robertson et al, 1994) were presented. These included: ‘elevator counting with distraction’ (a measure of selective attention), ‘elevator counting with reversal’ (a measure of attentional switching) and ‘lottery’ (a measure of sustained attention).

2.3.2.6. Working memory

A backward digit span task was used to assess complex working memory (Gathercole, 1999). After remembering numbers in order, participants had to reproduce the numbers in inverse order and press corresponding number keys on a keyboard. Number string length varied from 3 to 10.

2.3.2.7. Associative learning

Paired association learning which link two items is related to processes of memory encoding and retrieval (Buckner & Wheeler, 2001). 20 meaningless syllables (CVC) were paired with a number (1 or 2). The cued-recall test was repeated 4 times. A syllable was given as a cue on the screen, and participants had to say which number was paired with it.

2.4. Test procedure

Each participant was tested individually over 10 sessions, lasting between 30 and 50 minutes, carried out on different days. Testing and training took place in a sound-treated booth, with sounds presented via high-quality headphones at a comfortable listening level fixed across participants. At Session 1, participants carried out the syllable and word identification pre-tests and the attentional filtering task. At Session 2, they performed the acoustic-cue weighting test. At Sessions 3 to 6, they carried out their training sessions. At Sessions 7 and 8, they carried out the post-tests (same as pre-tests). At Session 9, they carried out the paired-association task, the working memory span task and the speech perception in noise task. At Session 10, participants were tested on the attention tasks and the F0/F2 difference limen tests.

3. RESULTS

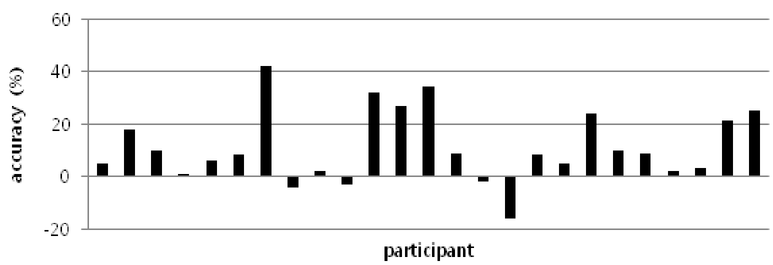
3.1. Pre and post-test results for Korean stop voicing contrast

The mean identification accuracy for the lenis-aspirated Korean voicing contrast in the pretest was just above chance level for both words and syllables. Significant individual variability was present in all measures as can be seen in Table 1. The mean percentage difference between pre- and post test was 11% for syllables and 7.4% for words. The degree of change between pre- and post-test varied across individual learners from -16% to +42% for the syllable test (see Fig. 1) and -7% to +28% for the word test.

Table 1. Accuracy of L2 phoneme identification in training sessions and tests

1.	Mean (%)	s.d.	Minimum	Maximum
syllable identification (pre-test)	55.5	11.9	36	72
syllable identification (post-test)	66.5	9.7	49	85
word identification (pre-test)	51.5	8.2	38	71
word identification (post-test)	58.8	9.0	44	81
training (first session)	66.1	8.6	50	79.9
training (last session)	74.8	11.2	52.1	96.5
training (change across four sessions)	8.7	5.6	-1.3	17.4

Figure 1: Change in syllable identification accuracy between the pre- and post-test for individual participants.



Next, we examined the rate of change between the first and final training sessions (Table 1). The degree of variability was such that, despite minimal previous exposure to this contrast, some participants, after four short training sessions were at ceiling (97%) while others were performing near chance.

Correlations analyses were carried out on the z-scores of pre/post data for the syllable and word tests and training performance. Syllable identification accuracy in the pre-test was only significantly correlated with word accuracy in the pre-test (r=0.566, p=.003), while post-test syllable accuracy was correlated with scores in the first (r=.671) and last (r=.547) training sessions. Pre-test scores of word identification were correlated with word identification post-test scores (r=.443). The performance at first training was highly correlated with the achievement of the last training session (r=.873), and the difference in score between the first and last training sessions was significantly correlated with accuracy in the final training session (r=.654).

3.2. Correlations with cognitive, auditory acuity and speech processing tasks

The next phase of the analysis involved looking at which of the additional measures collected correlated with measures related either to initial ability, to the rate of learning of the novel contrast, or to ultimate attainment. Pearson's correlations were applied to the normalised data (see Table 2).

First, let's examine what measures correlated with the initial sensitivity to the novel contrast. Syllable pre-test scores were significantly correlated with frequency acuity in the F2 region. Word-level pre-test scores were correlated with accuracy of sorting by the VOT dimension (attentional filtering test) and with attention switching ability. Performance at the initial and final training sessions was correlated with performance on the paired association task, with attentional switching ability and with the response time of sorting by the F0 dimension (attentional filtering test). Ultimate attainment (post-test scores) was again correlated with accuracy of sorting by the VOT dimention (both word and syllable tests), with frequency acuity in the F2 region (word test only) and with working memory (syllable test only). There were no significant correlations between performance measures for the novel contrast and measures reflecting speech processing ability in the L1, F0 limen and measures of sustained and selective attention.

In summary, the general measures that were the most often correlated with the ability to learn the novel phonetic contrast were measures of attentional switching (i.e. the ability to reallocate attention), the ability to sort stimuli according to a particular dimension, which is also somewhat linked to allocation of attention, frequency acuity and the ability to associate two unrelated events.

Table 2. Correlation between measures related to the L2 phonetic contrast and measures of speech, auditory or cognitive processing abilities.

	SID_pre	SID_post	WID_pre	WID_post	TR_first	TR_last	TR_diff
F2_limen	**0.547****	0.270	0.296	**0.397***	0.246	0.143	-0.094
F0_limen	-0.037	0.096	-0.045	0.230	0.204	0.244	0.173
PB_slop	-0.291	0.156	-0.197	-0.084	-0.024	-0.175	-0.316
VOT_acc	0.366	**0.412***	**0.414***	**0.693****	0.273	0.250	0.079
F0_acc	0.196	0.386	-0.081	0.219	0.260	0.235	0.068
VOT_rt	-0.047	0.065	-0.079	-0.183	0.020	0.065	0.099
F0_rt	-0.097	-0.285	0.068	-0.030	**-0.487***	**-0.437***	-0.122
Cind	0.223	-0.192	0.092	0.108	-0.316	-0.273	-0.058
Cdep	-0.245	-0.311	-0.031	-0.237	-0.265	-0.364	-0.320
PA_last	0.225	0.090	0.108	0.006	**0.473***	**0.468***	0.206
PA_diff	0.019	0.105	-0.063	0.035	0.352	0.302	0.060
WM	0.157	**0.397***	0.005	0.150	0.355	0.246	-0.058
TEA_sel	0.025	-0.074	-0.137	0.150	0.203	0.243	0.173
TEA_swi	0.023	0.272	**-0.413***	0.048	**0.464***	**0.446***	0.176
TEA_sus	-0.018	0.167	0.176	0.141	0.164	0.082	-0.089

1. * p<.05, ** p<.01
2. SID_pre (syllable identification pre-test), SID_post (syllable identification post-test), WID_pre (word identification pre-test), WID_post (word identification post-test), TR_first (training first session), TR_last (training last session), TR_diff (difference between the first and the last training session), PA_last (paired association last session), PA_diff (difference between the first and the second session of paired association task), WM (working memory span), Cind (context-independent rate of SPIN perception), Cdep (context-dependent rate of SPIN perception), F2_limen (F2 discrimination limen), F0_limen (F0 discrimination limen), PB_slop (slope of categorical perception for L1 voicing contrast), VOT_acc (accuracy of sorting by VOT dimension), F0_acc (accuracy of sorting by F0 dimension), VOT_rt (response time of sorting by VOT dimension), F0_rt (response time of sorting by F0 dimension), TEA_sel (selective attention test), TEA_swi (attentional switching test), TEA_sus (sustained attention test)

4. DISCUSSION

The results of our study confirmed our expectations that English listeners with no experience of Korean would vary significantly in their ability to both initially perceive and learn a novel phonetic contrast which is difficult as it involves two Korean phonemes that both assimilate to a single English category. At the pre-test, some English listeners identified the lenis-aspirated contrast with a high degree of accuracy while many were at chance; after four short training sessions, a few listeners were at ceiling while others had not learned anything about the contrast. In order to identify the Korean voicing contrast, it is necessary to use acoustic cue information from VOT and F0 (Francis and Nusbaum, 2002; Kim, 2007; Kang and Guion, 2006).

Indeed, some learning measures were correlated with VOT sorting accuracy or F0 sorting time. Successful acquisition of L2 speech sounds might depend on whether L2 learners can adjust their sensitivity to small acoustic differences through attention. Indeed, we found correlations between accuracy and frequency limens as well as with measures of attentional switching. Finally, the learning of a new contrast is dependent on being able to associate a novel sound with a new category. There again, learners who were successful at a paired-association task tended to be more successful at learning the novel contrast. It should be noted that the correlations between cognitive or frequency acuity measures and the learning of a novel contrast, although significant, are not necessarily very strong. It is therefore clear that no single dimension can predict successful learning, and that the picture is complex given the multiplicity of factors that contribute to the learning of the sounds of a new language. This study though provides a further contribution to the understanding of which factors can contribute to successful language learning at the phonetic level.

5. REFERENCES

Bradlow, A., Akahane-Yamada, R., Pisoni, D., Tohkura, Y. 1999. Training Japanese listeners to identify English/r/and/l: Long-term retention of learning in perception and production. *Perception & Psychophysics,* 61, 977-985.

Bradlow, A. R. and Alexander, J. A. 2007. Semantic-contextual and acoustic-phonetic enhancements for English sentence-in-noise recognition by native and non-native listeners. *Journal of the Acoustical Society of America,* 121(4), 2339-2349.

Buckner, R., Wheeler, M. 2001. The cognitive neuroscience of remembering. *Nature reviews. Neuroscience* 2(9). 624-634.

Flege, J., Frieda, E., Nozawa, T. 1997. Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics*, 25(2), 169-186.

Flege, J., Munro, M., MacKay, I. 1995. Effects of age of second-language learning on the production of English consonants. *Speech Communication*, 16(1), 1-26.

Flege, J., Takagi, N., Mann, V. 1996. Lexical familiarity and English-language experience affect Japanese adults' perception of/i/and/I. *Journal of the Acoustical Society of America,* 99(2), 1161-1173.

Garner, W. 1974. *The processing of information and structure*: Erlbaum Potomac, MD.

Gathercole, S. 1999. Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences*, 3(11), 410-419.

Goldstone, R. 1998. Perceptual learning. *Annual Reviews in Psychology*, 49(1), 585-612.

Hazan, V., Messaoud-Galusi, S., Rosen, S., Nouwens, S., Shakespeare, B. 2009. Speech perception abilities of adults with dyslexia: is there any evidence for a true deficit? *Journal of Speech, Language, and Hearing Research*, 52(6), 1510-1529.

Hazan, V., Sennema, A., Iba, M., Faulkner, A. 2005. Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication*, 47(3), 360-378.

Iverson, P., Evans, B. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. The Journal of the Acoustical Society of America, 126, 866-879.

Iverson, P., Kuhl, P., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., Siebert, C. 2003. A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition,* 87(1), 47-57.

Jia, G., Strange, W., Wu, Y., Collado, J., Guan, Q. 2006. Perception and production of English vowels by Mandarin speakers: Age-related differences vary with amount of L2 exposure. *Journal of the Acoustical Society of America*, 119-1118.

Kang, K., Guion, S. 2006. Phonological systems in bilinguals: Age of learning effects on the stop consonant systems of Korean-English bilinguals. *Journal of the Acoustical Society of America*, 119, 1672-1683.

Kim, Y. H. 2007. The Effect of Selective Attention on Discrimination between Lenis /t/ and Aspirated /th/ of Korean Alveolar Stops as shown by the Korean and the Japanese. Unpublished doctoral dissertation, Seoul National University, Seoul, Korea.

Logan, J., Lively, S., Pisoni, D. 1991. Training Japanese listeners to identify English/r/and/l: A first report. *Journal of the acoustical society of America,* 89(2), 874-886.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A., Jenkins, J., Fujimura, O. 1975. An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics,* 18(5), 331-340.

Robertson, I. H., Ward, T., Ridgeway, V Nimmo-Smith, I. 1994. *The Test of Every day Attention.* BurySt Edmunds: Thames Valley Test Company.

Schmidt, A. M. 1997. Cross-language consonant identification: English and Korean. In O.-S. Bohn M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 185-200). Philadelphia: John Benjamins Publishing Company.

Wong, P., Perrachione, T. 2007. Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, 28(04), 565-585.

6. ACKNOWLEDGEMENTS

This work was supported by the Korea Research Foundation Grant funded by the Korean Government [KRF-2008-356-H00004] and UCL Department of Speech, Hearing and Phonetic Sciences

Learning Vowel Sounds in a Migrant Setting: The Case of Polish Children and Adults in Ireland

Romana Kopečková

Trinity College Dublin, Ireland
kopeckor@tcd.ie

ABSTRACT

The present paper examines second language speech learning in 40 Polish children and adults after about three years of residence in the Republic of Ireland. These participants' performance in an oddity discrimination task and a delayed repetition task was compared to their perception of similarities between the relevant Polish and Hiberno-English vowel sounds. Situating itself within the general purview of research into age-related differences in second language speech learning (Flege, 1995; Baker et al., 2002; Jia et al., 2006), this paper presents new insights from a very specific situation. It attempts to shed light on the patterns of age-related differences in performance accuracy presented by the Polish participants in the context of recent migration flows within the European Union, and to explore the basis of the different learning outcomes, with a specific focus on the nature of the language input received by the migrant participants.

Keywords: age, second language experience, vowel perception and production, EU migration.

1. INTRODUCTION

There is wealth of empirical evidence suggesting that most children, at least in a long-term perspective, are more successful learners than adults in terms of perception and production of a second language (L2). Most recent studies in the area, motivated by the tenets of Flege's Speech Learning Model (1995), indicate that it may be the level of 'identification' with the sound system of the native language (L1) that distinguishes child L2 learners from those of adults. While children's representations of their mother tongue are still developing, adults perceive and produce the sound system of their L1 with a considerable level of stability. As a result, adults may be more likely to relate the sounds of an L2 to their existing L1 representations and therefore to perceive and produce the L2 less accurately (Baker et al., 2002; 2008). The claim that the ability to perceive differences between L1 and L2 sounds drives successful L2 perception and production has been well-established theoretically, however, rigorous empirical evidence based on comparisons between child and adult L2 learners has been scarce. The present study seeks to address this gap by examining how accurately Polish children and adults perceive and produce Hiberno-English vowels after about three years of residence in Dublin, and whether relationships can be established between their perceptuo-motor skills in the L2 and cross-language perception.

At least two more factors have been shown to influence L2 learners' success in segmental perception and production. Some L2 sounds may be more difficult to learn because they are too similar to their L1 counterparts and as such are categorized within L1 segmental representations (Best, 1995; Flege, 1995). Even young L2 learners after an extended period of residence in the target language country may experience difficulties in discriminating between such sounds (Tsukada et al., 2005). Also, the amount and quality of L2 experience with the target language may affect the development of perceptual and productive skills in an L2 (e.g. Flege et al., 1997; Baker et al., 2002; Jia et al., 2006), although the exact impact of L2 experience on non-native speech learning is difficult to evaluate given the methodological challenges surrounding its measurement (cf., Flege, 2008). While acknowledging the inevitable limitations of measuring and comparing L2 (phonological) input across diverse L2 learners, this study attempts to gain a better understanding of what possibly constitutes a common L2 experience for child and adult migrants, and to determine the relative effect of such experience on learning specific L2 vowel sounds.

The L2 vowels under scrutiny in the present study include /i/ - /ɪ/ and /u:/ - /ʌ/ contrasts in a /bVt/ phonetic context. These segments were chosen since they were hypothesized to present different levels of learnability to Polish speakers. According to the predictions of the Speech Learning Model (Flege, 1995), English /i/ and /ɪ/ will be difficult to learn by Polish speakers because both are likely to be perceived as highly similar to the native vowel /i/. The same will hold true for the perception and production of English /u:/, which is hypothesized to be mapped onto the Polish /u/. However, some differences in the perceptual L1-L2 match can be expected across the three vowels since English /ɪ/ may in fact be perceived as close to yet another vowel, Polish /ɨ/. A specific case is that of Hiberno-English /ʌ/, which is typically produced as a mid centralized back somewhat rounded vowel, in some Dublin accents approximating to the production of /ʊ/ (Hughes et al., 2005:115). It is likely to be perceived accurately by Polish learners since it is a salient L2 sound which does not interfere with the representation of any of their L1 categories; however, its production may be challenging, particularly for those learners without substantial L2 experience (Flege et al., 1997).

Finally, it may be worth briefly mentioning the background to the Polish migration to the Republic of Ireland, since it represents a unique case of linguistic and cultural integration in today's Europe. With Poland's accession to the European Union in May 2004, large numbers of Poles arrived in the Republic in search of a better life and career prospects. In fact, Poles came to form the largest 'new migrant' group in Ireland at one moment in time, to the extent of outnumbering its population of native speakers of Irish (Inglis, 2008:108). According to Singleton et al. (2007), the Poles arriving in Ireland in the years following Poland's EU accession were exceptionally motivated learners of English, frequently with ambitions of acquiring English to native-like levels, although they initially faced a struggle with the sound system of the Irish variety of English they encountered on arrival. Also, they tended to be exposed to frequent contact with their native language and English produced by other migrants. It thus seems worthwhile to examine how accurately Polish children and adults perceive and produce their L2 after about three years of stay in Ireland.

2. METHOD

In order to test the predictions presented above, the participants in the present study performed three language tasks and completed an extensive background questionnaire.

2.1. Participants

40 native Polish (NP) children and adults were recruited via advertisements in Polish newspapers and shops in Dublin in 2009. A condition of their selection was that they had to have arrived in Ireland after Poland's accession to the EU. This feature of the study made it possible to gain insights into a particular stage of their L2 speech learning in a specific type of migrant environment. A control group of 20 native English speaking (NS) children and adults were chosen on the basis of their place of birth (Dublin) and of not having learnt Polish as a foreign language. Another control group of 19 native Polish (NPP) children and adults living in Poland and having no English immersion experience were also included in the study. These participants were recruited on the basis of bearing something of a linguistic resemblance to a group of Polish migrants as they might be on their first day of arrival in the host country. Table 1 below provides an overview of the main characteristics of the participant groups.

Table 1: Participants in the study - means for age of arrival (AOA), chronological age (Chronage), length of residence in years (LOR), and L2 proficiency ranging from 1=beginner to 6=native-like (Proficiency) are provided in the brackets.

Groups	AOA	Chronage	LOR	Proficiency	Number of participants
Polish children in Ireland (NP)	7-12 (9.9)	12-15 (13.2)	1-5 (3.4)	1-5 (3.5)	20
Polish adults in Ireland (NP)	21-55 (26.4)	24-53 (29.6)	1-5 (3.2)	1-6 (3.2)	20
Irish children (NS)		9-14 (11.4)			10
Irish adults (NS)		25-42 (29.6)			10
Polish children in Poland (NPP)		10-12 (11.1)			10
Polish adults in Poland (NPP)		27-48 (33.1)			9

2.2. Language tasks

The tasks for NP and NS adults were administered in a quiet language laboratory room at Trinity College Dublin. The young NP and NS participants as well as all NPP participants were visited in their school or home, where the tasks were completed in a quiet study room. The tasks taken by the participants in their individual sessions are described below in terms of their complete form; however, as noted earlier, further analyses in this paper will focus only on the perception and production of the vowels /i/, /ɪ/, /u:/ and /ʌ/.

2.2.1. Cross-language identification

In order to determine whether child and adult L2 learners perceive the similarity between L1 and L2 vowels differently, NP and NPP participants in the study completed a cross-language identification task. The participants heard eight English vowels in a bVt format one at a time. The moment the L2 stimulus was presented, six Polish keywords in their orthographic form (only) could be seen on the screen as well: *bity* /bitɨ/, *byty* /bɨtɨ/, *buty* /butɨ/, *bety* /bɛtɨ/, *baty* /batɨ/, and *boty* /bɔtɨ/. The participants heard the target words (*beat*, *bit*, *boot*, *but*, *bat*, *bet*, *boat* and *bought*) twice. First, they heard the English word and matched it to one of the six Polish keywords to which they believed it was most similar. Second, they rated the degree of similarity between the L1 and L2 vowels using a 7-point Lickert scale ranging from '1' indicating that the two sounds were not similar at all and '7' indicating that they were a complete match.

2.2.2. Categorical discrimination

In order to test the discrimination abilities of the Polish participants in terms of non-native vowels, a categorical discrimination task in an oddity format was administered. Each of the tested L2 contrasts /i/-/ɪ/, /ɛ/-æ/, /ɪ/-/ɛ/, /u:/-/ʌ/ and /ɔ:/-/əʊ/ was presented in all possible combinations. In total, 38 triad items (5 contrasts x 6 change combinations x 8 no-change tokens) were presented to all the participants in the study.

For the purpose of the analysis A prime (*A'*) scores were calculated. These scores were based on the proportion of "hits", i.e. instances of correct identification of odd items out in change triads, and "false alarms", i.e. instances of incorrect identification of odd items out in no-change triads for each contrast, using the formula by Snodgrass et al. (1985). An *A'* score of 1.0 represents perfect discrimination of a contrast while an *A'* score of 0.5 and lower suggests insensitivity to discrimination of a sound contrast.

2.2.3. Delayed repetition

The production abilities of the Polish and Irish participants were tested in a delayed repetition task which elicited the same vowels that were used in the previous tasks. Seven native speakers of Hiberno-English subsequently identified the productions in a self-paced online presentation in a forced-choice task. An intelligibility score was based on the percentage of times they identified each of the L2 vowels as intended.

2.3. L2 input data

An extensive background questionnaire elicited general biographical, socio-psychological and language-educational data from the participants. As for the L2 input data discussed in this paper, the participants reported on the frequency of their English use in diverse contexts (with their family, with friends, in leisure time, at work/at school, in passive activities, such as reading and watching TV). Also, they reported on what kind of speakers they tended to communicate with in Ireland. Similarly, the participants reported on their use of Polish in their everyday life in Ireland. The contrasting characteristics of the Polish children and adults in terms of their L2/L1 contact are displayed in Table 2 below.

Table 2: Reported L2/L1 contact by Polish participants in Ireland (mean values and standard deviations; 1='very often', 2='often', 3='sometimes', 4='rarely', 5='never').

Group	L2 use	L1 use	L2 use with native speakers	L2 use with non-native speakers
adults	2.77 (.437)	2.45 (.510)	1.75 (.786)	2.10 (.852)
children	2.42 (.467)	2.66 (.699)	1.42 (.607)	2.42 (1.170)

3. RESULTS

3.1. Cross-language identification

Analyses of the cross-language identifications revealed that Polish children and adults chose the same primary response when categorizing the English vowels /i/, /ɪ/ and /u:/, but not when classifying the Hiberno-English /ʌ/. Whereas NP children were likely to select Polish /u/ as the closest L1 vowel to the L2 /ʌ/, NP adults judged this sound as closer to Polish /o/. Interestingly, the patterns of cross-language identification were reversed for the NPP participants. Confusion matrices for the proportion of times each English vowel was classified as its primary Polish response alternative are provided in Table 3 below. The table also includes mean similarity ratings and fit indices for each English vowel in terms of its Polish counterpart. Fit indices represent a useful metric which combines the identification and goodness-of-fit data into a single value. The fit indices in this study spanned from a low value of 1.09 (the fit of Hiberno-English /ʌ/ to Polish /u/ as judged by NP adults) to a high of 5.1 (the fit of English /i/ to Polish /i/ as categorized by NPP children), the assumption being that the higher the fit index for an L2 sound, the more readily this sound is accepted as an instance of the L1 category.

A series of ANOVA analyses revealed that even though L2 vowels /i:/ and /ɪ/ as perceived by NP children received lower fit indices than those by NP adults, the difference did not reach statistical significance. The L2 vowels /u:/ and /ʌ/ were, in fact, more readily assigned as instances of Polish /u/ by NP children than adults, although these differences were not statistically significant either. Likewise, no significant differences were found between the judgements of similarity by NPP children and adults; however, some interesting tendencies can be observed in the data. NPP children were more likely to assign higher similarity ratings to all the tested vowels, with the exception of Hiberno-English /ʌ/ which received low fit indices in general.

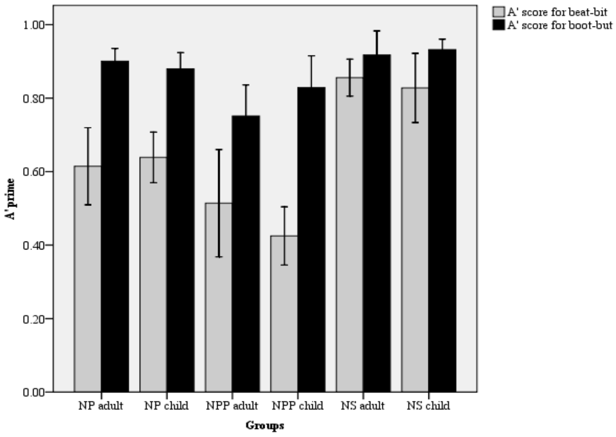
Table 3: Fit indices derived for English vowels in terms of Polish vowels. Only those identifications common to at least 30% of the participants are included.

L2 vowel	Common identifications	Proportion of identifications	Goodness ratings	Fit index
i:	NP adult /i/	0.90	5.11	4.59
	NP child /i/	0.95	4.42	4.19
	NPP adult /i/	0.89	4.00	3.56
	NPP child /i/	1.00	5.10	5.10
ɪ	NP adult /i/	0.95	4.68	4.45
	NP child /i/	0.70	4.57	3.20
	NP child /ɨ/	0.30	5.50	1.65
	NPP adult /i/	0.89	2.88	2.56
	NPP child /i/	1.00	4.30	4.30
u:	NP adult /u/	0.90	4.06	3.65
	NP child /u/	0.95	4.11	3.90
	NPP adult /u/	1.00	3.11	3.11
	NPP child /u/	1.00	4.20	4.20
Hiberno-English ʌ	NP adult /u/	0.35	3.14	1.09
	NP adult /o/	0.55	3.36	1.84
	NP child /u/	0.60	4.25	2.55
	NPP adult /u/	0.67	3.33	2.23
	NPP adult /o/	0.33	3.67	1.21
	NPP child /u/	0.40	3.00	1.20
	NPP child /o/	0.60	3.83	2.30

3.2. Categorical discrimination

Analyses of the perception of the two English vowel contrasts in Polish and native speaker participants revealed that NP children performed as accurately as NS children in the discrimination of the /i/ - /ɪ/ as well as the /u:/ - /ʌ/, while NPP adults were comparable to the NS adults in their discrimination of the latter contrast only. All Polish participants received low *A'* scores for discriminating /i:/ - /ɪ/, indicating low levels of sensitivity for this contrast (Figure 1). The *A'* scores for this contrast were submitted to a group x contrast ANOVA, yielding a significant main effect for group F (5,73) = 9.2, $p<0.01$. Also significant was the main effect for group for the /u:/ - /ʌ/ contrast, F (5,73) = 4.57, $p< 0.01$. A Tukey's HSD test ($\alpha = 0.05$) revealed that the native speakers differed from the NPP in all instances, with the exception of NPP children who discriminated /u/ - /ʌ/ contrast as accurately as both NP and NS participants. NP children and adults did not differ significantly from one another in the discrimination of either of the tested contrasts.

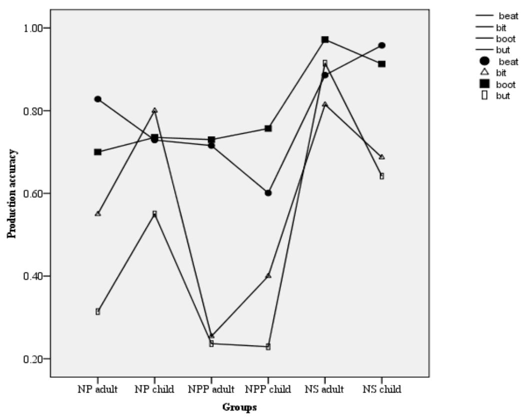
Figure 1: Mean *A'* scores obtained for the two tested contrasts. Error bars represent standard errors. *A'* score of 1.0 indicates perfect discrimination and a score of 0.5 or below indicates insensitivity to a contrast.



3.3. Production

In production, similar ANOVA analyses (group x vowel) yielded a significant main effect for group F (5,73) = 2,48, $p<0.05$ for /i:/, F (5,73) = 6.43, $p< 0.01$ for /ɪ/, F (5,73) = 2.52, $p<0.05$ for /u:/, and F(5.73) = 12.32, $p<0.01$ for /ʌ/ vowel sounds. Overall, NP children and adults produced the tested vowels more accurately than NPP participants, and comparably to native speakers, with the exception of /ʌ/ production. This L2 vowel was produced less accurately by NP adults as compared to NS adults as well to NP children, who produced the vowel concerned as accurately as NS children (Figure 2).

Figure 2: Production accuracy for all six groups of participants for the four tested vowels



4. DISCUSSION AND CONCLUSION

Unlike in Baker et al. (2002), the results of this study have not corroborated the hypothesis of the Speech Learning Model (Flege, 1995) that L1 sound categories are less powerful attractors of L2 sounds in child than in adult L2 learners. Polish children living in Ireland for three years performed comparably in a cross-language identification task to Polish adults residing in the country for the same period of time. In other words, NP children in this study were not less likely than NP adults to identify the tested L2 sounds with corresponding L1 categories. On the other hand, the findings of this study support the predictions of the model about similar versus dissimilar L2 and L1 vowels insofar as even children were challenged in the discrimination and production of highly similar L2 vowels while performing to native-like levels in the production of novel L2 sounds. It is to be noted that the adult L2 learners did not produce the specific L2 sounds as accurately as L2 children; their performance was comparable to that of adult learners without any immersion experience.

These findings suggest that the experience of learning the L2 sound system in the target language country may benefit learners to a different extent, at least in terms of the 'more learnable' L2 sounds (Baker et al., 2002; Aoyama et al., 2008). The NP children of this study reported a more intensive L2 contact with native speakers in both formal and informal contexts than NP adults did, which may have supported their learning of such novel sounds as Hiberno-English /ʌ/ more. What distinguished the participants' reports on L2 input related to the nature of relationships with native speakers. While Polish children reported frequent use of English with their Irish friends (90%), Polish adults reported using English in the same context in just 55% of cases. Thus, as Moyer (2008) has recently highlighted, it is the quantity and quality of L2 input together (received and actively sought out) that seems to make a difference for early L2 learners.

There are several limitations to the current study which need to be borne in mind when interpreting its results. Only a limited number of L2 vowels were involved in the investigation and these in addition were highly controlled in terms of consonantal surrounding. Also, estimates of L2 input were based on participants' reports, rather than on measurement. Despite the limitations, this study hopes to add to the growing body of research into L2 acquisition which directly compares children and adults in terms of their experience of L2 speech learning in diverse learning contexts.

5. REFERENCES

Aoyama, K., Guion, S. G., Flege, J. E., Yamada, T., Akahane-Yamada, R. 2008. The first years in an L2-speaking environment: A comparison of Japanese children and adults learning American English. *IRAL* 46(1). 61-90.

Baker, W., Trofimovich, P., Mack, M., Flege, J. E. 2002. The effect of perceived phonetic similarity on non-native sound learning by children and adults. *Proceedings of the Annual Boston University Conference on Language Development* 26(1). 36-47.

Baker, W., Trofimovich, P., Flege, J. E., Mack, M., Halter, R. 2008. Child-adult differences in second language phonetic learning: the role of cross-language similarity. *Language and Speech* 51(4). 317-342.

Best, C. T. 1995. A direct realist view of cross-language speech perception. In Strange, W. (ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press. 171-204.

Flege, J. E. 1995. Second-language speech learning: theory, findings, and problems. In: Strange, W. (ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press. 229-273.

Flege, J. E., Bohn, O.-S., Jang, S. 1997. Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics* 25(4), 437-470.

Flege, J. E. (2008). Give input a chance! In: Piske, T., Young-Scholten, M. (eds.), *Input matters in SLA*. Bristol: Multilingual Matters. 175-190.

Hughes, A., Trudgill, P., Watt, D. J. L. 2005. *English accents and dialects: An introduction to social and regional varieties of English in the British Isles* (4th ed.). London: Hodder Arnold.

Inglis, T. 2008. *Global Ireland: Same difference*, New York: Rutledge.

Jia, G., Strange, W., Wu, J.,Collado, J. 2006. Perception and production of English vowels by Mandarin speakers: Age-related differences vary with amount of L2 exposure. *The Journal of the Acoustical Society of America* 119(2). 118-1130.

Moyer, A. (2008). Input as a critical means to an end: quantity and quality of experience in L2 phonological attainment. In: Piske, T., Young-Scholten, M. (eds.) *Input matters in SLA*. Bristol: Multilingual Matters. 159-174.

Singleton, D., Skrzypek, A., Kopeckova, R., Bidzinska, B. 2007. *Attitudes towards and perceptions of English L2 acquisition among Polish migrants in Ireland.* Paper presented at the Royal Irish Academy research symposium "In / Difference", Limerick.

Snodgrass, J., Levy-Berger, G., Haydon, M. 1985. *Human Experimental Psychology*. New York: Oxford University Press.

Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., Flege, J. E. 2005. A developmental study of English vowel production and perception by native Korean adults and children. *Journal of Phonetics* 33(3), 263-290.

Top-down processes and experience in second-language speech perception

Michal Kubánek, Václav Jonáš Podlipský

Department of English and American Studies, Palacký University in Olomouc, Czech Republic
michal.kubanek01@upol.cz, vaclav.j.podlipsky@upol.cz

ABSTRACT

Speech perception has been described as a bi-directional process. This is to say that listeners make use of both bottom-up and top-down cues. Research of second-language (L2) speech perception, however, suggests that non-native listeners do not make efficient use of top-down cues when experiencing problems with bottom-up processing. This study investigates whether the relative degree of importance of bottom-up and top-down cues is related to the amount of L2 experience. Our subjects were Czech learners of English as a foreign language (EFL). Three experiments were conducted that all involved a conflict between segmental (bottom-up) and contextual (top-down) information. Contextual cues were provided either via the same (audio) mode or a different (visual) mode than the segmental cues. Results showed that the effect of conflicting cues was very similar across experience levels in audio mode. On the other hand, when listeners were forced to process audio (bottom-up) and visual (top-down) information simultaneously, less experienced listeners relied more strongly on top-down cues than more experienced listeners. Although the results of the experiments do not all point in the same direction, we conclude that with growing EFL experience the importance of top-down contextual cues declines.

Keywords: Speech perception, non-native, top-down cues, bottom-up cues, L2 experience

1. INTRODUCTION

1.1. Background

There is evidence that native speakers of a particular language make use not only of the acoustic information contained in the speech signal (the bottom-up direction) but also of contextual information such as sentence semantics or situational cues (top-down direction) when perceiving speech (see e.g. Davis and Johnsrude 2007, for a review). This is why advanced models of speech perception have incorporated both directions of processing. For example, the TRACE model uses the principle of bidirectional interactive activation (McClelland and Elman 1986), while Polysp (Hawkins and Smith 2001) builds a broad polysystemic model drawing on different sources of information utilized in speech perception.

The relative degree of listeners' reliance on the bottom-up cues on the one hand and top-down cues on the other is naturally not constant. In some circumstances, context cues are scarce (for instance when you are hearing the name of an unknown street, firm, or person for that matter) and bottom-up cues have to be relied on heavily. On the other hand, if context cues are available and bottom-up cues are degraded by external noise (of any type), a stronger reliance on top-down processing will be induced. This effect has been demonstrated in several, now classic, studies such as those by Warren and Warren (1970), Garnes and Bond (1976), or Kalikow et al. (1977). Not only is the relative importance of top-down and bottom-up cues modulated by these immediate conditions of communication, listener characteristics also play a role. In particular, it is well known that listeners suffer from age-related declines in speech perception capability which has the consequence that older adult listeners make greater use of top-down contextual constraints than younger adult listeners (for a review, see Sommers 2005).

Another listener characteristic that affects the relative degree of reliance on the two types of cues is listeners' language experience, by which we here mean whether they are native speakers of the language or not. In suboptimal (i.e. noisy) listening conditions, a difference in speech recognition ability between native and even highly proficient non-native listeners becomes apparent (e.g. Gat and Keith 1978; Takata and

Nábělek 1990). There is evidence suggesting that this difference results from non-native listeners' failure to use top-down information as efficiently as it is used by native listeners (Mayo et al. 1997; Cutler et al. 2004).

This is not to say, however, that non-native listeners do not use contextual cues at all. Utilization of linguistic sources was investigated for example by Hahne (2001) who measured event related potentials (ERP) in her subjects as a specific reaction on semantic incongruence in sentences presented to them. Comparing the data obtained for native and non-native listeners, she came to the conclusion that semantic integration, i.e. top-down information utilization, is slower and consumes more mental resources for non-native listeners. Nevertheless, this integration is done in a comparable manner, in other words the difference in performance of these two groups seems to be quantitative rather than qualitative. Similarly, Bradlow and Alexander (2007) tested how the availability of good bottom-up cues (clear speech) and of good top-down cues (high semantic predictability of a word) contributed to an enhanced perceptual performance in native and non-native listeners. They found that both groups of listeners employed essentially the same strategies (both bottom-up and top-down); the difference was that while the native listeners took advantage of both types of cues separately and in combination, the non-native listeners benefited from the bottom-up and top-down cue only in their combination.

1.2. Present Study

To sum up so far, previous studies suggest that non-native listeners utilize both segmental and contextual information during speech recognition. What makes non-native perception distinct from native perception is the different efficiency of processing and the potentially different balance between bottom-up and top-down cues.

The aim of our study will be to explore whether there is an interaction between (i) the just-mentioned proportion of the importance of bottom-up and top-down cues for speech recognition and (ii) the amount of foreign-language experience. We will focus on learners of English as a foreign language (EFL). The study by Bradlow and Alexander (2007) referred to above did examine EFL learners, but it did not evaluate the effect of L2 experience. To our knowledge, there is at least one study in which the level of L2 experience was considered as a factor affecting the bottom-up vs. top-down reliance (Mayo et al. 1997, who found that more experienced learners could make better use of context than less experienced learners), however this study tested only highly proficient natural-settings bilinguals (differing in their age of onset of learning) rather than EFL students.

Two researchers have put forward hypotheses specifically concerning EFL learners' employment of top-down (contextual and co-textual) cues in listening tasks. Field (1998) maintained that less advanced learners can efficiently employ compensatory techniques (e.g. context-based inferences) when experiencing segmental-level problems. Nevertheless, Field based his judgment on an experiment testing listeners' ability to guess the overall meaning of a text passage and not context-cues employment for word recognition in the narrower sense which, in fact, was found to be rather poor. Jenkins (2000), who confined her observations to word recognition in our narrow sense, asserted that even relatively advanced EFL learners do not make efficient use of top-down processing. She proposed two main explanations for this. First, most communication by non-native speakers of English happens in an environment where all of the speakers are non-native and, consequently, their production is more careful (suppressing coarticulation and assimilations) which ultimately favours bottom-up processing on the part of the listeners. Second, Jenkins claims that the decreased top-down efficiency emerges because learners fail to ascribe incomprehension to sources other than the introspectively experienced ineptitude of speech-sound processing and that, in turn, causes an over-reliance on bottom-up strategies.

In our experimental design, we draw on the observations and examples of failure in communication offered by Jenkins (2000). For that reason, two layers of sources of top-down cues are distinguished in this study. First, the failure to use linguistic, or 'co-textual', source is illustrated by the sentence *The table was surrounded by chairs* where a listener substituted the final word with a semantically incongruent, even though similarly sounding word *chess* (ibid.: 85) Second, the inability to employ extralinguistic, or 'contextual', information (i.e. what the speaker knows about the situation spoken about from any source outside of language as such) exemplified by the case of a listener disregarding the picture of a *red* car and substituting the attribute with the word *let* for which there was no visual evidence (ibid.: 81-2).

Exploitation of semantic (co-textual) cues by non-native listeners was explored in several studies mentioned above. Integration of visual (contextual) cues with acoustic signal, i.e. utilization of the extralinguistic top-down source, was attested for native listeners for example by Tanenhaus et al. (1995) or Kamide et al. (2003). Both studies employed eye-tracking for investigating anticipatory eye movements over visually depicted objects in a picture while the subjects were listening to auditory input. The subjects moved their eyes in anticipation of the completion of a sentence to the appropriate object in the picture even before the onset of the word which denoted the object as early as they received sufficient information derived from the acoustic signal. As for visual cues integration in non-native listeners, it has been shown to manifest itself in McGurk effect (lip-reading) both on the level of phonemic contrasts in non-words (Hazan et al. 2006) and word recognition in different speech styles (Hardison 2005).

2. EXPERIMENTS

To assess the relative importance of bottom-up and top-down cues for non-native listeners of different amount of L2 experience, all our experiments compared conditions of conflict between these two types of cues with conditions of accord between them. We measured listeners’ reaction times (RT) and the ‘segmental fidelity’ (SF) which was defined as the percentage of cases in which a listener’s response agreed with the segmental (bottom-up) cues and agreed *or* disagreed with the top-down cues. We used repeated-measures ANOVAs with experience as a between-subject factor and condition as a within-subject factor and RT and SF as dependent variables; these analyses were supplemented by post-hoc Scheffé’s tests when necessary. Generally, we hypothesized that if listeners do integrate bottom-up with top-down cues, a conflict between these two types of cues should result in higher RTs and lower SF.

2.1. Subjects

Thirty Czech learners of English as a foreign language (EFL) participated in all three experiments: half in a lower-experience (LE) group and half in a higher-experience (HE) group. The level of experience was based on the length of EFL learning period, the LE group subjects being grammar school students learning English for 3-4 years, the HE group subjects being university freshmen majoring in English who had been learning this language for at least 9 years. Table 1 below summarizes other statistical data. None of the subjects reported any known hearing or language difficulties.

Table 1: An overview of the length of learning, age and sex of listeners in the lower-experience and higher-experience group.

Group	EFL learning period (min./mean/max)	Age (min./mean/max)	Males/Females
Lower Experience	3 / 3.33 / 4	12 / 13.60 / 16	7/8
Higher Experience	9 / 10.87 / 14	20 / 20.73 / 24	6/9

2.2. Materials

All the experimental target words were taken from a dictionary accompanying the course book used by the LE group subjects and pre-tested with them for familiarity. The experimental pattern ensured that each target word of a given experiment appeared in all experimental conditions and, at the same time, that each subject heard a given target word only once. This potentially enabled detection of poorly recognized (possibly unfamiliar) target words across conditions and at the same time eliminated cross-condition priming. Pictures for the visual mode experiments depicted scenes (a park, an office, a wedding, a kitchen table, etc.) which were composed of 6 to 8 easily recognizable objects in a unified visual style.

2.3. Experiment 1

2.3.1. Design

The first experiment explored the interaction between segmental and semantic (co-textual) cues. The experimental design was inspired by FitzPatrick and Indefrey (2007) who used semantically rich sentences with controlled target word congruence and phonemic overlap to induce acoustic priming. In the present experiment, subjects heard 16 sentences in which the final (target) words were masked with fade-in noise. The target words were always polysyllabic; the fade-in portion was set to cover the initial fifth of their duration after which the noise covering reached its full intensity. Pink noise (also called 1/f noise) with signal-to-noise ratio (SNR) of 0 dB was used.

In half of the sentences the target words were semantically congruent with the rest of the sentence ('congruent' condition), in the other half they were incongruent but shared the initial phoneme with a co-textually congruent (but absent) word ('competition' condition). This experimental pattern is illustrated by the swap of the final words in this set of sample sentences (words in italics serve as semantic primers): "The *season* which comes after *summer* is autumn/author. / The *book* was *written* by two authors/autumns." Subjects were asked to repeat each target word.

2.3.2. Results and discussion

In both groups, RTs were significantly longer in the 'competition' than in the 'congruent' condition (LE: $p<.05$, HE: $p<.001$ [post-hoc Scheffé's]). As for SF, this was significantly lower in the 'competition' than in the 'congruent' condition for both groups (LE: $p<.01$, HE: $p<.05$ [post-hoc Scheffé's]). Accordingly, no interaction between experience and condition was found either for RT or SF.

The delay in RTs apparently occurred because listeners did consider context cues (disagreeing with segmental cues). The results are in accordance with the findings of FitzPatrick and Indefrey (2007) revealed by ERP which showed that non-native listeners are sensitive to semantically incongruent sentences in the L2 and that they start semantic integration quite early, even before complete phonetic information about the word is available.

Interestingly, post-experimental analysis discovered several examples in which members of both groups substituted the semantically incongruent target word with a more fitting one, e.g. in the sample set of sentences, different subjects "swapped back" the target words and provided the semantically congruent one. One HE group member even substituted "author" by semantically quite suitable and similarly sounding "awful". These instances, however, were scarce in the response pool and did not play a significant role in the overall results.

2.4. Experiment 2

2.4.1. Design

When studying the integration of top-down and bottom-up cues we wanted to compare single-mode (audition) with cross-mode (audition and vision) access to cues. Experiment 2 explored the degree to which EFL learners employ contextual information available via the visual mode. Subjects heard 16 sentences the final (target) word of which was noise-masked in the same way as in experiment 1, only the SNR was increased to 3 dB to balance for the additional visual information. The target word was always unpredictable from the sentence, nevertheless, it was always semantically congruent (there was no semantic clash between the sentence and the target word).

Each sentence was accompanied by a picture which appeared 2 seconds prior to the onset of the sentence accompanying and describing the picture, so that the subjects had time to inspect the picture. Half of the sentences contained target words which referred to objects present in the picture ('congruent' condition) and the other half contained target words which referred to objects not depicted in the picture. That is, the environment was the same, but the target object was substituted by a different one ('competition' condition). Like in experiment 1, subjects were asked to repeat each target word.

2.4.2. Results and discussion

The results did not reveal any interaction between condition and experience for either of the two dependent variables measured, RT and SF. No significant effect of condition on either RT or SF in either group was found.

The intended conflict of visually provided contextual cues with the segmental cues incoming in audio mode in the ‘competition’ condition did not apparently create a disturbance for the listeners. It seems that they might have accommodated for the speech related task and concentrated predominantly on auditory information the semantics of which produced no conflict between segmental and co-textual cues. Not even the higher SNR forced the subjects to derive compensatory information from the visual mode which might have been recognized as not always ‘reliable’ and subsequently ignored altogether.

This design might be further developed in a future experiment employing minimal pairs manipulated so as to create several instances across a continuum in a similar fashion used already by Garnes and Bond (1976) in audio mode only. In this manner, sentences like “There is a pin / bin on the floor.” could be prepared and combined with pictures depicting one of these two objects. The aim would be to find out whether the perceptual boundary between the two phonemes represented by the value of the relevant distinctive feature shifts under the influence of the target presented in the visual mode, in other words, whether a trade-off between acoustic and visual information could be detected.

2.5. Experiment 3

2.5.1. Design

The final experiment was intended to further investigate forced cross-modal processing. Similar sentences as in experiment 2 were used, this time without any noise masking. Again, 2 seconds prior to the onset of each sentence a picture appeared to enable the subjects to inspect it. The sentence-to-picture relation fell into three categories: (i) ‘congruent’ in which the whole sentence was about the picture, (ii) ‘partially incongruent’ in which the sentence was about the picture, but deviated in one word/object, and (iii) ‘incongruent’ in which the sentence completely mismatched with the picture. Subjects were asked to decide whether the sentence they heard described the picture (true/false responses; response accuracy was measured in percentages).

2.5.2. Results and discussion

The HE listeners responded with no significant difference in accuracy across conditions. Their RTs were significantly shorter in the ‘incongruent’ condition compared to both the remaining two conditions ($p<.05$ for both [post-hoc Scheffé’s]). In contrast, the response accuracy of LE group dropped significantly in ‘partially incongruent’ condition relative to both the remaining conditions ($p<.01$ for both [post-hoc Scheffé’s]) while their RTs were not affected by condition. Consequently, unlike in experiments 1 and 2, a significant interaction between experience and condition was found ($F[2, 56]=4.3943$, $p=.01688$).

The results of this experiment reflect the different level of experience of the two groups. The listeners of the more experienced group were able to detect cross-modal deviations in detail in the ‘partially incongruent’ condition successfully which might be ascribed to their more developed segmental processing abilities. On the other hand, LE listeners seem to put less weight on the information incoming in the acoustic signal the processing of which is not so developed and therefore not considered quite “reliable” by them. Thus, they seem to prefer employing compensatory top-down strategy of generalization (namely, if most of the words recognized in the sentence can be mapped on the picture, then it is likely that the answer is “true”).

Continuous integration of information retrieved from the acoustic signal with information provided visually is apparent from the significant drop of RTs in HE group listeners: early detected incongruence speeded up their responses. It seems that listeners with lower experience process the two modes separately and the integration is done in the post-processing phase. These findings, however, are difficult to assess, since another variable could be at play here; namely the different mean age of the two experimental groups. It might be the case that the different results stemmed from uneven age-related level of development of general cognitive strategies. This is a methodological shortcoming that we have to acknowledge.

3. CONCLUSIONS

Based on the results of experiment 1, which involved a conflict between bottom-up and top-down cues (both provided auditorily), we can conclude that EFL listeners at both tested levels of language experience utilize top-down cues in speech perception. This is because the conflict delayed their responses and because it led to a decrease in responses agreeing with bottom-up cues. This observation corroborates the results of previous, such as Hahne (2001) or FitzPatrick and Indefrey (2007). In experiment 1, no interaction between condition and experience was found and therefore no direct evidence of the influence of L2 experience on the use of top-down cues in speech perception was obtained.

The attempt to assess the relative degree of importance of top-down cues provided visually in experiment 2 did not result in any statistically significant differences between conditions. It appears that listeners disregarded the additional information provided in the form of pictures and concentrated on acoustic information only. Corruption of segmental cues by noise masking did not induce compensatory utilization of top-down visually provided cues.

Unlike in experiment 1, in experiment 3, which involved a conflict between auditory bottom-up cues and visual top-down cues, a significant interaction between condition and experience was found. The reliance on top-down cues was stronger in less experienced than in more experienced listeners.

Although the results of the experiments do not all point in the same direction, we conclude that with growing EFL experience the importance of top-down contextual cues declines. When bottom-up cues are less readily available due to lower language experience, the listeners seem to draw on contextual information and impose it over the gaps caused by segmental level deficiencies; with growing experience, the need for this strategy decreases and the listeners rely on bottom-up cues more heavily.

4. REFERENCES

Bradlow, A. R., Alexander, J. A. 2007. Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *Journal of the Acoustical Society of America* 121(4). 2339-2349.

Cutler, A., Weber, A., Smits, R., Cooper, N. 2004. Patterns of English phoneme confusions by native and non-native listeners. *Journal of the Acoustical Society of America* 116(6). 3668-3678.

Davis, M. H., Johnsrude, I. S. 2007. Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research* 229(1-2). 132-147.

Field, J. 1998. Skills and strategies: towards a new methodology for listening. *ELT Journal* 52(2), 110–118.

FitzPatrick, I., Indefrey, P. 2007. Effects of sentence context in L2 natural speech comprehension. *Nijmegen CNS Journal* 2(1). 43–56.

Garnes, S., Bond, Z. S. 1976. The relationship between acoustic information and semantic expectation. *Phonologica* 3. 285–293.

Gat, I. B., Keith, R. W. 1978. An effect of linguistic experience. *Audiology* 17(4). 339-345.

Hahne, A. 2001. What is different in second-language processing? Evidence from event-related brain potentials. *Journal of Psycholinguistic Research* 30(3). 251–266.

Hardison, D. M. 2005. Variability in bimodal spoken language processing by native and nonnative speakers of English: A closer look at effects of speech style. *Speech Communication* 46. 73–93.

Hawkins, S., Smith, R. 2001. Polysp: a polysystemic, phonetically-rich approach to speech understanding. *Italian Journal of Linguistics - Rivista di Linguistica* 13. 99-188.

Hazan, V., Sennema, A., Faulkner, A., Ortega-Llebaria, M. 2006. The use of visual cues in the perception of non-native consonant contrasts. *Journal of the Acoustical Society of America* 119(3). 1740–1751.

Jenkins, J. 2000. *The Phonology of English as an International Language*. Oxford: Oxford University Press.

Kalikow, D. N., Stevens, K. N., Elliott, L. L. 1977. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America* 61(5). 1337-1351.

Kamide, Y., Altmann, G. T. M., Haywood, S. L. 2003. The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language* 49(1). 133–156.

Mayo, L. H., Florentine, M., Buus, S. 1997. Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research* 40(3). 686-693.

McClelland, J., Elman, J. 1986. The TRACE model of speech perception. *Cognitive Psychology* 18. 1-86.

Sommers, M. S. 2005. Age-related changes in spoken word recognition. In Pisoni, D. B., and Remez, R. E. (eds.), *Handbook of Speech Perception*. New York: Blackwell. 469-493.

Takata, Y., Nábělek, A. K. 1990. English consonant recognition in noise and in reverberation by Japanese and American listeners. *Journal of the Acoustical Society of America* 88(2). 663-666.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., Sedivy, J. C. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science* 268(5217). 1632–1634.

Warren, R. M., Warren, R. P. 1970. Auditory illusions and confusions. *Scientific American* 223. 30–36.

Revisiting final-devoicing in the L2 acquisition of final voicing contrasts

Nancy C. Kula

University of Essex
nckula@essex.ac.uk

ABSTRACT

Although the facts of final obstruent devoicing in L2 phonology have been argued to follow more naturally from the emergence of the unmarked in Optimality Theory this paper shows that enriched representations at the sub-segmental level can provide an equally elegant analysis that unlike the OT case is supported by different models of first language acquisition. It is argued that final devoicing follows from the depleted licensing potential of empty nuclei that occupy the final position in Government Phonology. This analysis is sustainable under the assumption that segmental features are privative and voice is either present or absent in segmental representation. The former entails more complex voiced sounds and the latter less complex voiceless sounds. The effects of the emergence of the unmarked in the acquisition of final voicing contrasts in L2 phonology therefore follow from the interaction of licensing potential and complexity.

Keywords: Final devoicing, TETU, empty nuclei, elements, representation.

1. INTRODUCTION

Final obstruent devoicing (FD) in L2 phonology is one of the processes that provides some of the strongest evidence for the interlanguage hypothesis, namely that L2 acquisition involves grammars that are on a par with native language grammars. By this token it follows that contemporary phonological theories, otherwise created on the basis of native language grammars, should be able to handle interlanguage grammars and the accompanying attested learning patterns. Broselow, Chen & Wang (1998) and Grijzenhout and van Rooy (2000) respond to this challenge for final obstruent devoicing in L2 phonology, arguing that the devoicing facts can only be accounted for by the emergence of the unmarked (TETU) as proposed in Optimality Theory (OT). Broselow et al. (1998) further argue that non constraint-based approaches are particularly challenged by these data because the rules proposed in such approaches are deemed as emerging from data encountered in the language-learning situation. In the case of L2 FD no such learning situation is ever present. This paper aims to show that representation-based frameworks like Government Phonology (GP) can equally well handle these data and in contrast to being hindered by representations established on L1 input they are actually aided by them. The proposed analysis furthermore predicts the strategies L2 learners use in handling final voicing contrasts as well as the order in which these strategies will emerge.

2. FINAL DEVOICING IN L2 PHONLOGY

FD in L2 phonology is particularly illuminating in those cases where neither the native nor target language show FD. The typical case where FD is attested, as Grijzenhout and van Rooy (2000) (G&R, henceforth) point out, is where the native language disallows final obstruents while the target language allows final obstruents with contrastive voicing. This is exactly the case for Chinese (Wang, 1995) and Zulu (G&R, 2000) learners of English. Chinese only allows glides and nasals in final position while Zulu does not allow any codas. Zulu in contrast to Chinese shows contrastive voicing in non-final word positions.

2.1. Mandarin and Zulu data

The L2 FD facts for Mandarin Chinese are based on Wang (1995) who conducted an experiment with 10 Mandarin speakers aged between 23-30 and who had been studying EFL for about 6-7 years in their home country. The subjects had spent less than a year in an English speaking country at the time of testing. A recall production task based on memorized nonce words (with the target codas) accompanied by a definition was used. A total of 180 final stop tokens were collected. The results were as presented in table 1 below.

Table 1: Mandarin speakers production of English final voicing contrasts (Broselow et al. 1998: 264)

	Voiceless stops (*n* = 90)	Voiced stops (*n* = 90)
Correct	19% (17)	3% (2)
Epenthesis	36% (32)	36% (32)
Deletion	45& (41)	43% (39)
DEvoicing	_______	19% (17)

The results show repair strategies ordered by frequency as; first deletion, then epenthesis, then devoicing.

The Zulu FD facts as discussed in G&R (2000) are based on van Rooy (2000). Although no statistical data is given in G&R (2000) it is argued that at least two L2 phenomena are attested in coda position in Zulu; simplification of consonant clusters and final devoicing.

2.2. TETU and L2 devoicing in OT

Broselow et al. (1998) and G&R (2000) argue that final devoicing in L2 present a case of TETU. The idea is based on the assumption in OT that first language acquisition proceeds from an initial high ranking of markedness constraints over faithfulness constraints and that acquisition unfolds with the gradual demotion of markedness constraints based on input. G&R (2000) emphasize that markedness constraints for which no evidence for demotion is provided by the L1 input, remain highly ranked in the L1 grammar. Thus for both Chinese and Zulu speakers, who get no input for final devoicing, the markedness constraints NOCODA (*do not have codas*) and NOVOICECODA (*do not have voiced codas*) remain highly ranked in their L1 grammars. Assuming that L2 acquisition in OT starts from the L1 ranking then the L2 FD facts can be accounted for as a case of TETU, i.e. the emergence of NOVOICECODA by demotion to a position below a faithfulness constraint requiring the preservation of laryngeal contrasts. The analysis of L2 FD as presented in G&R (2000) is given in (3) below with the relevant general constraint definitions given in (1-2).

(1) Markedness constraints

NOVOICE: Obstruents are voiceless
NOCODA: Syllables do not have coda consonants
NOVOICECODA: Syllable codas may not contain voiced obstruents

(2) Faithfulness constraints

MAX-IO: Every segment in the input has a correspondent in the output (No deletion)
DEP-IO: Every segment in the output has a correspondent in the input (No epenthesis)
IDENT-F: Every segment in the output has identical values for a feature F as the corresponding segment in the input.

(3) TETU in L2 FD

a. **Adult Zulu/Mandarin grammar**
MAXCONS, NOVOICECODA, NOCODA » DEPVOWEL » IDENT(LAR) » NOLAR

b. **Interlanguage grammar**
NOVOICECODA » MAXCONS » DEPVOWEL » NOCODA, IDENT(LAR) » NOLAR

c. **Adult English grammar**
MAXCONS, DEPVOWEL, IDENTLAR » NOVOICECODA, NOCODA, NOLAR

The partial grammars in (3) reflect that NoCoda in the native grammar (3a) gets demoted below the faithfulness constraint DepVowel and a formerly absent ranking between NoVoiceCoda and MaxCons is established resulting in the interlanguage grammar (3b).

Notice that it is crucial that NoCoda is demoted before NoVoiceCoda as otherwise outputs with final voicing would win. There is nothing internal to the architecture of OT that would preclude such a demotion which also involves a markedness constraint being ranked below a faithfulness constraint. Another part of this analysis, which not commented on is that the constraint crucial to the emergence of FD NoVoiceCoda must first be promoted to being ranked with respect to MaxCons in the interlanguage grammar and then be demoted to achieve the target L2 grammar (illustrated by the native English grammar (3c)). This means that a constraint must both be promoted and demoted (and not just demoted) in the course of acquisition.

The OT TETU analyses of the L2 FD facts support a view of L1 acquisition where all constraints are assumed to be present in all grammars and remain accessible through out the lifespan i.e. support a Full Transfer, Full Access view of L2 acquisition. If an *emergentist* view of L1 acquisition is adopted (Fikkert & Levelt, 2008) namely, where constraints emerge from input experience then TETU could not be evoked and it is difficult to see how the constraint ranking in (3b) could be motivated. Let us turn to an alternative representational account of the L2 FD facts.

3. GP REPRESENTATIONS AND L2 FD

Government Phonology (Kaye et al., 1990) is a principles and parameters framework that assumes a set of universal principles accompanied by parameters whose settings reflect individual language grammars. The assumption is that at the initial stage parameters are set to the most unmarked setting from which L1 learners can make changes based on L1 language input. The unmarked settings assumed reflect typological markedness universals on language complexity. Thus for example, a parameter on final codas would be switched to *off* to reflect the fact that CV is the simplest syllable type.

One of the central tenets of GP is the notion of licensing and the assumption that every constituent must somehow be licensed in order to be licit in a structure. Licensing within a domain is regulated by the *licensing principle*, which states that ‘every position within a domain must be licensed apart from the head’. The source of licensing within a domain is a realised nucleus. Another principle of interest here is the *coda licensing principle*, which requires every coda to be followed by an onset that can act as its licensor. The net effect of this principle is that there are no word-final codas as these must be represented as followed by an empty nuclear position since a coda can only occur if it is followed by a licensing onset. This implies that the relevant parameter on word-final consonants is one that licenses final empty nuclei (FEN) or not. Given below in (4) are some principles and parameters of GP (and their implications) that are crucial for explaining the L2 FD facts.

(4) *GP Principles*

All positions/constituents/elements must be licensed

The optimal licensor is a realised vowel (implies empty nuclei are weak licensors)

GP Parameters

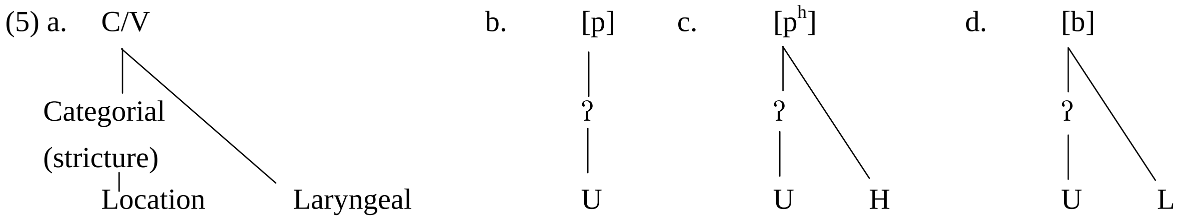
FEN are allowed (initial setting OFF)

Branching structure are allowed (initial setting OFF)

What the principles and parameters in (4) predict for first language acquisition with respect to coda complexity is the pattern CV > CVC > CVCC meaning that a realised final vowel is the most optimal. Languages like English where FEN can license a preceding onset are therefore higher on the complexity scale than Zulu that cannot. Given the case of Mandarin which licenses nasals and glides in final position and the contrast between CVC and CVCC we can conclude that complex structures are harder to license both at the syllabic and the sub-segmental level. The difference in licensing potential of realised versus empty nuclei is well established in GP. Empty positions within structure are regulated by the empty category principle (ECP) whose details we do not go into here (see Charette, 1992). The contrast between the licensing potential of realised and empty vowels is illustrated by Charette's (1992) discussion of Mongolian. Mongolian is a language that allows word internal empty nuclei to occur as long as they satisfy the ECP. Like in other languages consonant clusters must be licensed by a following vowel in Mongolian. Charette shows that a vowel that must act as licensor for a preceding consonant cluster fails to be subject to the ECP even though the conditions for this are met, i.e. a vowel that must license a consonant cluster is realised rather than be empty. In this case at least we see that a realised vowel is preferred to an empty one.

3.1. Laryngeal contrasts in GP

GP assumes that segments are composed of privative primes called *elements* that are manipulated in representing phonological contrasts and alternations. Elements differ from features in that they are acoustically defined and are interpretable at every level of derivation. I assume without much discussion the element set {A I U ʔ H L} with the main characteristics; pharyngeality, coronality/palatality, edge/stopness frication/voiceless and nasality/voicing, respectively. One way of extending this 6-member element set in order to express all the possible language contrasts is to have enriched representations so that elements may contribute different characteristics to a representation depending on the position that they assume in the phonological expression. Based on ideas developed in Kula (2002) I assume that elements are organized in element geometries as in (5a) below.



The structure in (5a) assumes two types of dependency relations within the segment; immediate dominance relations (indicated by vertical lines) and a branching dominance relation reflecting the difference in status of the elements involved. Thus while immediate dominance involves elements that form a core part of the segment, branching dependency usually hosts laryngeal features such as aspiration, glottalisation and voice, which form an outer layer of the segment. The difference between [p], [p^h] and [b] is as given in (5b-c) where aspiration and voice assume branching dependency positions where as voiceless [p] only has immediate dominance relations. The unmarked status of voicelessness is represented in (5b) by its being underspecified, i.e. the absence of a laryngeal element is interpreted as voiceless. In terms of complexity, a branching structure is more complex than a non-branching structure (on a par with the syllabic level) and will therefore in terms of acquisition be later to be acquired. How this follows in the word-final position is discussed presently.

3.2. Explaining Mandarin and Zulu L2 FD

The setting for final empty nuclei in Zulu is *off* since codas are never allowed. When the Zulu learner of English encounters word-final consonants in words like *dog* and *cat* their phonological representations are [*dogØ*] and [*catØ*], respectively. Since realized vowels are the best licensors we expect that epenthesis would be an option that would be considered in the initial stages. Once FEN can be licensed in the L2 grammar by a positive setting of the parameter on FEN the licensing abilities of FEN are also on grounds of complexity predicted to develop gradually with less complex structures preferred over more complex structures, thus (5b) over (5d). It is instructive that in Zulu, which has a contrast between lenis (aspirated) voiceless stops and fortis (ejective) voiceless stops (in addition to fortis (pain) and lenis (implosive) voiced stops), voiceless aspirated stops are never produced in L2 FD: they involve an equally complex structure (5c). In terms of licensing this correctly reveals the following licensing trajectory where *a* stands for any vowel and CC stands for complexity both at the syllabic and sub-syllabic level:

(6) *a* licences C > *a* licenses CC > Ø licenses C > Ø licenses CC

Under this understanding, the Mandarin facts raise interesting questions since Mandarin allows codas albeit a restricted set. Interestingly, only glides and the nasals {n, ŋ} are allowed in coda position. Glides in Chinese have been variously argued to be represented within nuclei (see e.g. Yip, 2006) and we will assume this position here as well. This then leaves us with the nasals. As already alluded to above, element theory assumes that the element |L| represents both nasality and voicing (Kula, 2002, Botma et al., to appear). The contribution that L makes to a segment follows from it's position in elemental representation. I follow Kula (2002) in assuming that |L| in an immediate dominance relation within the categorical gesture represents nasality in contrast to its position in a branching position where it contributes voicing. This implies that nasals in Mandarin consist of simplex non-branching structure that can be licensed by a following empty position. This means that for the Mandarin learner the setting that needs to be changed in the L2 English grammar is one that allows FEN to license more complex structures: i.e. structures involving |L| in a branching dependent position.

The proposed analysis where FD results from the failure of element licensing is in the same spirit as that developed for German in Brockhaus (1995) with the added advantage that some motivation based on complexity is provided. Thus representations based on the L1 in addition to the licensing principles and requirements of GP neatly account for the facts without ever having to make reference to a rule that is unattested in neither the L1 nor the L2. In addition, this analysis is able to stand up to emergentist views of L1 acquisition as all the structures assumed are based on L1 input.

The analysis makes the strong prediction that languages that have non-obstruent codas in final positions but which have no voicing contrasts in this position will also show devoicing effects. This prediction is supported by Spanish and Portuguese where studies have shown that despite having more contrast in codas, including continuants in the case of Spanish, these languages also show L2 FD effects. There are also, however, studies such as of Altenberg and Vago (1983/87) that show that both Hungarian and Farsi learners of English show FD despite the fact that both L1s involved have word-final voicing contrasts. These data can be interpreted in either of two ways. Either voicing is not phonologically contrastive in the two languages i.e. does not involve a branching structure or the development of the L2 does not involve full transfer of the L1 grammar in this sense supporting a Partial Transfer/Full Access hypothesis (Vainikka & Young-Scholten 1996 a, b). A full investigation of these languages would have to be made in order to ascertain the status of voicing. A final point I would like to briefly touch on is the status of perception in L2 FD.

3.3. Role of perception in L2 FD

A question that one must inevitably ask with processes such as FD is whether the L2 learners actually perceive the voicing contrasts or whether the results reflect misperception. Wissing & Zonneveld (1996) accompanied their study of L2 FD in Tswana learners of English with a perception experiment. Their finding was that the correct production of final voicing contrasts is always preceded by correct perception. Thus while only 52% of Tswana speakers attempts to produce final voiced obstruents were perceived as voiced by English speakers (the remaining 48% were perceived as voiceless), Tswana speakers classification of English speakers final obstruents as voiced or voiceless was significantly better at 70% accuracy for voiced obstruents. Voicing effects in English such lengthening of a vowel preceding voiced sounds could also not have played a role because the Tswana speakers failed to reliably control either vowel length or voicing in consonant closure. This suggests that at least in this case, the attested production problems cannot exclusively be explained by appeal to perception.

4. CONCLUSION

The emergence of FD in L2 acquisition has been accounted for within GP as resulting from the failure of an empty position to license a complex representation in a configuration where voiced segments involve branching structure at the sub-segmental level. It has been noted that L2 FD is particularly prevalent in those cases where the L1 has no codas. This has been shown to follow from the fact that codas are represented as followed by an empty nuclear position; languages where FEN are not parametrically licensed will have difficulty in producing final consonants. The privativity of features that is assumed also helps us capture the fact that voiceless obstruents, because they have no laryngeal specification, are structurally less complex than voiced obstruents and therefore unmarked. The analysis works under the assumption that empty nuclei are weak licensers, which even when licensed in a language only gradually begin to license complex structures.

5. REFERENCES

Altenberg, E.P., Vago, R.M. 1983. Theoretical implications of an error analysis of second language phonology production. *Language Learning* 33. 427–447. *Reprinted in* G. Ioup, G., Weinberger, S.H. (eds.), 1987. *Interlanguage Phonology*. Cambridge: Newbury House Publishers. 148–164.

Botma, B., Kula, N.C., Nasukawa, K. to appear. Phonological primes In: Kula, N.C., Botma, B., Nasukawa, K. (eds.), *Continuum Companion to Phonology*. London: Continuum.

Brockhaus, W. 1995. *Final Devoicing in German*. Tübingen: Niemeyer.

Broselow, E., Chen S., Wang, C. 1998. The emergence of the unmarked in second language acquisition. *Studies in Second Language Acquisition* 20. 261–280.

Charette, M. 1992. Mongolian and Polish meet Government Licensing. *SOAS Working Papers in Linguistics* 2. 275–291

Fikkert, P., Levelt, C. 2008. How does place fall into place? The lexicon and constraints in the developing phonological grammar. In: Avery, P., Dresher, B.E., Rice, K. (eds.), *Contrast in Phonology: Perception and acquisition*. Berlin, New York: Mouton de Gruyter, 231–270.

Grijzenhout, J., van Rooy, B. 2000. Second language phonology: acquisition through gradual constraint demotion. Ms. Heinrich-Heine Universität, Düsseldorf & Potchefstroome Universiteit vir CHO, South Africa.

Kaye, J.D., Lowenstamm, J., Vergnaud, J-R. 1990. Constituent structure and government in phonology. *Phonology* 7. 193 – 231.

Kula, N.C. 2002. The phonology of verbal derivation in Bemba. PhD thesis, University of Leiden, LOT dissertation series 65.

Rooy van, B. 2000. The consonants of BSAE: current knowledge and future prospects. *South African Journal of Linguistics*. Supplement 38.

Vainikka, A., Young-Scholten, M., 1996a. Partial transfer, not partial access. *Behavioural and Brain Sciences* 19. 744–745.

Vainikka, A., Young-Scholten, M. 1996b. Gradual development of L2 phrase structure. *Second Language Research* 12. 7–39.

Wang, C. 1995. The acquisition of English word-final obstruents by Japanese speakers. PhD thesis , State University of New York at Stony Brook.

Wissing, D., Zonneveld, W. 1996. Final devoicing as a robust phenomenon in second language acquisition: Tswana, English and Afrikaans. *South African journal of Linguistics, Supplement* 34. 3–24.

Yip, M. 2006. Casting doubt on the onset-rime distinction. *Lingua* 113.8. 779–816

Features, cues, and syllable structure in the acquisition of Russian palatalization by L2 American learners

Vladimir Kulikov

University of Iowa
vladimir-kulikov@uiowa.edu

ABSTRACT

The current study investigates the interaction between features, phonetic cues, and syllable positions in L2 acquisition. It tests the acquisition of Russian palatalized consonants by 40 beginning and advanced American learners. Results in particular groups of consonants are problematic for both 'feature-based' and 'cue-based' theories of generative SLA. Both proficiency groups show very high performance in an onset condition despite that fact that English lacks a contrastive V-place node. Coronal sounds are argued to provide more prominent cues (e.g. release burst) than labial sounds (e.g. Padgett 2001), yet the contrast in labials is discriminated at a higher rate. The results suggest that acquisition of a new contrast is sensitive both to position of segments and combination of features. Learners filter L2 sounds through L1 grammar; the constraint against a combination of coronal segments and [j] (*Coronal+j) in American English affects perception of palatalized coronal segments in L2. In a coda position, there are no active restrictions on such combinations in L1. L2 learners can also use additional phonetic cues (e.g. the 'robust' feature [+strident]) to facilitate perception in coronal stops.

Keywords: generative SLA, phonological features, phonetic cues, palatalization.

1. INTRODUCTION

In generative SLA, it has been argued that acquisition of a phonological contrast is governed solely by features available in L1 (Brown 1998) or by a combination of phonological features and 'robust' acoustic cues (Archibald 2007, 2009). Brown (1998) claims that L2 learners cannot acquire a phonological contrast in L2 if the feature of contrast is not contained in their native grammar. Archibald (2009) argues that L2 learners can use phonetic cues to perceive a new contrast even if their L1 lacks a contrastive feature.

Although Brown's model predicts that phonological features in L1 must determine perception of the L2 acoustic signal, it does not explain what acoustic cues must be filtered out and what cues can be processed by learners. This issue is not fully resolved in Archibald's model, either. If L2 learners can acquire sounds that are perceptually 'robust', to what extent does the L1 grammar allow perception of such sounds? Allowing phonology to govern perception of new sounds must have natural limitations, e.g. phonological constraints in the sense developed in Optimality Theory (Prince and Smolensky 2004).

This paper tests the two hypotheses in the case of acquisition of Russian palatalization by American learners. In Section 2, I present phonetic and phonological properties of Russian palatalized sounds. Section 3 introduces the hypothesis and predictions; Section 4 describes the experiment and presents the results. The discussion and conclusion are given in Section 5.

2. RUSSIAN PALATALIZATION: PHONOLOGICAL REPRESENTATION AND PHONETIC CUES

The palatalization contrast is one of the most important phonological contrasts in Russian. Most consonants in Russian are articulated as either non-palatalized ('hard') or palatalized ('soft'). This contrast can be phonologically represented as a secondary V-place node with the feature [coronal] (Hume 1994). Russian palatalized consonants have this node in their underlying feature geometric representation (Fig. 1a), whereas English consonants, which do not have contrastive palatalization, lack this node (Fig. 1b).

Palatalized consonants are produced with an obligatory co-articulatory gesture of the front of the tongue raised toward the hard palate; velarized consonants are produced with the back of the tongue raised toward the velum (Bolla 1981). The contrast is maintained in onsets and codas (1). Unlike many phonological contrasts, the palatalization contrast in Russian is never neutralized word-finally (Avanesov 1972).

(1) /tok/ 'current' /t^jok/ 'flowed'
/krof/ 'shelter' /krof^j/ 'blood'
/tsel/ 'intact' /tsel^j/ 'target'

Figure 1: Phonological representation of (a) palatalized and (b) non-palatalized sounds.

a.
C
|
C-place
… V-place
|
[coronal]

b.
C
|
C-place
|
…

Perception of the palatalization contrast requires specific cues. According to Kochetov (2006: 114-116), the formant transition between a vowel and a consonant and release burst are of the greatest importance among the acoustic properties of palatalized consonants. Vowels adjacent to a palatalized consonant were found to have a higher F2 than vowels adjacent to a velarized consonant (e.g. Jacobson et.al. 1963; Fant 1970; Bondarko 1977). This transition is often represented as a *j*-glide between a consonant and a vowel. However, proper timing between the two articulatory gestures is crucial. Russian palatalized consonants are pronounced with *simultaneous* tongue raising. Moreover, Russian has a minimal contrast between palatalized consonants and consonant+j clusters in onsets (2).

(2) /s^jel/ 'sat down' *3sg.masc.* /s^jjel/ 'ate' *3sg.masc.*
/pol^ju/ 'weed' *1sg.* /pol^jju/ 'water' *Fut.1sg.*

Release is an important condition in the production of Russian stops. Unlike in English, stops are most often released in Russian even in word-final position (Bondarko 1977). Palatalized coronal stops have a more distinct audible release than velarized stops. The feature that characterizes release of palatalized coronal stops in Russian is affrication (Fant 1970). According to Kochetov (2006: 122, 129), palatalized [t^j] has a release 46% longer than velarized [t] when pronounced before another stop both word-internally and across a word boundary. Both the F2 transition and release burst were found to be critical for identification of palatalized stops in a coda by native speakers. However, native speakers were more sensitive to differences in release burst than to F2 transition in distinguishing between word-final palatalized and non-palatalized [t] (Kochetov 2006: 126, 132).

Other studies (e.g. Steriade 1997; Padgett 2001) also suggest that palatalized coronal sounds provide, in general, more prominent cues (e.g. release burst) than labial sounds. According to Padgett (2001: 209), the palatalization contrast is less common cross-linguistically in labials than in coronals because the contrast in labials is perceptually less prominent. Very few studies looked into L2 acquisition of palatalization. However, results reported in Larsson-Hall (2004) show that while some contrastive pairs with labials (e.g. [f]-[f^j]) are indeed difficult and were acquired at a low rate (84%), other contrastive pairs with labials (e.g. [m]-[m^j], [p]-[p^j]) are perceived with 96% accuracy, even by beginners. I am not aware of studies that specifically tested perception of palatalized coronals vs. palatalized labials by L2 learners. Nevertheless, several issues have to be addressed if we are going to predict the results of acquisition of the palatalization contrast in Russian by American learners. First, L1 speakers and L2 learners may not weight the same perceptual cues equally. In coda position, release burst is a more important cue for native speakers than F2 transition. American L2 learners of Russian may not be as sensitive to release burst in final stops as native speakers. Final stops are often pronounced without release in American English (Ladefoged 2005); therefore, learners may find F2 transition to be a more useful cue.

Second, the same phonetic cues may not be weighted equally pre- and post-vocalically due to parsing. While release burst is more important in the perception of palatalization than F2 transition in codas, F2 transition may be a major cue to distinguish between palatalized and non-palatalized consonants in onsets Recent studies (e.g. Kulikov 2009) also suggest that L2 learners are sensitive to constraints that ban some combinations of sounds or features in L1 grammar. English does not allow [tl]/[dl] in onsets; consequently, American learners of Russian cannot distinguish between [t] and [d] in such word-initial clusters in Russian whereas they have little difficulty in perceiving the voice contrast in single segments. Similarly, American English does not allow [j] after a coronal (*tune* [tun] *AmE*. vs. [tjun] *BrE.*); therefore, American L2 learners may have difficulties perceiving the palatalization contrast in Russian coronals.

3. PREDICTIONS

The two models of acquisition differ in their predictions about the results of acquisition. Brown's 'feature governed' model will predict that acquisition of Russian palatalization will be difficult for American learners. The V-place node is absent in underlying representation of American sounds; L2 learners will fail to acquire the contrast between palatalized and non-palatalized sounds. This model is mute on differences in the perception of palatalized sounds before and after a vowel. F2 transition affects vowel quality in both cases. It is seems hard to predict which transition, CV or VC, will facilitate L2 learners' perception and which will not. Archibald's 'features+cues' model predicts that some perception cues (e.g. a robust phonetic feature [+strident] in affricated palatalized coronals) will facilitate acquisition of the palatalization contrast in some groups of segments (coronal stops). Consequently, we must expect higher perception rate in coronal stops than in all other categories. However, it is hard to predict whether release burst will equally facilitate perception of prevocalic and post-vocalic stops. Neither model makes clear predictions about differences between sonorants and obstruents in the perception of palatalization.

4. EXPERIMENT

4.1. Participants and design

In order to determine which model of acquisition more accurately explains the results of the acquisition of the palatalization contrast, I tested 40 participants who were American students of Russian at the University of Iowa. 27 beginning learners and 13 advanced learners participated in the experiment. Beginner learners studied Russian for two months prior to testing and had not been previously exposed to Russian; advanced learners had three years of instruction prior to testing. The learners in both groups had native speakers of Russian as instructors and were exposed to naturalistic input. Nine native speakers of Russian were used as a control group. They were students and professionals who had lived in the US for 2-6 years and maintained everyday communication in Russian with friends and family.

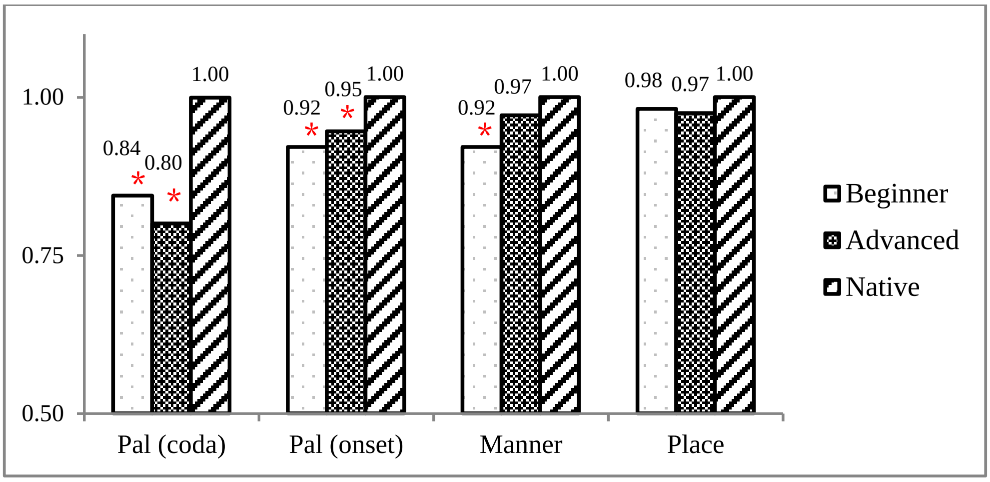
The stimuli in the AX discrimination task included minimal pairs with contrastive non-palatalized and palatalized sounds in 2x2 blocks. Contrastive pairs included labial obstruents and nasals [p]-[p^j], [b]-[b^j], [f]-[f^j], [v]-[v^j], [m]-[m^j], as well as coronal obstruents, nasals, and liquids [t]-[t^j], [d]-[d^j], [s]-[s^j], [z]-[z^j], [n]-[n^j], [r]-[r^j], [l]-[l^j]. The pairs were tested in word-initial (onset) and word-final (coda) positions. In addition, minimal pairs that contrasted in the manner and place of articulation were used as control categories. According to Brown (1998: 177-178), learners of a language that has the same contrastive feature as their L1 should not have difficulties in the acquisition of this contrast. Both Russian and English distinguish between stops and fricatives, and between labial, coronal, and dorsal places of articulation; however, English does not have a voiceless velar fricative [x], which is specified as [continuant] and [dorsal]. Contrastive pairs with the sounds [f]-[x] and [s]-[x], which minimally differ in a place feature ([dorsal] vs. [labial] and [coronal]), and with the sounds [k]-[x], which minimally differ in the feature [continuant], were used to test perception of familiar contrasts. The stimuli were randomized and mixed with distractors. Two versions of the test in the reverse order were randomly administered to the participants to reduce the possibility of fatigue effects.

4.2. Results

4.2.1. Contrasts

Figure 2 presents the results of the three contrasts: palatalization, manner, and place. As predicted, the manner and place contrasts, implemented by the familiar features [continuant] and [dorsal], were perceived with high accuracy, whereas the palatalization contrast was difficult for L2 learners. A repeated measures ANOVA with (3) Contrast and (3) Proficiency levels showed a significant effect for Contrast [$F(2,92)=14.04$, $p < 0.001$], Proficiency [$F(2,46)=10.55$, $p<0.001$], and interaction between Contrast and Proficiency [$F(4,92)= 4.34$, $p < 0.01$]. A *post-hoc* Tukey test showed that both beginner and advanced learners performed significantly differently from the control group ($p < 0.01$), but they were not significantly different from each other.

Figure 2: Mean accuracy in perception of the Palatalization, Manner, and Place contrasts by beginner and advanced learners (An asterisk indicates significant difference from other columns in a group).



For the Place contrast, no significant difference in perception was obtained [$F(2,46) = 1.034$, $p = 0.364$]. L2 learners perceived the contrast implemented by the familiar place features in a native-like fashion (98%). The effect was significant for the Manner contrast [$F(2,46) = 3.05$, $p < 0.05$] and the Palatalization contrast [$F(2,46) = 22.51$, $p < 0.001$]. *Post-hoc* Tukey tests showed that beginners and advanced learners performed differently on the Manner contrast, but there was no significant difference between the two proficiency groups on the Palatalization contrast. These results suggest that there were no problems with Manner and Place contrasts for the advanced learners of Russian. Beginning learners had problems with all contrasts other than Place. Both proficiency groups of L2 learners also had difficulties with perception of the Palatalization contrast.

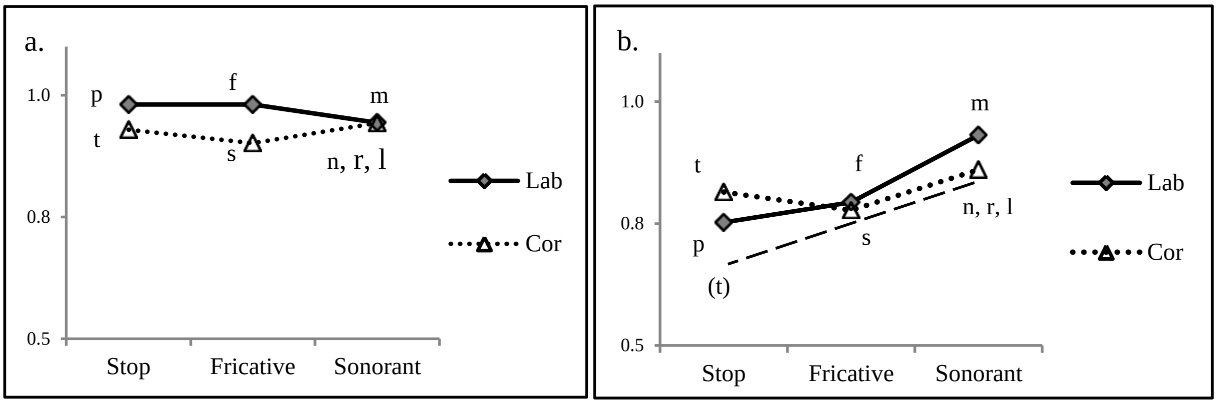
4.2.2. Position in a syllable: onset vs. coda

In order to determine whether L2 learners' perception of new sounds was affected by syllable position, a factorial 2 (Proficiency) x 2 (Position) x 3 (Contrast) ANOVA was performed. The main effect of Proficiency was not significant ($p = 0.35$). A significant main effect of Contrast [$F(2,628)= 20.92$, $p < 0.001$] and Position [$F(1,628) = 20.71$, $p < 0.001$], and a significant interaction between Contrast and Position [$F(2,628)=11.12$, $p<0.001$] were obtained. A *post-hoc* Tukey test showed that L2 learners' performance for the Palatalization contrast was significantly different from their performance for the Place and Manner contrasts. Learners did not have difficulties with new contrastive sounds in either syllabic position when the feature of contrast was present in L1. In contrast, perception of a new contrast was affected by the position of a contrastive sound when the contrastive feature was new. A significant effect of syllable position was found only for the Palatalization contrast [$F(1,478)=107.59$, $p<0.001$]. The perception rate of the Palatalization contrast in a coda was low in beginning (84%) and advanced learners (80%) in comparison with perception in an onset position (92% and 95%). These results suggest that onset position provides better cues for Russian palatalized consonants, which is consistent with the 'Licensing by Cue' hypothesis (Steriade 1995).

4.2.3. *Place and manner of articulation of palatalized sounds*

Figure 3 presents results for the Palatalization contrast in separate categories of sounds. The effect of place and manner of articulation of palatalized sounds on perception was tested using a factorial 2 (Place) x (3 (Manner) ANOVA. In onset position, only a significant main effect of Place was obtained [$F(1,234)=14.76$, $p < 0.001$]. L2 learners had more difficulties perceiving coronal sounds (92%) than labial sounds (97%). The difference was more distinct in fricatives (90% vs. 98%) and stops (92% vs. 98%) than in sonorants.

Figure 3: Mean accuracy in perception of the Palatalization contrast in onsets (a) and codas (b) by L2 learners split by place and manner of articulation. The dashed line in (b) represents the tendency for coronal stops based on results for fricatives and sonorants.



In coda position, the effect of place of articulation was not significant ($p=0.587$). The test found a significant effect of manner of articulation [$F(2,233)=13.88$, $p < 0.001$] and a significant interaction between Place and Manner [$F(2,233)=3.92$, $p < 0.05$]. L2 learners had greater difficulties perceiving the palatalization contrast in coronal fricatives (77%) than in labial fricatives (80%), and in coronal sonorants (86%) than in labial sonorants (93%). Only the contrast between coronal stops was perceived at a higher rate (81%) than the contrast between labial stops (75%). The results suggest that position in the syllable and the category of a sound play an important role in perception of Russian palatalization. Place, manner, and position in the syllable can explain a significant proportion of variance in perception scores ($R^2=0.211$, $p < 0.001$). Each of the three independent variables brings a significant contribution to the model ($p < 0.05$).

5. DISCUSSION AND CONCLUSION

The goal of this paper was to evaluate two models of the acquisition of a phonological contrast: a 'feature-based' model proposed by Brown and a 'cue-based' model proposed by Archibald. The analysis shows that nether model can fully account for the results of the experiment; however, the 'cue-based' model provides a more adequate explanation.

The overall results support Brown's 'feature-based' model. A new contrast implemented by a phonological feature not present in the L1 grammar is indeed more difficult to acquire than new contrasts with features present in the L1 grammar. Russian palatalization was difficult for beginning and advanced learners, whereas the sound [x] does not seem to be hard for American learners and is accurately perceived at a near-native level. However, other results cannot be explained within the 'feature-based' model. One of the issues which Brown's hypothesis does not address is the coda-onset asymmetry. It seems that some contrasts in languages can be easily perceived only in restricted prosodic positions. Japanese L2 learners in Brown (1998) failed to discriminate between English [l] and [r] before a vowel; however, they had near native accuracy perceiving this contrast in a coda position. Contrasts with palatalized consonants, on the contrary, seem to be easy to perceive in a pre-vocalic position. Current results show that L2 learners perceived [p]-[p^j], [f]-[f^j], and [m]-[m^j] contrasts with near-native 98% accuracy, which is consistent with results reported in Larson-Hall (2004). If the results of acquisition are determined solely by the absence of a contrastive feature, it is not clear what ensures perception of a contrast in some syllabic positions.

Archibald's 'cue-based' model of acquisition provides answers to some of these questions. Certain phonetic cues are argued to be perceptually robust, which can override restrictions on perception imposed by the absence of a contrastive feature. The 'robust' feature [+strident], which specifies affricates, seems to be at play when American learners acquire Russian palatalization. Recall that [t]-[t^j] were perceived with a higher accuracy (81%) than labial stops (75%) in coda position where Russian coronal stops have markedly stronger affrication. The overall tendency (this is marked with a dashed line in Fig. 3b), however, predicted a very low rate around 60% for coronal stops in coda position. I argue that L2 learners used affrication of word-final palatalized [t] as a 'robust phonetic' cue.

Both models cannot, nevertheless, explain why L2 learners had more difficulties with coronal sounds than with labials in both onset and coda positions, and with word-initial coronal stops compared to word-initial labial stops. The 'cue-based' approach should predict that better cues in L1 must, as well, facilitate perception in L2. However, this does not seem to be an accurate prediction. The results suggest that L2 learners weight cues differently than native speakers. For American learners of Russian, the F2 transition in palatalized consonants appears to be a more important cue than the release burst. This can explain 1) why the results of acquisition are distributed along a sonority scale in a coda position where accuracy of perception gradually increases from stops and fricatives to sonorants, and 2) why labial stops, which are less noisy than coronal stops, tend to be perceived with less difficulty than the latter. Thus, higher sonority facilitates L2 learners' perception of vocalic properties in palatalized sounds.

Finally, L1 phonotactic constraints may intervene in L2 perception. Although coronals in general have better acoustic cues, Russian palatalized coronals provide an additional difficulty for American learners. Recall that speakers of American English decompose the phonological structure of palatalized sounds into a 'segment+j' sequence (c.f. the 'Feature Reassembly' hypothesis (Lardiere 2007) in morphology). English bans clusters of coronal segments with [j] in onsets; therefore, American learners of Russian have difficulties with the perception of such sounds. It seems, however, that in this particular case the phonological restriction is partly neutralized by the phonetic robustness of palatalized coronals. Hence, L2 learners can perceive palatalized coronals in Russian, although the perception rate for coronals is lower than that for labials. In other cases (e.g. Kulikov 2009), phonotactic constraints in L1 can block perception of a phonological contrast in L2.

6. REFERENCES

Archibald, J. 2007. The role of acoustic prominence in L2 phonology: Against the deficit model. GASLA 9. Iowa City, IA. May 18–20.

Archibald, J. 2009. Phonological feature re-assembly and the importance of phonetic cues. *Second Language Research* 25(2). 231–233.

Avanesov, R.I. 1972. *Russkoie literaturnoie proiznoshenie*. Moscow: Prosveshchenie.

Bolla, K. 1981. *A conspectus of Russian speech sounds*. Cologne: Bölau.

Bondarko, L.V. 1977. *Zvukovoi stroi sovremennogo russkogo iazyka*. Moscow: Prosveshchenie.

Brown, C. 1998. The role of the L1 grammar in the acquisition of L2 segmental structure. *Second Language Research* 6(13). 287–317.

Fant, G. 1970. *Acoustic theory of speech production: with calculation bases on X-ray studies of Russian articulations*. The Hague: Mouton.

Hume, E.V. 1994. *Front vowels, coronal consonants, and their interaction in nonlinear phonology*. Garland Publishing.

Jacobson, R., Fant, G, Halle, M. 1963. *Preliminaries to speech analysis: the distinctive features and their correlates*. Cambridge: The MIT Press.

Kochetov, A. 2006. Testing Licensing by Cue: A case of Russian palatalized coronals. *Phonetica* 63. 113–148.

Kulikov, V. 2009. Features, acoustic cues, and prosodic positions in the L2 acquisition of Russian. LSA. San Francisco, CA. January 8–11.

Ladefoged, P. 2005. *A course in phonetics*. (5th ed.) Boston: Thompson Wadsworth.

Lardiere, D. 2007. Feature assembly in second language acquisition. In J. Liceras, H. Zobl, & H. Goodluck (eds.), *Role of features in second language acquisition*. Lawrence Erlbaum Associates. 107–140.

Larson-Hall, J. 2004. Predicting perceptual success with segments: a test of Japanese speakers of Russian. *Second Language Research* 20. 32–76.

Padgett, J. 2001. Contrast Dispersion and Russian Palatalization. In Hume, E., and K. Johnson (eds.), *The role of speech perception in phonology*. Academic Press. 187–218.

Prince, A., Smolensky, P. 2004. *Optimality Theory: Constraint interaction in generative grammar*. Blackwell Publishing.

Steriade, D. 1997. *Phonetics in phonology: the case of laryngeal neutralization*. Ms. UCLA, Los Angeles.

Factors in Early L2 Learners' Perception of Nonnative Sounds

Shinsook Lee

Korea University
leesseng@korea.ac.kr

ABSTRACT

This study investigated early Korean L2/EFL learners' perception of English fricative sounds in order to find out whether learners' different L2 experience influenced the way fricatives were perceived. For this purpose, three groups of children with different L2 experience took part in the study: 20 Korean EFL students, 21 bilingual returnees currently enrolled in an English-immersion program and 19 English-dominant bilinguals residing in the U.S. Their mean age was 9. They took a discrimination test with 48 English nonce words containing voiceless fricatives with four places of articulation (labiodentals, interdentals, alveolars, alveopalatals) before the front vowel /i/ and back vowels /ɔ/ and /a/.

The results showed that perception of fricative sounds was affected by participants' different English-language experience. This is because EFL students outperformed Returnees or English-dominant bilinguals on discriminating non-identical stimuli but the latter groups were almost native-like in discriminating identical stimuli. The results also revealed that overall participants with more L2 experience surpassed those with less L2 experience in discrimination and that identical stimuli were better perceived than non-identical ones. Moreover, there was some effect of place of articulation and there was also a back vowel advantage for some of the target sounds

Keywords: fricatives, L2 experience, stimulus type, place & vowel effects

1. INTRODUCTION

It has been well-established that native language (L1) influences the way sounds in a second language (L2) or a foreign language (FL) are perceived (Best 1995; Flege 1995). In addition to L1 influence, other factors such as age of the onset of L2/FL acquisition, amount of L2 use, length of exposure to L2 input, target sounds' place of articulation and neighbouring sounds feature in the perception of nonnative speech sounds (Sundara and Polka 2008).

However, the interaction among the factors above in the perception of nonnative phonemic contrasts by early language learners has not been much investigated. Thus, this study examined early Korean-English bilinguals' and early Korean EFL learners' perception of English fricative sounds in two vowel contexts.

Specifically, the study explored the following questions: 1. Does learners' different L2 experience influence perception of L2/FL sounds? 2. If so, do the learners show different error patterns depending on their L2 experience? 3. Do target consonants' places of articulation and the following vowels affect the way target sounds are perceived? In order to investigate the research questions, total 48 stimuli were presented in an AX discrimination task.

2. RESEARCH DESIGN

2.1. Participants

Three groups of young children participated in the study: 20 Korean EFL children who learned English in a classroom setting in Korea (hereafter, called EFL students), 21 bilingual returnees currently enrolled in an English-immersion program (hereafter, called Returnees), and 19 English-dominant bilinguals residing in the U.S. (hereafter, called ED bilinguals). Their mean age was 9. The present study recruited three groups of children with different English-language experience in order to examine whether L2 experience affected the way fricative sounds were perceived. Participants' background information is provided in Table 1.

Table 1: Participants' background information

Group	Mean Age (years)	Mean Initial Age of Exposure to English (years)	Mean Length of Learning English (years)
Korean EFLstudents	9.0	7.6	2.3
Bilingual returnees	9.8	4.1	5.8
English-dominant bilinguals	8.8	1.9	6.9

An ANOVA conducted on the mean comparisons between subject groups revealed that each group was significantly different from one another with respect to length of English learning as well as initial age of exposure to English (all p<.05), which suggests that each group had different English-language experience.

2.2. Stimuli

Twenty-four disyllabic English nonce words with a voiceless fricative onset were first constructed. The syllabic structure of the words was either CVC.CVC or CVC.CVCC stressed on the first syllable. Based on the twenty-four nonce words, two sets of stimuli were created: 24 identical pairs (e.g. *fiktom-fiktom*) and 24 non-identical ones (e.g. *fiktom-thiktom*). Thus, the total number of stimuli was 48. The target fricatives with four places of articulation —labiodental /f/, interdental /ɵ/, alveolar /s/, and alveopalatal /š/—were presented in two vowels contexts. In particular, the high front vowel /i/ and the low back vowel /a/ and the mid-low back vowel /ɔ/ were used (e.g. *findert-findert*, *findert-thindert*, *farbin-farbin*, *farbin-tharbin*). This study specifically examined these fricatives which differ in terms of place of articulation, since Korean has only two alveolar fricative sounds, the lax /s/ and the tense /s'/ and the alveopalatal fricative /š/ occurs as an allophone of /s/ before /i/ in Korean. Thus, Korean EFL learners often palatalize the /s/ sound in words like *sip*, which may cause some perceptual difficulty in the distinction between /s/ and /š/. Further, first language learners who acquire English as their mother tongue often confuse /f/ with /ɵ/, while Korean EFL learners have difficulty distinguishing between /s/ and /ɵ/. Consequently, it is expected that learners may show different error patterns in perceiving English fricative sounds depending on their L1/L2 experience.

2.3. Procedure

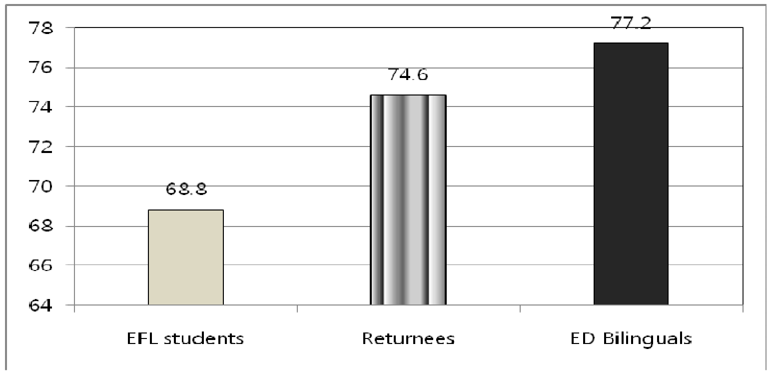
A discrimination task was administered using E-prime 2.0. The presentation order of the 48 stimulus items was randomized across participants. The participants sat at a computer in a sound-treated room and wore headphones. In particular, they were asked to press the corresponding key (i.e., same or different) as fast as possible on the keyboard after a stimulus presentation. Participants were given a maximum of 3s to respond in each trial. The next trial began after a 2s inter-trial interval.

3. RESULTS

3.1. Effects of group and stimulus type

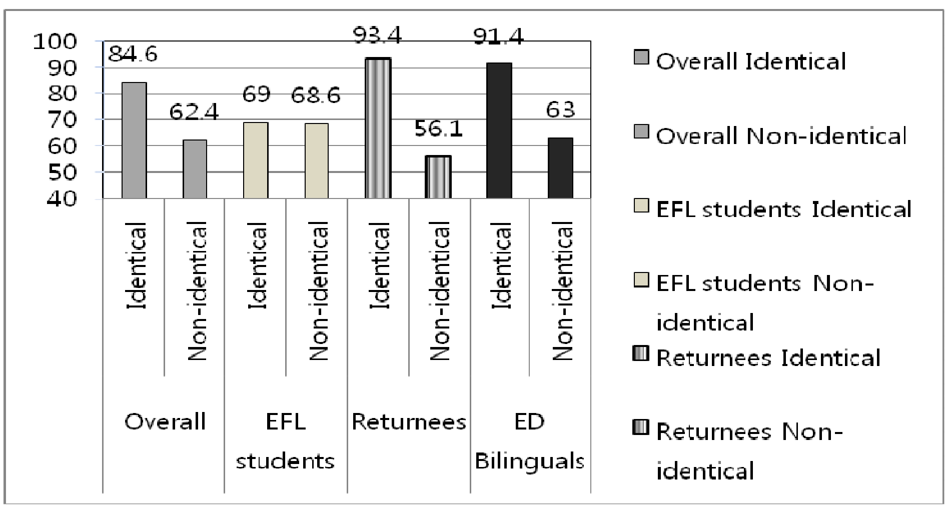
The overall results were analyzed in terms of stimulus type, as there were 24 identical (e.g. *findert-findert*) and non-identical (e.g. *findert-thindert*) stimuli each. A mixed ANOVA was conducted on mean accuracy and reaction times (henceforth, called RTs) with group (i.e. EFL students, Returnees, ED bilinguals) as a between-subjects factor and stimulus type (i.e. identity vs. non-identity) as a within-subjects factor. The results on mean accuracy showed that there was a main effect of group (F(2,57)=5.574 , p=.006), as ED bilinguals performed slightly better than Returnees, who in turn performed much better than EFL students, as shown in Figure 1.

Figure 1: Mean rates (%) of correct perception by group



Post hoc comparisons (Bonferroni) showed that the difference in correct percentages between EFL students and ED bilinguals was significant (p=.006). The effect of stimulus type was also significant ($F(1,57)=110.796$, $p<.0001$) and there was a significant interaction between stimulus type and group ($F(2,57)=28.481$, $p<.0001$). This indicates that the way the stimuli were perceived was influenced by the participant group, even though identical stimuli were overall perceived much better than non-identical ones (84.6% vs. 62.4%). In particular, EFL students' mean rate of accuracy between identical and non-identical stimuli was almost the same, unlike that for other participant groups, as shown in Figure 2. Unexpectedly, EFL students outperformed not only Returnees but also ED bilinguals in perceiving non-identical stimuli, even though the opposite pattern was obtained for identical stimuli.

Figure 2: Correct percentages of identical vs. non-identical stimuli by group



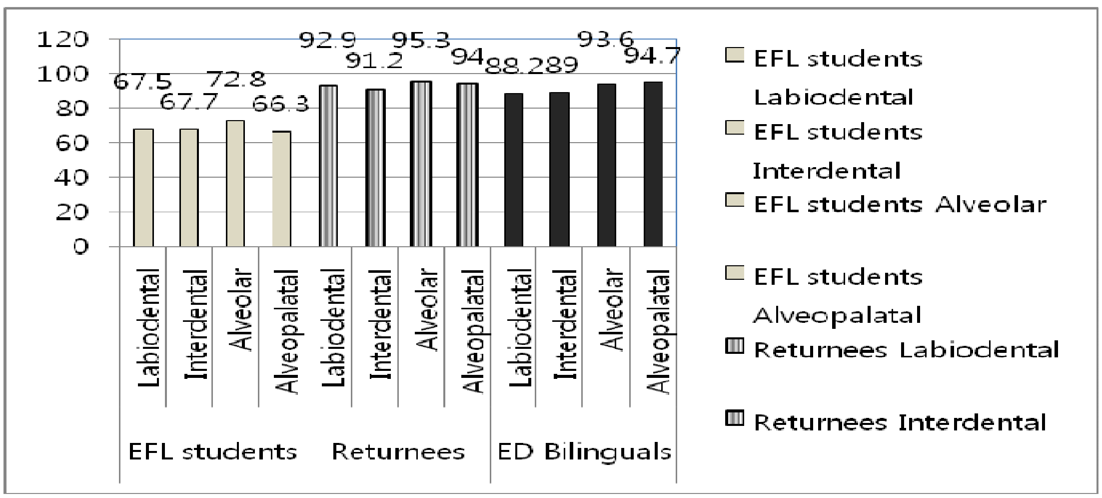
The results on mean RTs revealed that there was no main effect of group ($F(2,57)=.037$, $p>.05$) and the interaction between stimulus type and group was not significant ($F(2,57)=.039$, $p>.05$). Yet, there was a significant effect of stimulus type ($F(1,57)=7.928$, $p<.05$), as identical stimuli were responded much faster than non-identical ones (467ms vs. 525ms) across all the participant group.

3.2. Effects of place of articulation and vowels

The general results were also analyzed in terms of fricatives' places of articulation and the following vowels. As for the place of articulation, fricatives with four places of articulation were first subdivided into identical and non-identical stimuli. More specifically, there were four places of articulation for identical stimuli: labiodentals (e.g. *fiktom-fiktom*), interdentals (e.g. *thiknet-thiknet*), alveolars (e.g. *siknet-siknet*), and alveopalatals (e.g. *shipkin-shipkin*). Yet, there were three places of articulation for non-identical stimuli. This is because labiodentals were matched with interdentals (e.g. *fiktom-thiktom*), interdentals with alveolars (e.g. *thiknet-siknet*), and alveolars with alveopalatals (e.g. *sipkin-shipkin*). Consequently, the effect of the place of articulation was calculated separately for identical and non-identical stimuli.

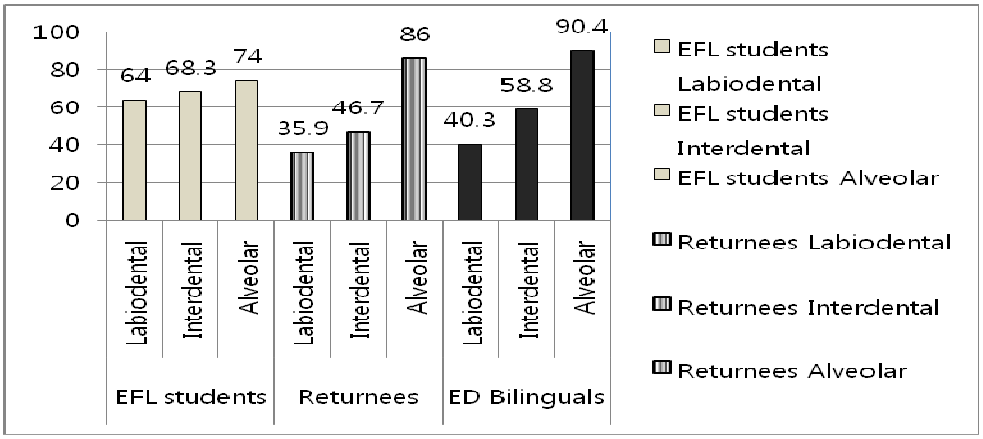
For identical stimuli, a mixed ANOVA was run on mean accuracy, with group as a between-subjects factor and place of articulation as a within-subjects factor. The effect of group was significant ($F(2,57)=29.620$, $p<.0001$). But there was no main effect of place ($F(3,171)=1.289$, $p>.05$) and the interaction between place and group was not meaningful ($F(6,171)=.370$, $p>.05$). This is due to the fact that Returnees and ED bilinguals outperformed EFL students in perceiving identical stimuli regardless of place of articulation (EFL students: 68.6%, Returnees: 93.4%, ED bilinguals: 91.4%). However, the way fricatives were perceived seems to be affected by the participants' L2 experience, even though there was no significant interaction between place of articulation and group. Namely, EFL students perceived alveolar fricatives better than other places of articulation, while ED bilinguals perceived alveolars and alveopalatals better than labiodentals or interdentals. Returnees showed the same pattern as ED bilinguals, even though the former performed slightly better than the latter, as shown in Figure 3. The results on RTs showed that there was no significant effect of group or place. The interaction between place and group was not significant, either (all $p>.05$).

Figure 3: Correct percentages of identical stimuli by group and place of articulation



Now, let us consider the effect of place of articulation for non-identical stimuli. A mixed ANOVA run on mean accuracy showed a main effect of group ($F(2,57)=5.816$, $p<.05$), as EFL students outperformed ED bilinguals, who in turn surpassed Returnees (EFL students: 68.8%, Returnees: 56.2%, ED bilinguals: 63.2%). Bonferroni pair-wise comparison revealed that the difference in correct percentages between EFL students and Returnees was significant ($p=.004$). There was also a main effect of place ($F(2,114)=58.997$, $p<.0001$) and the interaction between place and group was significant ($F(4,114)=8.360$, $p<.0001$). In particular, the participants perceived the difference between alveolars and alveopalatals much accurately (83.5%) than that between interdentals and alveolars (57.9%). The participants were below chance in perceiving the difference between labiodentals and interdentals (46.7%). Moreover, the participants' L2 experience was related to the perception of fricatives, as shown in Figure 4. That is, ED bilinguals performed better than Returnees, who in turn outperformed EFL students in discriminating between alveolars and alveopalatals. However, EFL students surpassed ED bilinguals and Returnees in differentiating between other places of articulation. In fact, both ED bilinguals and Returnees were below chance in discriminating between labiodentals and interdentals. Returnees were also below chance in distinguishing between interdentals and alveolars, while ED bilinguals were above chance.

Figure 4: Correct percentages of non-identical stimuli by group and place of articulation



The fact that both ED bilinguals and Returnees had difficulty differentiating labiodentals from interdentals seems to suggest that they show the same error pattern as young L1 English listeners. Unexpectedly, however, Returnees and ED bilinguals were also poor at discriminating between interdentals and alveolars, even though it has been well-documented that Korean EFL learners have most difficulty with the distinction between interdentals and alveolars.

As for the results on mean RTs, there was no significant effect of group or place. The interaction between place and group was not significant, either (all p>.05).

Now, let us move onto the effect of the following vowels on the perception of fricatives. Returnees had higher accuracy for alveolars located before back vowels than before front vowels for identical stimuli. Likewise, ED bilinguals showed back vowel benefits for interdentals. As for non-identical stimuli, the participants tended to show back vowel benefits for labiodental vs. interdental pairs and for interdental vs, alveolar pairs in terms of accuracy and RTs. Therefore, the results indicate that the participants overall had more difficulty perceiving fricative sounds located before the front vowel /i/ than before back vowels, especially when the stimuli were non-identical, which is consistent with the findings of the previous study (Gay 1970).

4. FINDINGS AND IMPLICATIONS FOR MODELS OF L2 SPEECH PERCEPTION

Let us consider the results of the experiment in terms of each participant group. As for EFL students, they were not much different with respect to perceiving identical vs. non-identical stimuli, since their mean rate of accuracy for both types of stimuli was around 68%. EFL students' overall low accuracy seems to indicate that they had not formed separate L2 categories for the target fricative sounds due to their short experience with the English-language. Further, their performance for the identical alveolar /s/ vs. /s/ pairs and non-identical /s/ vs. /š/ pairs was much better than that for other places of articulation. This seems to indicate that EFL students might not be affected by the Korean palatalization process, as many of them were able to discriminate /s/ and /š/ even before the front vowel /i/ (73% before front vowels vs. 75% before back vowels). However, their low accuracy for the identical /š/ vs. /š/ pairs compared to that of the identical /s/ vs. /s/ pairs seems to suggest that L1 interference is evident. Namely, many of EFL students might not have formed a separate category for the English /š/, as [š] is a phonetic variant of /s/ in Korean, and thus they might overall have more difficulty perceiving the /š/ than the /s/. Moreover, the results seem to reveal that not all predictions of the SLM (Flege 1995) and PAM/PAM-L2 (Best 1995, Best and Tyler 2007) hold true. That is, according to the SLM and PAM/PAM-L2, new L2 phonetic categories can be formed if phonetic/acoustic and/or articulatory discrepancies between an L2 sound and the closest L1 sound are spotted. The English alveolar sound /s/ is very similar to the Korean sound /s/ in terms of phonetic/acoustic and articulatory properties, and thus it is expected that English /s/ would pose more problems to EFL students, which is not the case.

In addition, EFL students' perception of both identical and non-identical labiodentals and interdentals was around 67%. This seems to suggest that EFL students would notice that English /f/ and /ɵ/ are rather different from Korean sounds, yet they might not have established separate categories for these sounds due to the sound's similarity to Korean /p^h/ and /s'/ or /s/, respectively, in terms of air release and place properties.

Now, let us turn to Returnees and ED bilinguals, who overall exhibited almost the same perceptual patterns. For non-identical stimuli, Returnees' discrimination of labiodentals vs. interdentals and that of interdentals vs. alveolars were quite poor. The same holds for ED bilinguals. This seems to suggest that Returnees and ED bilinguals show similar learning patterns concerning English /f/ and /ɵ/. Namely, they seem to have established separate but not perfect L2 categories for the English /f/ and /ɵ/ sounds, as they perceived identical labiodentals and interdentals quite well, but as they exhibited great difficulty in discriminating between non-identical labiodentals and interdentals. Further, according to the PAM/PAM-L2, when two nonnative sounds are classified as a single L1 sound, discrimination for the nonnative sounds is poor. Consequently, the two English sounds might be partially overlapped in categorization in English because of phonetic/acoustic and articulatory similarities between the two sounds, and this might result in poor discrimination.

Further, Returnees and ED bilinguals showed difficulty in the distinction between the interdentals and alveolars, even though the performance of the former group was poorer than that of the latter one. This may be due to the fact that both /ɵ/ and /s/ are coronal fricatives and thus they share many acoustic and articulatory properties. Accordingly, Returnees and ED bilinguals might have difficulty discriminating between interdentals and alveolars. By contrast, both Returnees and ED bilinguals were quite good at discriminating between /s/ and /š/ and this seems to indicate that the role of palatalization in both L1 and L2 might be rather limited. That is, similar to Korean, the sound /s/ in English can be palatalized as [š] before the high front glide /j/ as in *I miss you*. Yet, /š/ also functions as a phoneme in English unlike in Korean. Thus, it appears that both Returnees and ED bilinguals might have formed a separate category for the sound /š/ (and also for /s/), as opposed to EFL students.

In sum, the results above seem to suggest that L2 experience, along with L1 interference, figures in the perception of nonnative sounds. In fact, the results showed that L1 interference was noticeable even for learners with much L2 experience. This seems to support Strange and Shafer's (2008) claim that L1 interference may be evident even for learners with many years of L2 experience. Additionally, the results seem to support Flege's (1995) claim that both L1 and L2 phonological systems may coexist within a single phonological space. In fact, MacKay et al. (2001) claimed that a merged category which contains an L2 and its closest L1 sounds would develop over time. Further, they noted that bilinguals' perception of the L2 sounds would be different from L2 monolinguals' perception since bilinguals' perception would partially reflect typical patterns for the matching L1 sounds. Importantly, however, it was shown that participants' discrimination of the target sounds could also be affected by phonetic/acoustic and articulatory properties of the sounds.

5. REFERENCES

Best, C. 1995. A direct realist view of cross-language speech perception. In: Strange, W. (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research.* Timonium, MD: York Press. 171-204.

Best, T., Tyler, D. 2007. Nonnative and second-language speech perception: Commonnalities and complementarities. In: Bohn, O. M Munro, M. (eds.), *Language Experience in Second Language Speech Learning: In honor of J. Flege*. Amsterdam/Philadelphia: John Benjamins. 13-34.

Flege, J. E. 1995. Second language speech learning: Theory, findings, and problems. In: Strange, W. (ed.). 233-272.

Gay, T. 1970. Effects of filtering and vowel environment on consonant perception. *Journal of Acoustical Society of America* 48(4). 993-998.

MacCay, I. R., Flege, J. E., Piske, T., Schirru, C. 2001. Category restructuring during second-language speech acquisition. *Journal of Acoustical Society of America* 110(1). 516-528.

Strange, W., Shafer, V. 2008. Speech perception in second language learners. In: Edwards, J.H., Zampini, M. (eds.), *Phonology and Second Language Acquisition*. Amsterdam/Philadelphia: John Benjamins. 153-91.

Sundara, M., Polka, L. 2008. Discrimination of coronal stops by bilingual adults: the timing and nature of language interaction. *Cognition* 106(1). 234-258.

The precedence of perception over production in the acquisition of English stress by EFL learners: An OT account

Shinsook Lee and Mi-Hui Cho

Korea University and Kyonggi University
leesseng@korea.ac.kr and mcho@kyonggi.ac.kr

ABSTRACT

This paper examines the relationship between perception and production by conducting an experiment on 20 Japanese learners' acquisition of English word stress. Overall, Japanese learners' perception accuracy (93.98%) was higher than their production accuracy (84.06%), confirming the precedence of perception over production across the factors examined. The precedence relationship as well as different learner variation patterns was accounted for by the perception-production model proposed by Pater (2004) within Optimality Theory.

Keywords: perception and production, English word stress, learner variation, OT

1. INTRODUCTION

Stress placement in English is known to be affected by factors such as syllable weight and lexical category, even though there are many exceptions (Guion et al. 2003). Namely, heavy syllables with a long vowel, a diphthong, or a coda consonant tend to attract stress (e.g. *machine, agenda, divide, correct*). While stress has a tendency to fall on the final syllable in verbs, it tends to fall on the initial syllable in nouns (e.g. *prodúce* vs. *próduce*, *digést* vs. *dígest*). In addition to syllable weight and lexical category, suffix types are also known to affect stress placement of words in English (Chomsky and Halle 1968). Specifically, level 1 suffixes which are of Latin or Greek origins are dubbed as stress-shifting suffixes, as stress falls on the syllable immediately preceding the suffixes when they are attached to the base words (e.g. *transpórt* vs. *transportátion*). By contrast, level 2 suffixes are of Germanic origin and they do not affect stress placement of the base words when they are attached to the base words (e.g. *góvern* vs. *góvernment*). Suffixes from French are known to be stress-bearing, as stress falls on the suffixes themselves (e.g. *bíllion* vs. *billionáire*). As for the relationship between perception and production, it has been claimed that first and second language learners' perception abilities in general precede their production abilities at both segmental and prosodic levels (Pater 2004; Flege 1991).

Recently, Lee (2006) showed that the acquisition of English stress by Korean college students was influenced not only by factors such as word type (i.e., real vs. nonce words) and lexical category (nouns vs. adjectives) but also by suffix type (level 1 vs. level 2 suffixes). She also reported that Korean EFL students' perception of English stress was better than their production of it. Thus, this paper examines the acquisition of English stress by Japanese learners of English with respect to the factors investigated in Lee. This is because only a few studies have examined the role of suffix type in stress placement, even though there are many studies on the impact of syllable weight and lexical category on stress placement in English. Moreover, not many studies have provided an Optimality theoretic analysis of EFL learners' acquisition of English stress in terms of both perception and production. Thus, this paper raises the following questions: 1) Do Japanese learners of English show the precedence of perception over production in the acquisition of English stress, even when considered the factors investigated in Lee (2006)?; 2) Does learner variation exist in terms of perception and production?; 3) If learner variation exists in terms of perception and production, can Pater's (2004) Optimality-theoretic model of first language acquisition account not only for the precedence relationship between perception and production but also for the learner variation in second language acquisition?

In order to answer the questions addressed, we investigated 20 Japanese learners' acquisition of English stress by conducting experiments on both perception and production. Based on the results, we show that second/foreign language learners' perception of English stress is ahead of their production of it regardless of the factors examined above. Interestingly, however, patterns of learner variation different from those in Lee (2006) emerged, since only two types of variation occurred, unlike Korean learners' patterns: Both good perception and production, and good perception but poor production. Moreover, we suggest that learners' English proficiency rather than their native language background may be more closely related to the precedence relationship and learner variability. Further, we propose that Pater's model should slightly be modified so as to account for second/foreign-language learner variability, especially for Korean learners who showed both poor perception and production patterns. Specifically, we show that the Initial Strengthening constraint is responsible for the learners' misperception of English stress, which in turn, may lead to the misproduction of stress. In addition, we show that lexical representations may encode prosodic structures of target words, as suggested by Pater (2004) and Gnanadesikan (2004).

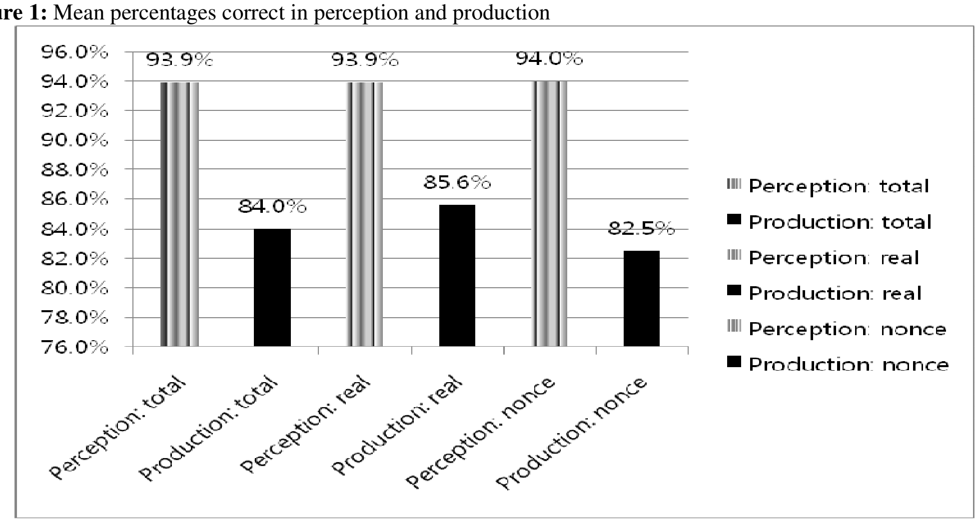
2. EXPERIMENTAL DESIGN

Twenty native Japanese speakers (6 male, 14 female) with a mean age of 21.6 participated in the experiment. The participants were students recruited from Kyushu University, Japan. A total of 64 real and nonce words (32 words for each) consisting of base words with more than 2 syllables and level 1 or level 2 suffixes, which can be attached to given base words, were used to test the participants' stress placement of English words (e.g. level 1 suffixed words: *uniformity, ironic* vs. level 2 suffixed words: *consciousness, effortless*). Specifically, the participants were asked to concatenate aurally presented base words and their matching level 1 or level 2 suffixes into single words with correct main stress possible. For the perception test, the participants were instructed to listen to the test words and to mark a syllable on an answer sheet, on which they perceived primary stress to be.

3. RESULTS

Figure 1 shows that participants' performance in perception was much better than that in production, similar to the result obtained by Lee (2006) with Korean EFL learners.

Figure 1: Mean percentages correct in perception and production



As given in Table 1, participants had great difficulty in producing correct stress patterns for class 1 adjectives regardless of whether the target is a word or a nonce word.

Table 1: Mean percentages correct in perception and production by word type, suffix type, and lexical category

Category	Perception	Production
Real Class 1 Noun	86.2	94.3
Real Class 1 Adjective	95.6	59.3
Real Class 2 Noun	97.5	96.8
Real Class 2 Adjective	96.2	91.8
Nonce Class 1 Noun	90.6	93.7
Nonce Class 1 Adjective	92.5	48.1
Nonce Class 2 Noun	95.6	93.1
Nonce Class 2 Adjective	97.5	95

As shown in Table 2, the effect of suffix type was significant and there was a three-way interaction among the variables in both perception and production.

Table 2: A Repeated Measures ANOVA

		Perception		Production	
Factor/Factor interaction	df	*F*	*p*	*F*	*p*
Word type	1, 19	.021	.88 6	11.176	.003
Suffix type	1, 19	11.217	.003	39.642	.000
Lexical category	1, 19	3.466	.078	170.242	.000
Word type*Suffix type	1, 19	.151	.702	3.763	.067
Word type*Lexical category	1, 19	1.208	.286	.568	.460
Suffix type*Lexical category	1, 19	2.397	.138	82.433	.000
Word type*Suffix type*Lexical category	1, 19	5.033	.037	6.974	.016

The general results were that participants' good perceptual abilities were not necessarily related to their good production skills. Specifically, the following learner-variation patterns emerged, as shown in Tables 3 and 4.

Table 3: Japanese EFL learners' variation patterns with respect to perception and production

<table>
<tr><td rowspan="5">Japanese participants (N=20)</td><td>Perception</td><td>Production</td><td>Number of participants</td></tr>
<tr><td>Good</td><td>Good</td><td>16 (80%)</td></tr>
<tr><td>Good</td><td>Poor</td><td>4 (20%)</td></tr>
<tr><td>Poor</td><td>Good</td><td>0 (0%)</td></tr>
<tr><td>Poor</td><td>Poor</td><td>0 (0%)</td></tr>
</table>

Table 4: Korean EFL learners' variation patterns with respect to perception and production: Lee (2006)

<table>
<tr><td rowspan="5">Korean participants (N=29)</td><td>Perception</td><td>Production</td><td>Number of participants</td></tr>
<tr><td>Good</td><td>Good</td><td>1 (3.4%)</td></tr>
<tr><td>Good</td><td>Poor</td><td>13(44.8%)</td></tr>
<tr><td>Poor</td><td>Good</td><td>0 (0%)</td></tr>
<tr><td>Poor</td><td>Poor</td><td>15 (51.8%)</td></tr>
</table>

Interestingly, the category of poor perception but good production did not emerge either for Japanese EFL learners or for Korean EFL learners, similar to the category patterns found in first language acquisition. Pater (2004) proposed a perception-production model which captures the discrepancies between perception and production in terms of differential rankings of the separate set of faithfulness constraints posited for the perception and production components with regard to the markedness constraints. He also assumed that the lexical form is the output to the perception grammar and the surface form is the output to the production grammar. In what follows, we adopt Pater's model and provide an OT account for the results of the experiment.

4. AN OT ACCOUNT FOR THE DISCREPANCIES BETWEEN PERCEPTION AND PRODUCTION

(1) Rhythm and faithfulness constraints

a. FootBinarity: Feet are binary.

b. Parse-syllable: Syllables are parsed by feet.

c. All-Foot-Left: Every foot stands at the left edge of some ProsodicWord.

d. NonFinality: No prosodic head (syllable, foot) is final in the ProsodicWord.

e. Max(perception): If the input is a perceived form, every segment of the input has a correspondent in the output.

f. Max(production): If the input is a lexical form, every segment of the input has a correspondent in the output.

g. Ident-Stress(perception): If the input is a perceived form, primary stress of the input has a correspondent in the output.

h. Ident-Stress(production): If the input is a lexical form, primary stress of the input has a correspondent in the output.

i. InitialStrengthening: The initial syllable should be most prominent in a word.

(2) Good perception and good production: *consciousness* (Nouns)

a. Perception:

Perceived form: kánʃəsnis	Max (per)	ID-Stress (per)	Max (pro)	ID-Stress (pro)	Ft-Binarity	Parse-σ	Non-Finality	All-Foot-Left
L_1☞ [kánʃəs] nis			NA	NA		*		
L_2 [kánʃəs][nis]			NA	NA	*		*!	**
L_3 [kánʃəs]	*!**		NA	NA				
L_4 kan[ʃə́snis]		*!	NA	NA		*		*

b. Production

Lexical form: [[kánʃəs] nis]	Max (per)	ID-Stress (per)	Max (pro)	ID-Stress (pro)	Ft-Binarity	Parse-σ	Non-Finality	All-Foot-Left
S_1☞ [kánʃəs] nis	NA	NA				*		
S_2 [kánʃəs][nis]	NA	NA			*		*!	**
S_3 [kánʃəs]	NA	NA	*!**					
S_4 kan[ʃə́snis]	NA	NA		*		*		*!

(3) Ranking for good perception and good production: *consciousness* (Noun), *ironic* (Adjectives):

Max(perception), Ident-Stress(perception), Max(production), Ident-Stress(production) >> FootBinarity, Parse-syllable, NonFinality, All-Foot-Left

(4) Good perception but poor production: *ironic* (Adjectives)

a. Perception

Perceived form: airánik	Max(per)	ID-Stress (per)	Ft-Binarity	Parse-σ	Non-Finality	All-Foot-Left	Max(pro)	ID-Stress (pro)
L_1 ai[rán] ik				**		*!	NA	NA
L_2 [rán] ik	*!*			*			NA	NA
L_3 [ai][rán][ik]			*		*	*!, **	NA	NA
L_4☞ ai[ránik]				*		*	NA	NA
L_5 [ái]ra[nik]		*!	*	*	*	**	NA	NA
L_6 [ái]ranik		*!		**			NA	NA

b. Production

Lexical form: [ai[ránik]]	Max (per)	ID-Stress (per)	Ft-Binarity	Parse-σ	Non-finality	All-Foot-Left	Initial Strengthening	Max(pro)	ID-Stress (pro)
S_1 ai[rán] ik	NA	NA		**		*!	*		
S_2 [rán] ik	NA	NA		*				**!	
S_3 [ai][rán][ik]	NA	NA	*		*	*!, **	*		
S_4 ai[ránik]	NA	NA		*		*	*!		
S_5 [ái]ra[nik]	NA	NA	*	*	*!	**			*
S_6 ☞ [ái]ranik	NA	NA		**					*

(5) Poor perception and poor production: *uniformity* (Nouns)

a. Perception

Perceived form: jù:nəfó:rməti	Initial Strengthening	Ft-Binarity	Parse -σ	Non-Finality	All-Foot-Left	Max (per)	ID-Stress (per)	Max (pro)	ID-Stress (pro)
L_1 [jù:nə][fó:rmə]ti	*		*		**!			NA	NA
L_2 ju:nə[fó:rmə]ti	*		***!		**			NA	NA
L_3 [fó:rmə]ti			*			***!*		NA	NA
L_4 ☞ [júnə][fo:rmə]ti			*		**		*	NA	NA

b. Production:

Lexical form: [[júnə][fo:rmə]ti]	Max (per)	ID-Stress (per)	Initial Strengthening	Ft-Binarity	Parse -σ	Non-Finality	All-Foot-Left	Max (pro)	ID-Stress (pro)
S_1 [jù:nə][fó:rmə]ti	NA	NA	*		*		**!		*
S_2 ju:nə[fó:rmə]ti	NA	NA	*		***!		**		*
S_3 [fó:rmə]ti	NA	NA			*			***!*	*
S_4 ☞[júnə][fo:rmə]ti	NA	NA			*		**		

5. DISCUSSION AND IMPLICATIONS

From the OT analysis provided in the previous section, the following ranking schema emerged:

(6) Ranking schema

a. Good perception, good production: Faith(perception), Faith(production)>>Markedness

b. Good perception, poor production: Faith(perception)>>Markedness>>Faith(production)

c. Poor perception, poor production:

(i) Markedness, Faith:Max(perception)>>Faith:Stress(perception)>> Faith(production-NA)

(ii) Faith(perception-NA)>>Markedness, Faith:Max(production)>>Faith:Stress(production)

First, in the category of good perception and good production in (6a), the constraint ranking shows that both faithfulness constraints of perception and production outrank markedness constraints, which reflects the ranking schema of adult native speakers of English. The category of good perception but poor production in (6b) indicates that markedness constraints are couched between Faith(perception) and Faith(production). This ranking is similar to that of children learning English as their mother tongue. However, in the category of poor perception and poor production in (6c), we have a rather different picture from Pater's (2004) proposed ranking schema for first language acquisition. This is because not all the faithfulness constraints relevant to perception and production components are uniformly dominated by markendess constraints. In other words, EFL learners may perceive the contents of segments intact, while they may misperceive the prominence of stress. Consequently, only stress related faithfulness constraints are low-ranked but other segment relevant faithfulness constraints are unranked with respect to markedness constraints. The same holds for the production grammar, thus only stress related production constraints are dominated by markedness constraints as opposed to segment related production constraints.

Also, the ranking schema presented in (6) shows that the same constraint rankings account for both nouns and adjectives. Importantly, it is proposed that the constraint of InitialStrengthening plays a crucial role in determining the optimal output, as most Japanese and Korean EFL learners tended to perceive and produce the initial syllable as most prominent. This may be derived from the fact that phrase-initial or prosodic-initial position is most salient in pitch-accented languages such as Korean and Japanese. And this can attract learners' attention to the first syllable of the target words, which reflects interference from the native language.

Further, the paper assumes that prosodic structure of a given word can be encoded in the lexical representation of the word, as proposed in Pater (2004), although stress is predictable in some languages with fixed stress like French where stress always falls on the last syllable in a word. Concerning the encoding of prosodic structure in the lexical representation, Gnanadesikan (2004) also maintained that children's inputs are adult-like, and thus the inputs should include prosodic structures, even though children's outputs are subject to constraints, resulting in different surface representations from adult forms (e.g. *mosquito* [fi-giɾo] *spaghetti* [fi-gɛɾi]).

Finally, Japanese EFL learners showed only two patterns of learner variation: Good perception and good production, good perception but poor production. By contrast, Korean EFL learners showed another pattern, that is, both poor perception and production, in addition to the two patterns shown by Japanese EFL learners. This might have resulted from learners' different English proficiency rather than their native language, as most Korean participants' English proficiency was low relative to Japanese participants' English proficiency, as Japanese participants were recruited from a high scholastic-level university. Accordingly, this might have contributed to the different patterns of learner variation between Japanese and Korean participants.

6. REFERENCES

Chomsky, N., Halle, M. 1968. *The Sound Pattern of English*. New York: Harper and Row.

Flege, J.E. 1991. Perception and production: the relevance of phonetic input to L2 phonological learning. In: Hueber, T, Ferguson, C. (eds.), *Crosscurrents in Second Language Acquisition and Linguistic Theories*. Amsterdam: John Benjamins. 249-289.

Gnanadesikan, A. 2004. Markedness and faithfulness constraints in child phonology. In: Kager, R., Pater, J., Zonneveld , W. (eds.), *Constraints in Phonological Acquisition*. Cambridge University Press. 73-108.

Guion, S., Clark, J., Harada, T., Wayland, R. 2003. Factors affecting stress placement for English non-words include syllable structure, lexical class, and stress patterns of phonologically similar words. *Language and Speech* 46. 403-427.

Lee, S. 2006. The relationship between perception and production in L2 English stress acquisition. *Studies in Phonetics, Phonology and Morphology* 12(3). 661-685.

Pater, J. 2004. Bridging the gap between receptive and productive development with minimally violable constraints. In: Kager, R., Pater, J., Zonneveld , W. (eds.), *Constraints in Phonological Acquisition.* Cambridge University Press. 219-244.

[pɪlɪpɪno], I know, but [fɪlɪpɪno] is what I say

Alex Ho-Cheong LEUNG

Newcastle University, School of English Literature, Language and Linguistics
h.c.leung@ncl.ac.uk/ alexlhc@gmail.com

ABSTRACT

As a response to a preliminary study (Leung to appear) of five HK learners of English which found that those who had grown up hearing Filipino-accented English showed no trace of this accent in their production, this study probes further to look for more subtle signs of exposure to Filipino English. Data were collected from 10 speakers aged 2 ½ to 25 who were divided into three groups. Both Groups A and B were initially exposed to Filipino-accented English input at home, and Group A continued to receive such input. Group C had not received any Filipino-accented English input at home. Findings from two perception tasks targeting English words with /p/, /t/, /k/, /f/, and /v/ onsets spoken in a Filipino accent showed that speakers with exposure to Filipino-accented English could better perceive these words than those who had none. A decline from Group A to C was found in their ability to recognise target phonemes, indicating that quantity and/or recency of input plays a role. These results raise the issue of incipient/ passive-bilingualism (Diebold 1964; Romaine 1995) and call for more detailed study of attitude, accommodation and identity with respect to the acquisition of a given second language variety.

Keywords: Input, perception, Filipino English, variety/varieties, Hong Kong Chinese

1. INTRODUCTION

How much do we really know about input in the second language acquisition of phonology? Given the possibility of not only internal but external factors impacting on the acquisition outcome, one can expect the end state be highly variable (Piske and Young-Scholten 2009). The literature tells us that native-like attainment is far from a guarantee; learners with near native competence coexist with speakers whose accents are markedly foreign even after prolonged residence in the target language country. While researchers have discussed the effect of foreign-accented input on second language phonology (e.g. Young-Scholten 1994; 1995) and other language contact phenomenon such as koineization and dialect levelling (e.g. Kerswill and Williams 2005), little research has been carried out on learners' choice when several varieties of the same language exist (see Rys 2007).

In Hong Kong, children are exposed to both Cantonese and English, but studies of bilingual/young second language learners (e.g. Matthews and Yip 2009) focus on the mental representation of two languages rather than on input factors. With respect to English, input may be an important factor where children's first exposure is from Filipino housekeepers. We therefore expect children's acquisition of features characteristic of Filipino English: unaspirated plosives and [p], and [b] for /f/, /v/ (Bautista 2000; Tayao 2008). Under such circumstances, many children appear not to have not grown up 'acquiring' a Filipino accent.

In spite of the intriguing nature of this phenomenon, such seeming dissociation between input and acquisition outcome has not been studied thoroughly and systematically in L2 phonology (Leung 2009a, b). This paper is an effort to address this phenomenon from a speech perception perspective. It will first summarise a study that looks at this apparent dissociation from a production point of view (Leung to appear). Building on that, the paper will proceed to report on the actual study that pertains to subjects' ability to perceive English words with /p/, /t/, /k/, /f/, /v/ onsets pronounced in a Filipino accent. This will then be followed by a section of discussion. Lastly, some relevant issues will be pointed out in the concluding section.

2. EXPLORATORY STUDY

In his exploratory study that taps into the purported dissociation between input and linguistic outcome in the context of Hong Kong, Leung (to appear) finds that speakers who grew up receiving Filipino-accented

English input do not produce English speech with such an accent. The study focuses on Hong Kong-Chinese English speakers' pronunciation of labio-dental fricatives /f/ and /v/, which are realised as [p] and [b] in Filipino English (Bautista 2000; Tayao 2008). Furthermore, the study also examines the production of /p/, /t/, /k/ which are often not aspirated in the English of Filipino speakers (ibid.).

Through two separate tasks, it is revealed that the pronunciation of five subjects (three with Filipino-accented English input, two Hong Kong controls without such input), aged from 12 to 23 at the time of data collection, differ from the Filipinos. In a paragraph reading task, marked differences were observed with regard to both the aspiration of /p/, /t/, /k/ (aspiration for subjects and HK controls: 100% vs Filipino Group: 17%) and the rendering of /f/, /v/ as [f] and [v] (subjects and HK controls: 100% vs Filipino Group: 50%). These findings are reinforced by data from a second task, which was a spontaneous speech production task set up in the format of a semi-structured interview. The relevant sounds analysed from the recordings of the subjects displayed no Filipino trace/ influence. These results accord with the anecdotal observation of the dissociation conundrum between input and acquisition in Hong Kong.

3. THE PRESENT STUDY

3.1 Speech perception

The study of speech production by Leung (to appear) seems to indicate that learners are not influenced by the Filipino-accented input which they are constantly exposed to from an early age. This seems to be at odds with a range of L2 phonology findings (Moyer 2009; Young-Scholten 1994). One might hence be led to thinking that learners do not acquire this variety, therefore, challenging the role that input plays in (second) language acquisition (cf. Piske and Young-Scholten 2009; Young-Scholten 1994; 1995). However, for one to claim that these learners do not acquire this particular variety of English, one has to show that they shun this accent or to demonstrate their insensitivity towards it. This is because learners could be bi-dialectal and possess passive knowledge of this variety even though the spoken form is not adopted. That is to say, these speakers could have built up implicit knowledge (perception) of this type of English through ongoing exposure (Wode 1997), but nonetheless fail to display it in production. In the light of that, this follow-up study with the focus of speech perception was conducted.

Against the backdrop of Leung (to appear), the study intended to find out whether subjects who grew up with Filipino-accented English input could perceive this variety, despite not actively producing it.

The study involved two tasks so that cross-validation of results is made possible. In the first task, ten subjects were asked to listen to stimuli of English words with /p/, /t/, /k/, /f/, and /v/ onsets pronounced with a Filipino accent and write them down. In the second task, participants were asked to pick the given pictures of the words that they heard.

It was found that speakers who were exposed to Filipino-accented English performed differently from the subjects who had not received such input. This leads us to question the tentative non-acquisition claim made by Leung (to appear) on the basis of production data.

3.2 Subjects

The subjects of this study ranged from 2 ½ to 25 years of age. Five of the ten subjects had also participated in Leung's (to appear) speech production study. Their data are crucial to analysing the possibility of the learner having acquired the Filipino-accented variety implicitly without conscious knowledge, as it was already shown by Leung that these people do not show traces of Filipino-accent in their production of English (cf. section 2).

Informants were divided into three groups. Group A consisted of speakers who had been and still were being exposed to Filipino-accented English input, while group B were speakers who had been exposed to this variety previously but no longer do. Finally, group C were speakers who had not received any Filipino accented English input. The speakers' language profiles are very diverse; it covers a wide-spectrum of variations in age, level of education and years of exposure to English (cf. Table 1).

Table 1: Subjects' profile.

Groups	Subjects	Age	Length of English instruction	Level of education	Length of exposure to Filipino-accented English
A	AL	23	12 yrs	Tertiary (undergrad)	23 yrs (from birth)
	CH	18	12 yrs	Tertiary (hi-diploma)	18 yrs (from birth)
	NH	12	6 yrs	Secondary (1st-yr)	12 yrs (from birth)
	JA	4 yrs and 5 months	1 yr	Kindergarten (2nd yr)	4 yrs and 5 months (from birth)
	KY	3 yrs and 2 months	2 months	Kindergarten (1st yr)	3 yrs and 2 months (from birth)
	KL	2 yrs and 7 months	0 yr	Pre-school	2 yrs and 7 months (from birth)
B	AW	24	12 yr	Tertiary (undergrad)	10 yrs (from birth – 10)
	TC	16	9 yr	Secondary (4th yr)	3 yrs (from age 7 -10)
C	KK	25	12 yr	Tertiary (postgrad)	N/A
	BN	24	12 yr	Tertiary (postgrad)	N/A

3.3 The tasks

All the recordings for the tasks in this study were recorded by a female Filipino English speaker whose production resembled the prototypical Filipino accent attested in the literature (Bautista 2000; Tayao 2008). Painstaking effort was put into the preparation of the recordings. The reader was, at the beginning, very conscious about the whole process even though she knew she was not the target of this study. She admitted paying constant attention to the words that begin with letters "f" and "v" since she was aware of the confounding nature of these sounds with [p], and [b] in Filipino English. She also thought seeing the spelling of the words affected her pronunciation to a certain degree (for the influence of orthography in pronunciation and acquisition, please refer to Bassetti 2009). However, it is believed that the final version chosen for the study adequately represents Filipino-accented English.

3.3.1 Task 1 (word spelling task) & Task 2 (picture choosing task)

Both tasks involved a word-listening procedure where informants identified the 15 words played to them. Subjects' ability to identify the words played is considered to be an indication of his/her knowledge of such word/ sound and vice versa. The first task required subjects to take dictation on a sheet of paper of the word they heard. In the second task, subjects were asked to select the picture representing the 15 words they heard from the images given. The option of "don't know" was available when they could not identify the word. Similarly, they could provide a word in the second task when they thought the word they heard was not among the pictures provided, even though in actuality the pictures given encompassed every word played.

Words chosen were mainly vocabularies that are related to the daily life of children such as, "foot", "van", "pen", "ten" and "king". Also, the words represented a number of different phonological environments. For instance, vowels of different heights were included following the onset /f/. Similarly both front and back vowels were included. Examples include, "fish" where /ɪ/ is a [+ high], [+ front]; "foot" where /u:/ is a [+ high], [-front]. This was to ensure the perception results obtained were not affected by the quality of the following vowel. In addition, subjects were tested for their knowledge of the words in the given pictures before the task begins. Words that the subject did not know were excluded lest the analysis was obscured. A training phase was also included; subjects were asked to listen to a few words other than the targets and practise writing them down and choosing them from the pictures. The actual experiment did not start until the informants were familiar with the procedures.

3.4 Results

3.4.1 Task1 word spelling task

The youngest group of subjects encountered difficulties in spelling the words since they were still in the process of acquiring literacy in English, therefore, their results are not reported here. In total 15 words with the various onsets /f/, /v/, /p/, /t/, /k/ were played to the subjects. The words were chosen on a random basis from the pre-selected list of words which mostly meet the criteria stated in section 3.3.1.

The results show a cline of varying capabilities in recognising the sounds. The subjects in group A have the highest sensitivity towards the different sounds, with group B being slightly less competent in identifying the words, and group C being the worst among the three groups. Group A is able to identify the words beginning with /p/, /t/, /k/ 85 % of the time, /f/, /v/ 94 % of the time. Group B is able to recognise words with /p/, /t/, /k/ onsets correctly 75 % of the time, and 44 % of the /f/, /v/. Group C identifies 21 % of the words starting with /p/, /t/, /k/ in the passage, and 27 % of the words with /f/, /v/ are correctly spotted. Table 2 gives a detailed account for the individual performance of each subject in the group.

Table 2: Subjects' performance in task 1.

Groups	Subjects	Plosive onsets (response/ no. of tokens)				Labio-dental fricative onsets		
		p [p^h/ p^h]	t [t^h/ t^h]	k [k^h/ k^h]	total	f as [f]	v as [v]	total
A	AL	3/3	4/4	3/3	10/10	3/3	2/2	5/5
	CH	3/3	1/3	3/3	7/9	3/3	3/3	6/6
	NH	3/3	1/3	3/3	7/9	2/3	3/3	5/6
B	AW	1/3	2/2	2/2	5/7	2/6	1/2	3/8
	TC	3/3	2/3	2/3	7/9	1/3	2/3	3/6
C	KK	1/4	0/2	0/2	1/8	1/5	1/2	2/7
	BN	0/2	0/3	2/3	2/7	1/6	1/2	2/8

3.4.2 Task 2, picture choosing task

Words containing /p/, /t/, /k/, /f/, /v/ as onsets were played. In line with the previous task, a general decline in sensitivity towards the target sounds is observed moving from the results of group A to group C. Group A has chosen the pictures accurately for words with /p/, /t/, /k/ onsets 89% of the time, while /f/, /v/ are correct 61% of the time. It has to be pointed out that the data obtained from subject KY were removed from the analysis since he was found to have certain learning disabilities. Group B has selected the correct pictures that represent the words with /p/, /t/, /k/ onsets approximately 40% of the time, while the accuracy rate for /f/ and /v/ is 50%. Lastly, group C has picked 24% of the correct images corresponding to words with /p/, /t/, /k/ onsets, and 46% for /f/, /v/ onsets. Table 3 lists the details of the subjects' responses.

Table 3: Subjects' performance in task 2.

Groups	Subjects	Plosive onsets (response/ no. of tokens)				Labio-dental fricative onsets		
		p [p^h/ p^h]	t [t^h/ t^h]	k [k^h/ k^h]	total	f as [f]	v as [v]	total
A	AL	3/3	3/3	3/3	9/9	2/3	3/3	5/6
	CH	2/3	3/3	2/2	7/8	2/4	0/1	2/5
	NH	3/3	3/3	2/2	8/8	1/4	0/1	1/5
	JA	2/4	3/3	2/2	7/9	4/5	1/1	5/6
	KY	1/5	1/4	1/2	3/11	1/3	1/1	2/4
	KL	3/4	4/4	1/2	8/10	3/4	1/1	4/5
B	AW	2/3	2/4	1/2	5/9	4/5	1/1	5/6
	TC	0/2	1/1	1/3	2/6	0/4	1/2	1/6
C	KK	1/4	0/3	1/1	2/8	1/5	2/2	3/7

	BN	0/3	0/4	2/2	2/9	1/4	2/2	3/6

4. DISCUSSION

A cline of decreasing ability to pick the appropriate words can be observed in both tasks (table 4 and 5).

Table 4: Correction identification of words with the following sounds in task 1.

Groups	/p/, /t/, /k/	/f/, /v/
A	85%	94%
B	75%	44%
C	21%	27%

Table 5: Correction identification of words with the following sounds in task 2.

Groups	/p/, /t/, /k/	/f/, /v/
A	89%	61%
B	40%	50%
C	24%	46%

These results can be taken as an indication of the decline in awareness towards the sounds that represent /p/, /t/, /k/, /f/, /v/ onsets among groups. The remarkable ability to identify Filipino /p/, /t/, /k/ is actually a very good indicator of knowledge of the Filipino variety, since non-aspiration is found to be one of the major barriers for intelligibility (Jenkins 2000). Therefore, the competence in these sounds (the non-aspirated p, t, k) is a reliable pointer telling us that subjects have in fact established these sounds in their grammar. However, one might wonder why subjects in Group C did not score zero percent despite not having received any Filipino-accented English input. Given that Filipino foreign-domestic helpers constitute approximately 1.5 % of Hong Kong's population (Visa and Policies 2007), group C's performance could possibly be due to subjects' occasional exposure to this variety of English from the ambient environment.

It is interesting to note that the ability of some subjects in choosing the correct words dropped from task 1 to task 2. This is perhaps a manifestation of task effects; subjects might have relied on the options they had from the images presented to them in the second task. Conversely, in task 1 they had to count on their own competence of the sounds and write down the words presented to them, this perhaps prompted them to write down what they actually heard without being affected by any external stimuli.

In all, results obtained from this study indicate that exposure to Filipino accented English at some point in the subjects' life is essential to their ability in recognising such sounds. In other words, this linguistic system is established in their grammar even though this variety is not adopted in production.

5. CONCLUSION

On the face of it, Leung's study (to appear) seems to be going against the traditional belief of language acquisition, since one would logically acquire the variety s/he is exposed to. The non-production of Filipino accent by the informants is apparently saying that they had somehow avoided acquiring the Filipino variety regardless of the continuous input they received, in some cases prior to starting school and being exposed to another variety of English. However, such a conclusion drawn without considering speakers' ability to perceive the variety is dubious or haphazard at best. It is possible that the subjects have acquired implicit knowledge of the variety, even though they did not use it in their production. This, in fact, is exactly what is found in the perception study reported in this paper. Subjects showed a decline in sensitivity towards Filipino accented English speech sounds according to their exposure profile to this variety. Ongoing exposure to such input leads to better ability in perceiving the sounds. Synthesising the results of Leung's study (to appear) as well as the ones reported here reminds us of the importance of including a wide-array of testing methods (in this case by looking at both production and perception) so that conclusions drawn are more empirically sound. This echoes the point made by Tench (1996) who calls for circumspection in the design of methodologies in L2 phonology studies. In fact, a number of papers in this volume (e.g. Oliver and Iverson;

Lee and Cho) have suggested the need to investigate both production and perception data in order to reveal the full profile of one's phonological competence.

Although the results of the current study should be viewed with caution due to the small sample size included, they do in turn opened up a new research question, namely: "What has impeded subjects' production of Filipino accent in their English speech?" Drawing on insights from related findings of acquisition studies in migration settings (e.g. the so called Ethan experience, where children filtered out their parents' accented input and acquired the community variety (Chambers 2005); and multi-dialectal exposure due to inter-clan marriages (Stanford 2008)), we could infer that sociolinguistic factors may be at work leading to the non-adoption of Filipino accent in speakers' English speech production. Yet, initial findings obtained through a verbal-guised experiment suggested a more complex picture than straightforward accommodation. Results showed that speakers expressed neutral attitudes towards Filipino English but negative ones towards Hong Kong English. This points to the need for further investigation into factors such as attitude and identity with respect to the acquisition of a given second language variety. It will also be interesting to find out when such divergence from Filipino-accented speech occurs if the children orientate to that at all (cf. Kerswill 1996; Kerswill and Williams 2000 for details of shift of linguistic orientation).

6. REFERENCES

Bassetti, B. 2009. Orthographic input and second language phonology. In T. Piske, and M. Young-Scholten (eds.), 191-206.

Bautista, M. L. S. 2000. Studies of Philippine English in the Philippines. *Philippine Journal of Linguistics* 31 (1). 39-65.

Chambers, J. K. 2005. Sociolinguistics and the language faculty. *Canadian Journal of Linguistics* 50. 215-239.

Diebold, A. R. 1964. Incipient bilingualism. In D. Hymes (ed.), *Language in culture and society: a reader in linguistics and anthropology*. New York: Harper & Row.

Jenkins J. 2000. *The phonology of English as an international language: new models, new norms, new goals*. Oxford: Oxford University Press.

Kerswill, P. 1996. Children, adolescents and language change. *Language Variation and Change* 8. 177-202.

Kerswill, P., Williams, A. 2005. New towns and koineisation: linguistic and social correlates. *Linguistics* 43(5): 1023-48.

Kerswill, P., Williams, A. 2000. Creating a new town koine: children and language change in Milton Keynes. *Language in Society* 29. 65-115.

Lee, S, Cho, M.-H. to appear. The precedence of perception over production in the acquisition of English stress by EFL learners: an OT account. *NewSounds 2010 Conference Proceedings*.

Leung, A., H.-C. to appear. An apparent dissociation between input and outcome in child L2 acquisition: L2 English-accent acquisition in Hong Kong. *LangUE 2009 Conference Proceedings*.

Leung, A., H.-C. 2009a. When acquisition defies input: a case of L2 English-accent acquisition in Hong Kong. Paper presented at Student Work in Progress Forum (SWiP). Newcastle University, Newcastle upon Tyne, U.K. 20th May, 09.

Leung, A., H.-C. 2009b. "Filipino" or "Pilipino": A look at L2 English-accent acquisition in the presence of foreign domestic helpers in Hong Kong. Paper presented NewRoutePhD Conference 2009, Brunei Gallery, U.K. 8th June, 09.

Matthews, S. and V. Yip. 2009. Contact-induced grammaticalization: evidence from bilingual acquisition. *Studies in Language* 33 (2). 366-395

Moyer, A. 2009. Input as a critical means to an end: quantity and quality of experience in L2 Phonological attainment. In T. Piske, and M. Young –Scholten (eds.), 159-174.

Oliver, G, Iverson, P. to appear. Individual variation in the production and perception of second-language phonemes: French speakers learning English /i-I/. *NewSounds 2010 Conference Proceedings*.

Piske, T., Young-Scholten, M. 2009. *Input matters in SLA*. Bristol: Multilingual Matters.

Romaine, S. 1995. *Bilingualism (2nd ed.)*. Cambridge: Basil Blackwell.

Rys, K. 2007. Dialect as a second language: linguistic and non-linguistic factors in secondary dialect acquisition by children and adolescents. PhD dissertation, Ghent University.

Stanford, J. 2008. Child dialect acquisiton: new perspectives on parent/ peer influence. *Journal of Sociolinguistics* 12 (5). 567-596.

Tayao, Ma. L. G. 2008. A lectal description of the phonological features of Philippine English. In MA. Lourdes, S. Bautista, and K. Bolton (eds.), *Asian Englishes today: Philippine English – linguistic and literary perspectives*. Hong Kong: Hong Kong University Press. Pp. 157-174.

Tench, P. 1996. Methodology in phonological interlanguage. *IRAL* 34 (4). 241-260.

Visa and Policies - Foreign Domestic Helpers. (2007). Immigration Department Annual Report 2005-2006 <http://www.immd.gov.hk/a_report_05-06/eng/ch1/frameset.htm>. Last accessed 20 January, 2008.

Wode, H. 1997. Perception and production in learning to talk. In S.J. Hannahs, and M. Young-Scholten (eds.). *Focus on phonological acquisition*. Amsterdam: John Benjamins. 17-46

Young-Scholten, M. 1995. The negative effect of 'positive' evidence on L2 phonology. In L. Eubank, L. Selinker, and M. Sharwood Smith. (eds.), *The current state of interlanguage*. Amsterdam: John Benjamins. 107-22.

Young-Scholten, M. 1994. On positive evidence and ultimate attainment in L2 phonology. *Second Language Research* 10 (3). 193-214.

Pronunciation for Communication: an Investigation of the Effects of Explicit and Selected Teaching of Pronunciation in the Brazilian EFL Classroom

Ronaldo Mangueira Lima Júnior

University of Brasilia, Brazil
ronaldo.limajr@gmail.com

ABSTRACT

This study has investigated the effects of explicit pronunciation instruction, and their durability, in the Brazilian English as a Foreign Language (EFL) classroom. An interventionist action research was carried out with two groups of basic level, adolescent learners of EFL. There was intervention of weekly explicit pronunciation lessons for one semester in one of the groups. All participants were recorded once before and twice after the interventions and all their recordings were phonetically transcribed and analyzed. The results indicate, among other conclusions, that there are positive effects of explicit pronunciation teaching and that these effects are durable.

Keywords: phonetics; phonology; second language acquisition; explicit instruction; EFL.

1. INTRODUCTION

Most English textbooks adopted by Brazilian schools and language institutes are produced by international publishers which make them targeting a worldwide audience, having as basis learners from various mother tongue (L1) backgrounds. With pronunciation teaching the situation is far from being different, and to generalize its teaching in a setting where all learners share the same L1, as is the case in Brazil, is not only an unnecessary investment of time but also a way of boring students with tasks that, more often than not, impose no challenge whatsoever to them. A good example is when textbooks present a pronunciation lesson that contrasts the segments [v] and [b] (e.g. *van* and *ban*). The choice to include such a lesson is based on the difficulty that Spanish-speaking learners have in distinguishing such sounds; however, Brazilian learners and teachers are not challenged when practicing this contrast, once the distinction [v] and [b] exists in Portuguese and causes no difficulty at all to its speakers learning English. Another example is a lesson contrasting [r] and [l] (e.g. *rate* and *late*), based on the difficulty that Japanese students have due to the lack of [l] in their L1, which, again, does not imply the slightest challenge to Brazilian learners of English.

As a result, Brazilian EFL teachers and students might end up discrediting the pronunciation tasks presented in their textbooks, and, thus, choosing to simply ignore those activities altogether. Ignoring the pronunciation activities might become a dangerous habit, though. Since the textbook fails in providing meaningful exercises to their students, EFL teachers may start looking at pronunciation as the least important part of the class, or the part that can be skipped if they run out of time, as Kelly (2000) and Yule and MacDonald (1994) point out. The latter even state that, unfortunately, “many in the language-teaching field seem to feel that pronunciation teaching has little observable effect and that classroom time can be more effectively devoted to fostering other aspects of the L2” (Yule and MacDonald 1994: 111).

Therefore, this research defends that Brazilian EFL teachers (and all EFL teachers with their students sharing the same L1) can, and must, benefit from the homogeneity their students have regarding their L1, especially where pronunciation teaching is concerned, selecting specific aspects that cause difficulty to their students due to L1 and L2 phonological differences. After all, “in classrooms [...] where teachers share a first language with their students, teachers need to carefully consider how they can best make use of their students’ first language to further their competency in English” (McKay 2005: 296-297). As a result, instigated on the issues mentioned above, this research had the objective of investigating the effects that the explicit teaching of the segmental features of English that are problematic to Brazilian EFL learners would have on their phonological interlanguage development.

2. METHODOLOGY

An interventionist action research was conducted in order to reach the research objective. The action research model, despite a few variations concerning its definitions, has as its basis the following steps (as in Andre 1995, Carr and Kemmins 1988, Chizzotti 2006, MacIntyre 2002, Moita Lopes 1996, Nunan 1992, to mention a few). In parentheses is how each step is represented in this research:

1. Identify a problem, a situation that requires change (pronunciation teaching in Brazil);
2. Collect data for analyses (recordings);
3. Come up with a hypothesis to plan the intervention (explicit and selected pronunciation teaching might be advantageous to students);
4. Implement the intervention (explicit pronunciation teaching);
5. Evaluate the effects of the intervention (data analyses);
6. Disseminate results (this paper).

In this research, after indentifying the problems in pronunciation teaching in Brazil already mentioned, two classes of Brazilian EFL students were selected to participate in the research. All students were between 11 and 13 years of age, in the basic level (third semester studying English) and both classes would have the same non-native teacher in the semester with the intervention (2/2007). The same-age, same-level and same-teacher criteria were adopted to avoid variants that could invalidate the results. Firstly, students from both classes were recorded reading a diagnostic test, which consisted of words and sentences that tested only the sounds of English considered difficult to Brazilian Portuguese speakers according to Avery and Ehrlich (1992), Collins and Mees (2008), Godoy et al. (2006), Kelly (2000) and Yavaş (2006), as follows:

Table 1: Difficulties Brazilians have with English sounds based on Avery and Ehrlich (1992), Collins and Mees (2008), Godoy et al. (2006), Kelly (2000) and Yavaş (2006).

DIFFICULTIES WITH CONSONANTS		
English sounds	**Possible mistake by Brazilians**	**Examples**
[p^h] [t^h] [k^h] in the beginning of stressed syllables	[p] [t] [k], sounding as [b] [d] [g] to native speakers;	*pin* [pɪn], sounding as [bɪn]
[t] and [d]	[tʃ] and [dʒ]	*teacher* [ˈtʃiːtʃər]
[tʃ] and [dʒ]	[ʃ] and [ʒ]	*catch* [kæʃ] *major* [ˈmejʒər]
[θ] and [ð]	[t], [s] or [f] and [d], [z] or [v], respectively	*think* [fɪnk] *this* [zɪs]
dark, post-vocalic [ɫ]	[w]	*will* [wɪw]
Initial consonant clusters	Vocalic prosthesis in the beginning	*special* [ɪspɛʃəl] *strong* [ɪstrɔŋ]
Final [m], [n] and [ŋ]	Replaced by the nasal vowel	*man* [mæ̃]
Final [ŋ], as in sing	[ŋg]	*sing* [sɪŋg]
Initial [h]	Not pronounced or added to words without it because of spelling	*house* [aws] *honest* [hɔnɪst]
[z] and [s]	[s] and [z]	*eyes* [ajs] *basic* [ˈbejzɪk]
[t], [d] and [ɪd] in final *-ed* of regular verbs	[ed]	*stopped* [stɑːped] *played* [plejed]
Initial [j] and [w]	[ɪ] and [ʊ]	*year* [ɪər] *want* [ʊʌnt]
Final [p], [t] and [k]	Vocalic paragoge [pi], [ti] and [ki]	*took* [tʊki] *and* [ændi]
[g]	[ʒ]	*getting* [ʒɛtɪŋ]
DIFFICULTIES WITH VOWELS		
English sounds	**Possible mistake by Brazilians**	**Examples**
[iː] and [ɪ]	[ɪ] and [iː]	*beat* [bɪt] *bit* [biːt]
[uː] and [ʊ]	[ʊ] and [uː]	*fool* [fʊl] *full* [fuːl]
[æ] and [ɛ]	[ɛ] and [æ]	*man* [mɛn] *men* [mæn]
[æ]	[ɑː]	*bad* [bɑːd]
[ə]	[eɪ], [jon] or [oʊ]	*chocolate* [tʃɔːkəlejt] *action* [ækʃjon] *dangerous* [dejndʒəroʊs]
[ʌ]	[ʊ]	*luck* [lʊk]
[ʌ]	[æ]	*fun* [fæn]

After the recordings, transcriptions and analyses of the diagnostic test, the interventions began. In the research group (17 students), besides their regular classes, learners also had weekly 15-minute pronunciation lessons for one semester, whereas in the control group (11 students), learners only had their regular classes. The pronunciation lessons were delivered by the researcher and students in both groups were recorded again reading the diagnostic test right after the intervention was over (immediate post-test) and once again 11 months later (delayed post-test). These recordings were also phonetically transcribed and analyzed.

3. DATA ANALYSES

The data discussed here is arranged according to the recordings: first, the diagnostic test recording will be analyzed, followed by the analyses of the immediate post-test and finally the delayed post-test.

3.1. The Diagnostic Test

This first analysis was necessary because, even though there is plenty information in the current literature (Table 1) regarding the difficulties Brazilians have with English sounds, it was important to see which ones caused the greatest difficulties to those specific students taking part in the research. By doing so, it was possible to determine which sounds would be approached in the intervention. Therefore, after the transcription of all diagnostic tests, the errors were counted and the sounds that had the greatest occurrence of errors among all research participants were considered the most difficult ones:

Table 2: Most difficult sounds for research participants according to the diagnostic test.

Most Difficult Sounds	Error Occurrence Rate
[æ]; [z] and [s]; [ŋ]	100%
[ɫ]; [uː]	95%
[θ] and [ð]	93%
[t], [d] and [ɪd] in final *–ed*	90%
[iː]	83%
Initial [p^h], [t^h] and [k^h]	81%
[ə]	70%

Having this information, I began the intervention classes, which were limited concerning time, for out of the 58 hours of regular instruction students had in that semester, only 4 were exclusively dedicated to pronunciation, that is, less than 7% of the total. This was done intentionally because this research did not have as a goal to investigate the effects of pronunciation teaching as it would be done in a pronunciation/speech course, but to investigate the effects of pronunciation teaching when incorporated to the regular foreign language lesson, in a realistic and feasible fashion. I understand that not all language teachers can afford to spend a lot of time with pronunciation instruction, especially if they have to follow a schedule established by the language institute, as is the case with most foreign language teachers in Brazil. So in this sense, this research aimed at investigating the effects of small pronunciation moments, which could be easily incorporated to the teaching agenda of most language instructors. It was for this reason that students in the research group did not have 15 extra minutes of pronunciation lessons, but their regular teacher gave me 15 minutes of their regular class once a week for the pronunciation instruction, in a way that the research and control groups had the same number of instruction hours in the semester.

The pronunciation lessons followed the same pattern: sounds were presented with recognition tasks followed by production ones. Recognition activities are important for the beginning because, while there is no knowledge of the new L2 sound, the words in the L2 with sounds that do not exist in the L1 are heard by learners through the sounds of their L1 (Avery and Ehrlich 1992, Naiman 1992). Burns (1992) states that this is actually a cycle, for the same way production helps comprehension the opposite is also true. Still about the importance of recognition exercises, Flege (2007) says that learners need to establish new phonological categories to the sounds of an L2 and that perceiving the phonetic differences between an L2 sound and the closest L1 sound is key to determine a new category.

After the recognition activities, the intervention moved on to a controlled practice of the new sounds, with repetition of them and of words containing and contrasting them. Finally, more communicative activities were used, which incorporated the sound being practiced in more meaningful and authentic discourse. In sum, the pronunciation classes followed the "communicative framework to teach pronunciation" proposed by Celce-Murcia et al. (1996: 52), which has the following steps: description of how the sound is articulated; practice distinguishing the sound from similar sounds; controlled practice; guided practice; and communicative practice.

3.2. The Immediate Post-Test

Right after the semester with the intervention was over, all students were recorded reading the same test used in the diagnostic phase. The words and sentences were transcribed and the error occurrence was again counted and then compared with the diagnostic test. It is possible to verify the positive effects of the explicit pronunciation instruction by analyzing the decrease in error occurrence. Table 3 shows the seven sounds with the greatest reduction in error occurrence between the diagnostic test and the immediate post-test in the research and control groups, respectively:

Table 3: Sounds with highest reduction of error occurrence between the diagnostic test and the immediate post-test.

RESEARCH GROUP		CONTROL GROUP	
Sounds	**Reduction Rate**	**Sounds**	**Reduction Rate**
[θ] [ð]	98% → 55% (**43 points**)	[t]	45% → 36% (**9 points**)
[æ]	100% → 71% (**29 points**)	[iː]	86% → 77% (**9 points**)
[ɫ]	100% → 73% (**27 points**)	[h]	24% → 15% (**9 points**)
[uː]	94% → 68% (**26 points**)	[ɾ]	39% → 30% (**9 points**)
[t] [d] [ɪd]	90% → 65% (**25 points**)	[ɪ]	18% → 9% (**9 points**)
[iː]	81% → 63% (**18 points**)	[æ]	100% → 95% (**5 points**)
[h]	18% → 6% (**12 points**)	[ɔ]	68% → 63% (**5 points**)

Although there was also some decrease in error occurrence in the control group – after all, these students were also having their regular English classes –, there was not one sound in the control group whose decline in error occurrence was higher than 10 percent. Likewise, with the research group, there was not one sound that was not taught in the intervention whose error occurrence drop was more than ten percent. These two facts, along with the significant diminution in error occurrence with the sounds taught in the intervention, as shown in table 3, demonstrate that students can benefit immensely from a few moments, even if time-limited, of explicit pronunciation teaching, as long as it is specific to their needs as speakers of a certain L1, in this case, Brazilian Portuguese.

Another way of looking at this data is comparing the two tests from the same participant. With this analysis, it is possible to see that learners with the best error reduction rate in the research group had decreases much higher than those with the best rates in the control group. All names used in this paper are codenames to protect participants' right of anonymity.

Table 4: Participants with highest reduction of error occurrence between the diagnostic test and the immediate post-test.

RESEARCH GROUP		CONTROL GROUP	
Participant	**Reduction Rate**	**Participant**	**Reduction Rate**
Lorena	59% → 35% (**24 points**)	Mario	62% → 54% (**8 points**)
Fabiola	50% → 29% (**21 points**)	Gabriela	65% → 58% (**7 points**)
Patrícia	52% → 32% (**20 points**)	Iolanda	67% → 63% (**4 points**)
Fabia	42% → 25% (**17 points**)	William	64% → 60% (**4 points**)
Thales	50% → 35% (**15 points**)	Yasmin	56% → 53% (**3 points**)
Sabrina	54% → 43% (**11 points**)	Leandro	54% → 52% (**2 points**)
Nivea	53% → 42% (**11 points**)	Alessandra	44% → 42% (**2 points**)
Mauricio	53% → 43% (**10 points**)	Beatriz	58% → 57% (**1 point**)
Mauro	62% → 53% (**9 points**)	Bernardo	49% → 48% (**1 point**)
Geovana	51% → 42% (**9 points**)	Carlos	55% → 55% (**0 points**)

As can be seen, ten participants in the research group had a reduction which was higher than that of the participant in the control group with the best reduction rate.

3.3. The Delayed Post-Test

In order to add a longitudinal aspect to the research, all students were recorded a third time 11 months after the second recording. A comparison in error occurrence between the immediate and delayed post-tests reveals the following significant reductions:

Table 5: Sounds with highest reduction of error occurrence between the immediate and delayed post-tests.

RESEARCH GROUP		CONTROL GROUP	
Sounds	**Reduction Rate**	**Sounds**	**Reduction Rate**
[ʌ]	38% → 12% (**26 points**)	[ɫ]	87% → 78% (**9 points**)
[t]	44% → 21% (**23 points**)	[θ] [ð]	81% → 72% (**9 points**)
[ɫ]	73% → 59% (**14 points**)	*Consonant clusters*	12% → 6% (**6 points**)
[ə]	64% → 54% (**10 points**)	[ɪ]	9% → 5% (**4 points**)

This data are somewhat surprising because in the two semesters between the two post-tests, none of the classes had specific and explicit pronunciation instruction, quite the opposite, they didn't even have the same teachers because several of them had transferred their class days/hours and were scattered in nine different classes in the school at the time of the delayed post-test. As a result, the expected result was that participants from both groups had similar reduction rates. The initial intention of the delayed post-test was only to check the durability of the instruction given, and not to evaluate the development of the participants without pronunciation instruction; nonetheless, the data in tables 5 and 6 show that, even without explicit pronunciation instruction for one year, the participants who had had the intervention classes continued progressing in the phonological area more significantly than the ones in the control group.

The explanation here proposed is that learners in the intervention group, due to the explicit pronunciation lessons, had their awareness raised to the importance of learning the L2 pronunciation. Also, the pronunciation lessons may have fostered, besides the immediate learning documented in the first post-test, a latent knowledge about the sounds that were taught, in such a way that it was only materialized in some fo the learners' production only when their interlanguages were ready for this acquisition. For some learners in the research group, the intervention lessons might not have helped them pronounce the sounds being taught immediately, but the lessons certainly assisted them to perceive these sounds and self-monitor in later stages of phonological acquisition.

Comparing both post-tests for each participant individually, it is also possible to see the higher development within the research group:

Table 6: Participants with highest reduction of error occurrence between the immediate and delayed post-tests.

RESEARCH GROUP		CONTROL GROUP	
Participant	**Reduction Rate**	**Participant**	**Reduction Rate**
Geovana	42% → 26% (**16 points**)	Mário	54% → 47% (**7 points**)
Murilo	62% → 47% (**15 points**)	Yasmin	53% → 46% (**7 points**)
Rita	64% → 53% (**11 points**)	Iolanda	63% → 58% (**5 points**)
Thales	35% → 24% (**11 points**)	Beatriz	57% → 52% (**5 points**)

Another way to analyze this data is looking at the individual lowest error occurrence in each one of the three tests. In the control group, the individual lowest error occurrences were: 44% in the diagnostic test, 42% in the immediate post-test and 41% in the delayed post-test. There was a difference of three points between the first and the last tests. Among the participants in the research group, the individual lowest error occurrences were: 42% in the diagnostic test, 25% in the immediate post-test and 24% in the delayed one, indicating a fall of 18 points.

Lastly, we can also look at the error occurrence average in each one of the three tests for each group. In the control group, the arithmetic average of error occurrence was 57% in the diagnostic test, 54% in the immediate post-test and 51% in the delayed post-test, which shows a decrease of 6 points between the first and last recordings. The research group, however, had an average of error occurrence of 56% in the diagnostic test, 46% in the first post-test and 43% in the last one, with a difference of 13 points between the first and last tests.

The analyses of the delayed post-test, thus, reveal that not only were there retention and durability of the positive effects of the explicit pronunciation lessons, but also the continuous development in the phonological interlanguage of the participants that had the intervention lessons, even without having specific pronunciation lessons for a year.

4. FINAL WORDS

As has already been explicated in the data analyses, and returning to the objectives of the research, it can be stated that there are positive effects in the explicit instruction of pronunciation in EFL classes and that these effects are durable not only in the retention but also in the continuous development in the phonological acquisition of the L2. Besides, the fact that these benefits were reached with a limited time for the intervention is astounding. There were 16 intervention sessions with 15 minutes each, within the 35 classes of 100 minutes each of regular classes, i.e. less than 7% of students' class time was used for explicit pronunciation instruction. This fact demonstrates that, for students to profit from a more intelligible pronunciation, there is no need for extra pronunciation lessons, since, as this research points out, small portions of explicit instruction, as long as they are specific and meaningful for students regarding their L1 background, can bring concrete and durable benefits.

As Collins and Mees (2008:212) argue, "many language teachers feel they do not have sufficient time to give their students prolonged pronunciation training" and, hence, "some, indeed, devote no time to it whatsoever". One of the consequences of this study is, therefore, to raise language teachers' awareness to the fact that pronunciation instruction does not need to be prolonged and does not require extra time in the teaching curriculum/agenda in order to be effective. Students not necessarily need to take pronunciation courses besides their regular language classes because their regular classes, if incorporated with consistent, explicit and specific phonological instruction and practice, even with apparently little time dedicated to it, will bring communicative benefits to the language learners.

5. REFERENCES

Andre, M. E. D. A. 1995. *Etnografia da prática escolar* [Ethnography of the school practice]. Campinas: Papirus Editora.

Avery, P., Ehrlich, S. 1992. *Teaching American English pronunciation*. Oxford: Oxford University Press.

Burns, I. M. Pronunciation-based listening exercises. In: Avery, P., Ehrlich, S. *Teaching American English pronunciation*. Oxford: Oxford University Press. 197-205.

Carr, W., Kemmis, S. 1988. *Teoría crítica de la enseñanza*: la investigación-acción en la formación del profesorado [Critical theory of teaching: action research in teaching development]. Barcelona: Martinez Roca.

Celce-Murcia, M., Brinton, D. M., Goodwin, J. M. 1996. *Teaching pronunciation*: a reference for teachers of English to speakers of other languages. New York: Cambridge University Press.

Chizzotti, A. 2006. *Pesquisa qualitativa em ciências humanas e sociais* [Qualitative research in human and social sciences]. Petrópolis: Vozes.

Collins, B., Mees, I. M. 2008. *Practical phonetics and phonology*: a resource book for students. Oxon: Routledge.

Flege, J. 2007. Language contact in bilingualism: phonetic systems interaction. In: Cole, J., Hualde, J. I. (Eds.) *Laboratory phonology 9*. Berlim: Mouton de Gruyter. 353-382.

Godoy, S. M. B., Gontow, C., Marcelino, M. 2006. *English pronunciation for Brazilians*: the sounds of American English. Barueri: Edtitora DISAL.

Kelly, G. 2000. *How to teach pronunciation*. Harlow: Longman.

Macintyre, C. 2002. *The art of action research in the classroom*. London: David, 2002.

Mckay, S. L. 2005. Sociolinguistics and second language learning. In: Hinkel, E. (Ed.) *Handbook of research in second language teaching and learning*. Mahwah: Lawrence Erlbaum Associates, Inc. 281-299.

Moita Lopes, L. P. 1996. *Oficina de linguística aplicada* [Workshop in Applied Linguistics]. Campinas: Mercado de Letras.

Naiman, N. 1992. A communicative approach to pronunciation teaching. In: Avery, P.; Ehrlich, S. *Teaching American English pronunciation*. Oxford: Oxford University Press. 163-171.

Nunan, D. 1992. *Research methods in language learning*. Cambridge: Cambridge University Press.

Yavaş, M. 2006. *Applied English Phonology*. Oxoford, Massachusetts and Victoria: Blackwell Publishing.

Yule, G., Macdonald, D. 1994. The effects of pronunciation teaching. In: Morley, J. *Pronunciation pedagogy and theory*: new views, new directions. Bloomington: TESOL. 111-118.

The development of early L2 phonology: a dynamic approach

Wander Lowie

University of Groningen
w.m.lowie@rug.nl

ABSTRACT

The central themes in the field of L2 phonology have been researched in detail over the past twenty years, but many questions remain unanswered. Additional insights can be gained from research that focuses on the developmental *process* rather than on the *products* of acquisition. Therefore, the current study has adopted a dynamic systems theory approach to second language development (de Bot, Lowie, & Verspoor, 2007; Van Geert, 2008). In this approach, language is regarded as a self-organizing system that constantly changes as a result of its internal reorganization in interaction with its environment, and develops in a nonlinear way. This paper elaborates on the usability of a dynamic systems approach to L2 phonology and reports on a longitudinal pilot study investigating the early L2 phonological development of a young Dutch learner of English. Using a dynamic variability analysis, a number of phonetic correlates were investigated, like the development of VOT over time. The data show how different components of the developing phonological system interact dynamically.

Keywords: early phonological development; dynamic systems theory; second language phonology; process-oriented research

1. INTRODUCTION

Research in the field of L2 phonology has strongly developed since New Sounds 1990, and the central questions in this field, like the age issue, the role of the mother tongue, and the relation between perception and production have been addressed extensively. Although much progress has been made, many questions remain unanswered and studies have sometimes even reached contradictory conclusions. What almost all studies have in common is that they have tested the products of language proficiency at one or two moments in time and that they implicitly assume a linear pattern of development. However, considering the increasing evidence for the nonlinear and dynamic nature of cognitive development, the assumption that language learning develops linearly may be seriously flawed. The constantly changing interactions between the learner's individual characteristics, and the availability of resources (including factors like attention, the availability of input, the opportunity to use the language, and proficiency in other languages) lead to a developmental pattern that cannot be fully predicted. Due to the constant change of all factors involved, the instable language system is highly individual and may be too complex to investigate in traditional experimental settings. In spite of the valuable knowledge we have gained about the grand sweep of L2 phonological development from a large number of studies, it is not surprising that no definitive conclusions can be reached about the very nature of L2 phonological development.

The current study has therefore adopted a dynamic systems theory approach to L2 phonology, which regards language development as a non-linear process. Dynamic Systems Theory (DST) aims at describing and explaining the ways in which complex systems change over time due to a process of self-organization. Development is conceived as an iterative process, in which each level of development critically depends on the previous levels (van Geert 1994). Due to variability and a constantly changing interaction of the system and its environment, the developmental pattern is nonlinear and long term development cannot be reliably predicted. A DST approach to language development is useful in that it provides tools and constructs that can clarify the interaction over time of various variables and accounts for non-linear development (Thelen & Smith, 1994; van Gelder & Port, 1995). The current study aims at investigating L2 phonological development by setting up a series of longitudinal case studies. Previous research using longitudinal case studies has revealed interesting complex interactions in spoken and written language production that would not have surfaced in a traditional quantitative study. For instance, Bassano and van Geert ((2007) have shown how a threshold level of one-word utterances operates as a precursor to grammatical development in a child's language development between the age of 14 and 36 months old. Verspoor, Lowie & van Dijk (2008) have shown how lexical variation and sentence length show a negative dynamic correlation in a case study of L2 writing. Since the application of DST to cognition does not make principled distinctions between different modules of language processing (Spivey, 2007), the application of a dynamic variability analysis to phonological L2 development is a potentially powerful tool in gaining insight into the process of development.

This paper summarizes the main premises of the DST approach to language development, and argues how this could be applied to L2 phonology. By way of illustration, some data are presented from a longitudinal study of a young Dutch learner of English. The data were recorded in two sets of 14 weekly sessions over a total period of two years. For this study, a number of phonetic correlates were investigated that are mentioned in the literature as potentially problematic for Dutch L2 learners of English, like the /æ/-/ɛ/ vowel contrast (Broersma, 2005), Voice Onset Time (Flege, 1991) and syllable-final voicing (Crowther & Mann, 1992). This contribution will focus on the development of VOT at two different task levels.

2. A DEVELOPMENTAL PERSPECTIVE ON L2 PHONOLOGY

Dynamic Systems Theory is a theory of change. Starting in the 1960s, the theory has extensively been used in a wide variety of disciplines, from physics to biology and from meteorology to demography. More recently, applications of DST have been developed within cognitive science (Port & van Gelder, 1995; Thelen & Smith, 1994) developmental psychology (Van Geert, 1998) and language development (Bassano and Van Geert, 2007). Since the late 1990s, the theory has also been applied to second language acquisition (de Bot et al., 2007; Herdina & Jessner, 2002; Larsen-Freeman, 1997). These and many other authors have argued that language can be seen as a complex dynamic system and that language development is a nonlinear, chaotic, and highly individual process that cannot be adequately described using traditional research methods that use a linear type of logic. The arguments to consider language as a dynamic system centre around three principles that can suitably be applied to language development, including L2 phonology: the occurrence of nonlinear, chaotic patterns of development, the existence of interconnected self-organizing subsystems, and the relevance of variability.

2.1. The dynamics of language development

A crucial characteristic of nonlinear development is that it is non-deterministic. Traditionally, second language acquisition is seen as a linear process, with a clear starting point (no knowledge of the second language) and a clear end-point (native-like proficiency in the second language). Not all learners will achieve the ultimate goal, but are assumed to fossilize somewhere along the line. Traditional research has focused on the individual contribution of each of the factors in achieving, or failing to achieve the ultimate goal. From a DST perspective, a person's language is a fully integrated system. Within this framework, it is simply impossible to identify factors affecting the process of language acquisition, because the system *is* the sum of all these factors. Motivation, aptitude, and all the other "factors" constitute one system that constantly changes as a function of the system's previous point of development within the developmental context. A change in "motivation" at one point in time may trigger a change in any of the other components of the system at the next point in time. If we regard the development of the system as a sequence of subsequent iterations, its long term development cannot be predicted due to the complex interaction of all its components over time. And since the system will continuously change (including both "acquisition" and "attrition"), there is no end state of this development. Consequently, the individual contributions of each of the interactions cannot explain the development. Or as van Geert (2003) puts it "the effect of a dynamic process differs from the sum of its parts" (2003: 657).

A dynamic system is characterised by the existence of interconnected subsystems. When cognition is seen as a system, the language system can be seen as one of its subsystems. Below that level, the syntactic system, the lexicon, and the phonological system can again be seen as subsystems. The organisation of the language system as consisting of embedded subsystems has great advantages over a static modular view of language components, as all levels can be presumed to interact dynamically, which can account for self-organisation (Spivey, 2007). Self-organisation is well attested in any naturally occurring phenomenon, including brain activation (Kelso, 1995) and is fed by the system's limitations and resources. The resources that can be assumed to shape L2 phonology include factors like the amount and the nature of the exposure to the target language, the amount of anxiety and social conformity, the learner's motivation and attitude, and the (perceived) distance to the L1, all of which will dynamically change in interaction. Due to its characteristics, a dynamic system may be drawn to a point or path that has been referred to as an attractor state, which is a (sub-)system's preferred state (Kelso, 1995). Attractors in the L2 phonological system can, for instance, be triggered by the L1 phonological system. Using the terminology of L2 phonology research, perceptual magnets (Iverson et al., 2003) could also be seen as attractors.

The nonlinear, erratic patterns of development are often manifested in a great deal of variation found in learner data. Traditional studies have often regarded variation as "noise" that stands in the way of finding significant differences between conditions. From a DST point of view, variability is seen as an essential characteristic of the dynamic process of self-organisation that may signal changes and transitions in the system. Longitudinal studies into second language development have shown the relevance of variability (Verspoor et al., 2008). The analyses show that increased variability leads to a subsequent increase of development, but not linearly and only as a function of the dynamic interaction of factors over time.

An important implication of the chaotic, nonlinear and highly variable nature of language development is that it never stops. While a native speaker variety may be a learner's ultimate goal and could be seen as one of the attractors of the system, it is not the system's only attractor. Neither is it likely that all of the components of a dynamic system reach an attractor state simultaneously. If this happened, we could speak of true fossilization, but an attractor state is not irreversible. Given the availability of enough energy and optimal circumstances (like attention and motivation), an attractor state is not an end state of development.

2.2. Dynamic theory, dynamic methodology

When we acknowledge that the process of language development is best approached from a dynamic perspective, this implies that we will need to adjust our research methodologies. Methods like analyses of variance and regression analyses may be very useful in determining the relative importance of effects at one or two moments in time, but are not very well suited to explore the dynamics of language development. To investigate dynamic systems, a range of methodologies have been developed. Since change over time is essential to DST approaches, DST studies usually make use of longitudinal data, preferably of a dense nature. To investigate the process at different time frames, data collection varies between yearly, monthly, weekly, daily and hourly studies. Ideally, studies should employ a combination of different time frames. Three main types of analyses have thus far been used in DST approaches to language development and cognition: variability analyses, computer modelling, and time series analyses. In variability analyses, a variety of descriptive techniques have been used to investigate the nature of the change over time of one or more variables. Descriptive techniques, like Moving min-max graphs and moving correlations have been used to make sense of graphically discerned tendencies in the data (Verspoor et al, 2008). Building on variability analyses, observed relationships between variables have been tested using dynamic simulations (Van Geert, 2008). Dynamically changing relationships between variables could further be analysed using non-linear time-series analyses (Hamaker, Dolan, & Molenaar, 2005).

2.3. The dynamics of L2 phonological development

The interacting dynamics of phonological processing has been recognized for some, but not all time frames. It has been argued that speech production and perception at the processing level are dynamic processes (Greenberg, Arai, & Grant, 2006). This view is compatible with task-dynamic models of articulation (Van Lieshout, 2004), in which speech production is not described in terms of movements of the individual articulators, but in terms of the coordinated actions of all articulators involved in making a particular constriction (*tract variables*). The combined use of several tract variables is referred to as a *gesture,* which can be seen as the dynamic equivalent of a phoneme. The dynamic specification of a gesture includes a spatial goal that can be related to an attractor in a dynamic system (Goldstein, Byrd, & Salzman, 2006).

This framework is in line with the well-attested observation that perception plays an important role in the development of L2 phonology within larger time frames (Best, 1994; Flege, 1995; Iverson et al., 2003). The starting point in learning additional languages will always be the learner's mother tongue. In pronunciation, the sound structure of the mother tongue will initially be applied to L2 production. The competing L1 perceptual targets can be seen as attractors in the L2 learner's phonology. This approach to L2 phonology is also in agreement with the observed success of pronunciation teaching strategies that use "holistic" methods, like paying attention to overall articulatory settings rather than focusing on the "correct" pronunciation of individual sounds (Collins & Mees, 1993).

3. A CASE STUDY

To test the development of L2 phonology within the framework of a dynamic subsystem, a research program was set up consisting of a series of multiple longitudinal case studies in which a variety of phonetic correlates was included in a number of different settings and speaking situations, for both perception and production, for both young and older learners of English. The present paper reports on a first pilot investigation of a young Dutch learner of English, whose phonological development was monitored over a period of two years. Due to space limitations, only the VOT data of voiceless plosives will be presented in this paper, concentrating on the dynamic difference between two task levels: spontaneous production and word repetition in the second year of production data. In view of the limitation of resources (like attention), it was hypothesized that the development of VOT in the word repetition condition would precede that of the spontaneous condition and that the repetition condition would therefore serve as a precursor to spontaneous production. This pilot consisted of limited data, and serves as an illustration of a dynamic investigation of L2 phonology rather than a full fledged study.

3.1. Participant, materials and procedures

The participant, Hannah, was seven years old at the onset of the study. Although growing up in the Netherlands implies a large amount of exposure to English, Hannah had had no previous instruction in English and her knowledge of English was rather minimal. Importantly, Hannah had not been exposed to written English. Data were collected during two 14-week periods, March–June 2007: and March–June 2008.

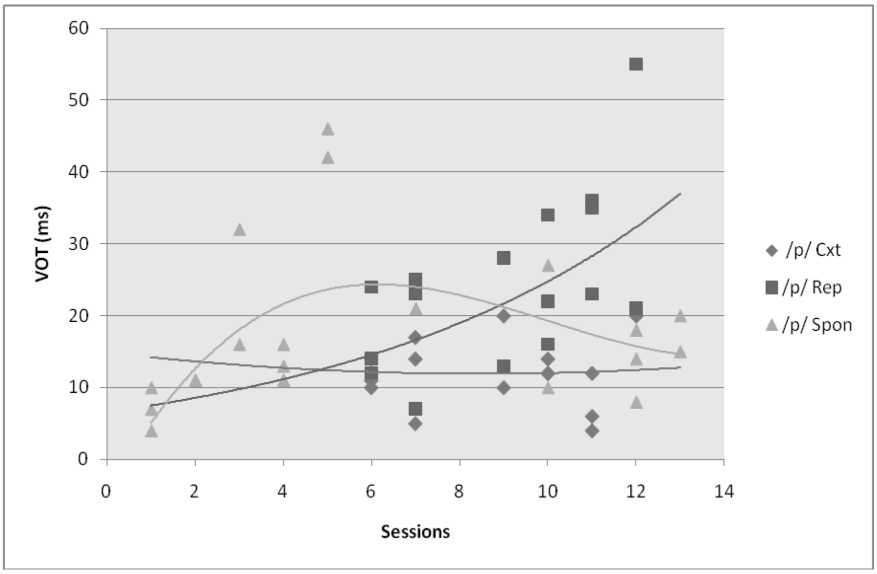
The recording sessions were embedded in English lessons, during which Hannah was exposed to a variety of native speaker data from sound recordings and DVDs. The focus of the lessons was always on fluency and vocabulary and explicit pronunciation instruction was consistently avoided. Data were gathered from three tasks, a picture description task, a sentence repetition task and a word repetition task, which were all presented as vocabulary tests. To ensure consistency in the repetition task, words and sentences were repeated from a recorded female native speaker. The different tasks levels were included to enable the comparison of different degrees of attention to pronunciation. In more cognitively complex tasks, like picture description, fewer resources are available for attention to pronunciation and a lower degree of accuracy may be expected.

To determine the quality of pronunciation, four measures were used: VOT of initial plosives, vowel quality of the /ɛ/ - /æ/ contrast, final voiced and voiceless obstruents and their preceding vowel length, and dental fricatives. In this paper, we will focus on the first measure, VOT. The recordings were all made in the child's own home. After the recordings, analyses were carried out with PRAAT. The VOT was measured for voiceless plosives only from the release of the plosive till the onset of the vowel.

3.2. VOT results

This section concentrates on the second series of recordings. The VOT values for the stop plosives in three places of articulation were analysed separately as well as combined. As can be expected from a beginning learner, the data were highly variable, especially in the spontaneous speaking condition (the picture description task). Figure 1 displays the development of /p/ in the three tasks and includes polynomial smoothers for each of the task levels (trendlines).

Figure 1: VOT values for /p/ (ms) for three task levels in the second data collection period of 13 weeks. The trendlines are polynomial smoothers of the data.



As is clear from this figure, the strongest development is found in the repetition task (Rep), which coincides with a large amount of variability. Towards the end of the data collection period, VOT values for /p/ vary between 20ms and 40ms. Average VOT values for Dutch /p/ are around 10ms, while average values for English native speakers are around 60ms (Cho & Ladefoged, 1999). The least development and the least variability is found in the sentence repetition task (Cxt), which remains around the values reported for native Dutch. Patterns found for other voiceless plosives (/t/ and /k/) were very similar, though the overall VOT values were higher across all conditions, which is in agreement with general observations showing that /k/ has the longest VOT, followed by /t/ and /p/ (Cho & Ladefoged, 1999).

Figure 2: Average VOT values of /ptk/ for all sessions in the second measurement period (2a). The lines show the development as a moving average (window =4). The top right graph (2b) shows the moving correlation between the two task levels. The bottom right graph (2c) shows the result of a growth model testing the precursor interaction between two task levels (300 iterations).

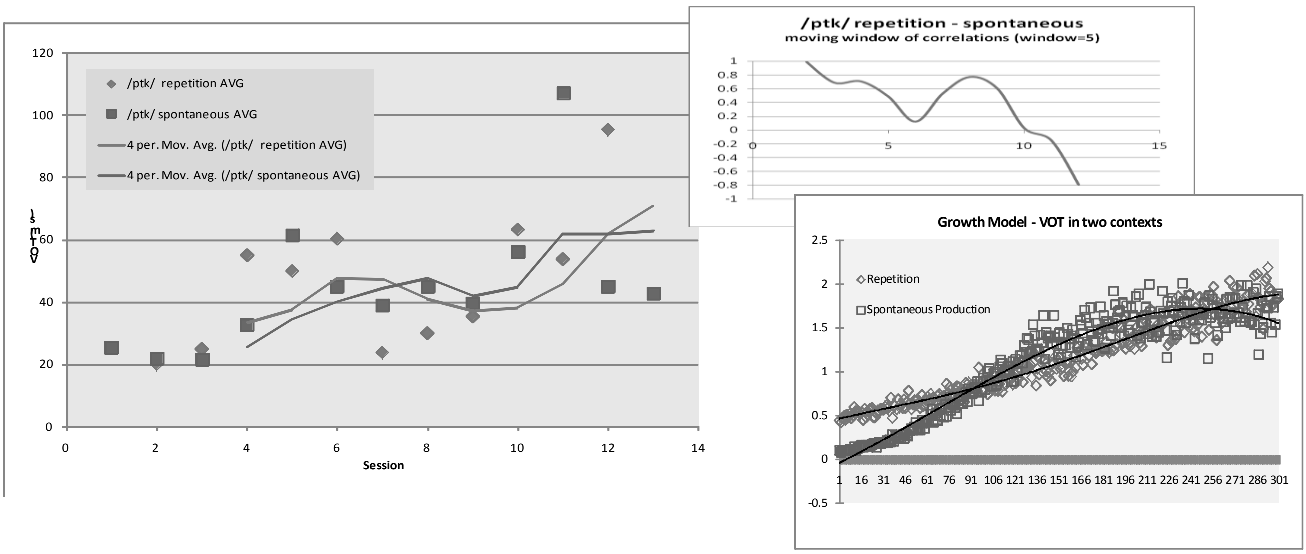


Figure 2a represents the average /ptk/ data for two of the conditions, word repetition and picture naming. In both conditions the VOT values improve significantly over time ($r=.60$, $p<0.05$). The data for both task levels were highly variable and the moving average suggests that the correlation between the two task levels seems to shift over time between positive and negative, which is further illustrated by the moving correlation graph in Figure 2b. The interaction between repetition and picture naming was tested by modelling it as a mutual precursor interaction in 300 dynamic iterations (Figure 2c). Based on the data, the model was used to verify the alternation between support and inhibition of the two tasks. The simulation displays the same pattern of interaction as the one found in the data, which confirms the precursor interaction. The simulation also shows that, due to the “chaotic” development, the amount of variability in the model increases over time and to some extent blurs the difference between the conditions. This is in fact what could be expected to occur in the data, had there been more data points.

4. CONCLUDING REMARKS

In spite of the limited data set, the analysis hopefully serves to illustrate the rich potential of detailed studies of variability in phonological development, especially when this is combined with computer simulations to flesh out the relationships between the developmental variables in more detail. The current illustration shows that, at least in the pilot data, the interaction between VOT production in the speaking situations is not as straightforward as was expected, as no unilateral precursor relation was found. In other aspects of phonological development in the same pilot, the difference between the situations was much more obvious. An example of this is found in the production vowels. In the repetition task, the formant frequencies of the vowel /æ/ move significantly toward native speaker values and away from those of /ɛ/, while in the picture description task a lot of variation is found, but no significant difference could yet be established between /æ/ and /ɛ/.

Studies that have investigated the difference in L2 phonology between young and older learners have attributed this difference to a variety of causes, like cerebral lateralization, cognitive maturation, socio-psychological factors, and the changing influence of the learner’s native language. The difficulty of reaching final conclusions on a topic like this and the seemingly paradoxical observation that some late learners under some circumstances are able to attain native-like phonological control (Bongaerts, 1999), can easily be accounted for if we accept that language is a complex dynamic system. The nonlinear self-organization of the phonological system, under the influence of self-organizing subsystems like the articulatory system, each of which have their own resources, attractors and repellors, and each of which are embedded in a dynamically changing environment, is too complex to allow for simple explanations. Moreover, one of the most universal and best documented properties of L1 child language is its variability. An in-depth examination of the development of a child’s sound system would enable researchers to map the variability in a child’s L2 and thereby reveal transitions in development. Such an approach can reveal the parameters shaping the L2 sound system. Much progress can therefore be expected from detailed longitudinal studies charting a variety of dynamic components to investigate issues in L2 phonology.

5. REFRERENCES

Bassano, D.,Van Geert, P. (2007). Modeling continuity and discontinuity in utterance length: A quantitative approach to changes, transitions and intra-individual variability in early grammatical development. *Developmental Science, 10*(5), 588-612.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In C. Goodman, H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 171-204). Cambridge, MA: MIT Press.

Bongaerts, T. (1999). Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 133-159). Mahwah/London: Erlbaum.

Broersma, M. E. (2005). *Phonetic and lexical processing in a second language*

Cho, T., Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics, 27*(2), 207-229.

Collins, B., Mees, I. M. (1993). Articulatory setting in English, Danish and Dutch and the implications for second language acquisition. *Dutch Working Papers in English Language and Linguistics, 29*, 1-12.

Crowther, C. S., Mann, V. (1992). Native language factors affecting use of vocalic cues to final consonant voicing in English. *Journal of the Acoustical Society of America, 92*(2), 711-722.

De Bot, K., Lowie, W. M., Verspoor, M. H. (2007). A dynamic systems theory approach to second language acquisition. *Bilingualism: Language and Cognition, 10*(1), 7-21.
Flege, J. E. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a 2nd language. *Journal of the Acoustical Society of America, 89*(1), 395-411.
Flege, J. E. (1995). Second language speech learning. theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience* (pp. 233-277). Timonium, MD: York Press.
Goldstein, L., Byrd, D., Salzman, E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. In M. A. Arbib (Ed.), *Action to language via the mirror neuron system*. Cambridge: Cambridge University Press.
Greenberg, S., Arai, T., Grant, K. (2006). The role of temporal dynamics in understanding spoken language. In P. Divenyi, S. Greenberg, G. Meyer (Eds.), *Dynamics of speech production and perception* (pp. 171-190). Amsterdam, etc.: IOS Press.
Hamaker, E. L., Dolan, C. V., Molenaar, P. C. M. (2005). Statistical modeling of the individual: Rationale and application of multivariate stationary time series analysis. *Multivariate Behavioral Research, 40*(2), 207-233.
Herdina, P., Jessner, U. (2002). *A dynamic model of multilingualism. perspective of change in psycholinguistics*. Clevedon: Multilingual Matters.
Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition, 87*(1), B47-B57.
Kelso, J. A. S. (1995). *Dynamic patterns : The self-organization of brain and behavior*. Cambridge, MA etc.: The MIT Press.
Larsen-Freeman, D. (1997). Chaos/complexity science and second language acquisition. *Applied Linguistics, 18*(2), 141-165.
Port, R., van Gelder, T. (1995). *Mind as motion: Exploration in the dynamics of cognition*. Cambridge (Ma): Bradford.
Spivey, M. (2007). *The continuity of mind*. New York: Oxford University Press.
Thelen, E., Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press.
Van Geert, P. (1998). A dynamic systems model of basic developmental mechanisms: Piaget, Vygotsky and beyond. *Psychological Review, 5*, 634-677.
Van Geert, P. (2008, A dynamic systems approach to the study of L1 and L2 acquisition: An introduction. *The Modern Language Journal, 92*, 179-199.
Van Gelder, T. J., Port, R. (1995). It's about time: An overview of the dynamical approach to cognition. In T. J. van Gelder, R. F.Port (Eds.), *Mind as motion: Explorations in the dynamics of cognition*. Boston MA: Bradford books/MIT press.
Van Lieshout, P. (2004). Dynamical systems theory and its application to speech. In B. Maasen, R. Kent, H. F. M. Peters, P. Van Lieshout, W. Hulstijn (Eds.), *Speech motor control in normal and disordered speech* (pp. 51-82). Oxford: OUP.
Verspoor, M. H., Lowie, W. M., Van Dijk, M. (2008), Variability in second language development from a dynamic systems perspective. *The Modern Language Journal, 92*, 214-231.

How native dialect shapes non-native perception: Cuban and Peninsular Spanish learners of English

Irina Marinescu

University of Toronto
irina.marinescu@utoronto.ca

ABSTRACT

The effects that experience and native dialect have on the perception of non-native vowel contrasts are investigated in this study. Two groups of Spanish native speakers from Cuba and Spain were tested in an AX discrimination task on how they perceived the contrasts among the English vowels /æ, ɑ, ʌ/. According to the hierarchy of contrast difficulty predicted in hypothesis 1, all groups of listeners (learners, non-learners of English and controls) show higher error rates for contrasts between back vowels than for contrasts between back and front vowels. Performance improves with experience (hypothesis 2), however the error rates drop less dramatically for the 'difficult' contrast /ɑ-ʌ/ than for the other 'easier' contrasts. Native dialect influences non-native perception (hypothesis 3), as error rates with the /ɑ-ʌ/ contrast indicate that L2 learners with different native dialects map the L2 vowels differently. The differences in performance between the advanced groups of the two dialects investigated are attributed to the perceptual strategy of boundary-shifting, whose extent and direction differs cross-dialectally.

Keywords: non-native perception, back vowels, Spanish, English, dialect, experience

1. INTRODUCTION

This study examines the role of perception in the reportedly different categorization of the non-native English vowels /æ, ʌ, ɑ/ by Spanish native speakers with different dialectal backgrounds. Previous studies with inexperienced Spanish learners of English (Guitart, 1985, 1996) found that each of the /æ, ʌ, ɑ/ vowels may be heard differently and, more importantly, that the learner's native variety influences L2 perception. Specifically, the English /ʌ/ tends to be perceived as /a/ by Peninsular listeners, but as /o/ by Caribbean listeners. Using a perceptual task, the present study aims to answer two questions: (1) In what way does non-native perception differ for listeners with distinct native varieties and what dialect-specific perceptual strategies are responsible for the differences? and (2) What are the role of experience with non-native contrasts and its relationship with the native dialect? Two groups of Spanish native speakers from Cuba and Spain, each including two subgroups of learners and non-learners of English, were tested in an AX discrimination task on the Canadian English contrasts /ɑ-ʌ/, /æ-ʌ/ and /ɑ-æ/. The perceptual difficulty of the contrast, the experience with L2 and the native dialect were the factors deemed to influence the perception of the L2 listeners. Given the reduced number of studies exploring the effects that the native variety has on non-native perception, this research aims to contribute information on such effects. The nature of the L2 contrasts analyzed here and the role of experience are also relevant to language acquisition and language teaching involving two widely spread linguistic varieties, English and Spanish.

2. PREVIOUS STUDIES

2.1. Low and mid back vowels and inherently difficult contrasts

The domain of low and mid back vowels represents a source of perceptual confusion even for native listeners. In English, for instance, /ʌ/ is mistaken for /ɑ/, /ɔ/ or /ʊ/ and is identified correctly only in 88.7% of the cases (Syrdal and Gopal, 1986). Also, the neighbouring vowels /ɑ/ and /ɔ/ are incorrectly classified as /ʌ/ by native speakers. On the whole, more confusion is observed for back vowels, which suggests that the acoustic cues that signal back articulations are weaker and more likely to produce perceptual ambiguity.

Lindblom (1986) attributes this effect to the reduced mobility of articulators and sensory control at the back of the mouth that correlates with less salient acoustic-perceptual phenomena (as compared to the front). Contrasts between low and mid back vowels may be difficult to perceive, particularly if the perceptual distance between them is small. Thus, the /ɑ-ʌ/ contrast, which involves two back vowels, is inherently more difficult to perceive than /ʌ-æ/ and /æ-ɑ/ contrasts involving back and non-back vowels. Native and non-native listeners alike obtain higher error rates and longer and more variable response times (Polka, 1995) when discriminating such 'difficult' contrasts. In a study that investigated cross-linguistic discrimination of the English vowel contrasts, Flege (1995) found that English /ɑ-ʌ/ and /æ-ɛ/ contrasts were those that generated the highest error rates among non-native listeners with various L1s, including Spanish. As a matter of fact, the /ɑ-ʌ/ contrast was the most problematic even for L2 listeners (the L1 Dutch and German groups) whose overall performance was otherwise similar to that of English natives. Performance on the /æ-ɑ/ contrast was also poor for the L1 Spanish group. That supports the idea that, even without considering factors like transfer from L1 and inventory size, contrasts involving back vowels are inherently more difficult.

Not only are such contrasts problematic for native and non-native listeners, but also performance with some difficult L2 contrasts improves slowly with experience or training. For instance, after a perceptual training experiment on five English vowels, Japanese learners succeeded in learning the temporal cues of the /ɑ-ʌ/ contrast, but failed to attend to spectral cues (Lambacher et al, 2005). In the same vein, Levy and Strange (2008) showed that perception of non-native contrasts between French rounded vowels improved with experience but unevenly as, occasionally, inexperienced listeners performed better than experienced listeners on some contrasts. Specifically, for the /u-y/ contrast in alveolar context, the difference in error rates between inexperienced and experienced listeners was only 4%, and in bilabial context the inexperienced actually had lower error rates than experienced listeners (8% versus 25%).

2.2. Cross-linguistic and cross-dialectal differences in perception

2.2.1. Cross-linguistic differences in perception

Perception is language specific, that is, native listeners refine their perception to recognize automatically the contrasts in their language. Since phones in languages differ in many ways, perceptual strategies that listeners adopt also vary cross-linguistically. Fox et al. (1995) showed that English and L2 Spanish listeners perceive English vowels differently. Specifically, whereas English monolinguals use 3 dimensions to categorize vowels (height, backness and central/non central distinctions), Spanish listeners use only 2 dimensions (height and proximity to a prototype vowel). Moreover, vowel height tends to be strongly correlated with duration for American English monolinguals but not for Spanish native speakers. The vowel inventory size also plays an important role in perception (Flege, 1995, Wagner and Ernestus, 2008). The greater the vowel inventory of a language is, the greater the number of dimensions necessary to perceive contrasts is. However, if a particular area in the perceptual vowel space is crowded, then perception is more sensitive to fine-grained differences among phones in that vowel space, so perception is warped by the native inventory. This is true for native listeners' perception of L1 contrasts. However, the task of non-native listeners is different and more difficult particularly if their native language inventory is smaller than that of L2. For instance, in Spanish the low central area in the perceptual space is committed only to the vowel /a/, whereas in English the corresponding area is occupied by three vowels /æ, ʌ, ɑ/. With fewer phonetic categories to attend to in L1, non-native listeners of L2 have to learn to reattune their perception to the specific contrasts in L2.

2.2.2. Cross-dialectal differences in perception

Perception of listeners from distinct varieties of a language may differ, too. Escudero and Boersma (2004) found dialectal differences between Scottish and Southern English in the perception of the /i-ɪ/ contrast. Whereas Scottish English listeners favoured the spectral cues, Southern English listeners perceived the contrast based on a combination of spectral and temporal cues. A similar finding is reported for the French vowels /o-ɔ/ and /ɑ-a/ in two dialects, Standard and Swiss French (Miller and Grosjean, 1997). In contrast with Standard French that uses mainly spectral cues, in Swiss French duration is given a more important weight in vowel identification. Thus, cross-dialectal differences are reflected in different perceptual strategies of weighting spectral and temporal cues.

2.2.3. Native dialect effects in non-native perception

A small number of studies showed that cross-dialectal differences in the perceptual strategies have repercussions on the processing of non-native phones. For instance, Holden and Nearey (1986) report such effect in three Russian varieties. Although these dialects have identical phonemic inventories, vowels display different distributions in the perceptual space, which seems to affect the listeners' perceptual behaviour in L2. Depending on the native variety, the non-native vowel /ʌ/ is perceived as [a], [o], or [e]. Morrison (2008) compared non-native perception in Mexican and Peninsular Spanish and found dialectal differences in the identification of the Canadian English front high and mid vowels. Guitart (1996, citing Valle, 1995) discusses an experiment in which the English /ʌ/ is identified as [a] in 83% of cases by Peninsular and as [o] in 71% of the cases by Caribbean learners. In his phonological interpretation, Guitart speculates that listeners with distinct dialectal backgrounds create different hierarchies of features based on the acoustic saliency of features like [+low] or [+round]. However, no clear justification for the listeners' preference for one or the other realization is given.

3. CURRENT STUDY – GOALS AND HYPOTHESES

The studies reviewed in **2.** point to the fact that some contrasts are inherently more difficult than others. The arguments are (i) the nature of the cues they encode, (ii) the fact that native and non-native listeners alike have higher error rates and longer response times with the 'difficult' contrasts and (iii) that experience or training may not result in great improvement in performance. In identification tasks, Spanish listeners assimilate the English /æ/ and /ɑ/ most often to [a] and /ʌ/ to [a] or [o]. In discrimination tasks Spanish listeners make more errors with the /ɑ-ʌ/ and /æ-ɑ/ contrasts. There are cross-linguistic and cross-dialectal differences in perception that can be attributed to different vocalic inventories and different perceptual strategies of cue weighting. A small number of articles supported the idea that the native dialect shapes non-native perception.

Based on these findings, the present study analyzes the influence that (1) the contrast inherent difficulty, (2) the experience with L2 and (3) the native variety have in non-native perception of Canadian English vowels /æ, ʌ, ɑ/. To test the first question regarding the difficulty of the English contrasts among low and mid back vowels, I hypothesized that these contrasts can be hierarchically ordered, with /ɑ-ʌ/ as the most difficult, followed by /ʌ-æ/ and /æ-ɑ/ (*hypothesis 1*). This hierarchical pattern, reflected in the discrimination error rates and response times, can be observed for all groups of listeners tested: non-native learners of English (Spanish experienced learners of English), non-native non-learners (Spanish monolinguals) and native English listeners.

Bearing on hypothesis 1, the second hypothesis addressed the role of experience with the L2 contrasts in discrimination performance. It is expected that experience with L2 contrasts will determine lower error rates for the learner groups especially with the 'easy' contrasts /ʌ-æ/ and /æ-ɑ/ whereas for /ɑ-ʌ/ contrast the error rates will drop less dramatically in the advanced group as compared to the monolingual group (*hypothesis 2*).

As I have argued that back vowels are perceptually confusable, I specifically investigated whether the /ɑ-ʌ/ contrast was more difficult for one dialectal group than for the other. I assume that in processing this contrast, listeners are likely to use categories situated in the low and back perceptual space of L1, that is, /a/ and /o/. Thus, *hypothesis 3* states that both groups of listeners use the perceptual strategy of shifting the boundary between their L1 vowels /a/ and /o/. The difference is that PS listeners shift their /a-o/ boundary towards /a/ whereas CS listeners towards /o/. If this is the case, different types of confusions with low and mid low vowels are expected for each group. Specifically, if Cuban listeners tend to perceive a back (rounded mid low) vowel for the L2 /ʌ/, they are more likely to err with back vowel contrasts /ɑ-ʌ/, as for them such contrasts represent a within-category contrast. Conversely, if Peninsular listeners tend to form fronted low unrounded percepts for the L2 /ʌ/, they will have a higher error rate with contrasts involving front vowels, /ʌ-æ/ and /æ-ɑ/.

4. METHODOLOGY

A two-talker AX discrimination task was chosen for the experiment. In this paradigm, subjects had to attend to salient phonetic differences between pairs of stimuli and answer whether the words they heard were the same or different. This task operates at the phonetic level rather than at the categorization (phonological) level and thus the discrimination task is suitable both for L2 learners and non-learners. The phonetic interpretation was also tapped into by setting the interstimulus interval at 1 second (Werker and Logan, 1985). The test was performed in one session that lasted 30 minutes and was part of a larger experiment that elicited L1 and L2 production data, which is not reported here.

4.1. Stimuli and materials

Two female native speakers of Canadian English (Southern Ontario) pre-recorded the stimuli, which were then extracted from the context and paired to create the set of perceptual testing material. The recording equipment used included an M-Audio Microtrack 24/96 professional 2-channel mobile digital recorder and a lavaliere unidirectional microphone. The tokens were real English words with a C_1VC_2 structure, with C_1 a stop or the glottal fricative /h/ and C_2 a stop. There were 6 minimal triads (e.g. *hat – hut – hot*) and several minimal pairs (e.g. *buck – back, duck – doc, tap – top*). A block of 72 pairs was created by pasting together various tokens separated by a one-second ISI. The set included: an equal number of 'same' and 'different' pairs (36); an equal number of target vowel pairs /æ-ɑ/, /ʌ-ɑ/ and /ʌ-æ/ (24); within the 'different' pairs, tokens produced by one speaker combined with an equal number of tokens produced by the other speaker (18); within the 'same' pairs, an equal number of identical tokens, same name – same speaker and same name – different speaker (18). Nine distracter pairs were also added; the items in these pairs have the same vowel but differ in voicing of the coda stop. The 81 pairs were randomized and uploaded twice to the perceptual testing software, thus yielding two blocks and a total of 162 audio files of token pairs. A short trial test was set up to familiarize participants with the software and testing material.

The perceptual testing software was developed in LabView 7.1 and runs under Windows on a portable computer. Two function keys of the computer (F1, F12) were assigned as decision buttons 'same' and 'different'. The answers and the corresponding response times were stored automatically.

4.2. Participants

Three groups totalling 34 participants were recruited for the study: 14 from León, Spain and 14 from Holguín, Cuba. Each of the two non-native groups, Peninsular and Cuban, consisted of two subgroups, one of advanced learners and one of monolinguals. The criteria used to assign participants to the experienced group were the formal training in English and the use of English in everyday activities. Participants in both advanced groups within the Peninsular and Cuban groups had obtained or were pursuing a university degree in English language and literature and used English more than 10 hours/week. The monolingual groups had little or no exposure to English.

The Peninsular Spanish group included seven subjects (6 female, 1 male), with ages 23-43, mean 31.4. They had extensive formal training in English as they were students in the 3rd year (5) or had graduated (2) and used English on a daily basis. The monolingual group from Spain included seven female subjects with ages 27-48, mean 35.7 with university education, who reported having minimal or no exposure to English.

The Cuban advanced group consisted of seven participants (6 female, 1 male) with ages 24-38, mean 31, with university degrees in English language and literature (6) and Teachers College (1). They all worked in the Public Relations Office in a tourist resort and reported using English on a daily basis with foreign tourists. The monolingual group consisted of seven participants (6 female, 1 male) with ages 26-41, mean 32.4. They worked in the same tourist resort but they had little or no exposure to English, nor did their jobs require any interaction with foreign tourists.

Participants reported no hearing problems and were paid Cnd$ 15 (or the equivalent in the local currency).

5. RESULTS

The number of incorrect answers was tallied for each contrast and converted into error rates. Participants' mean error rates and standard deviations are reported for each contrast and subgroup in Table 1.

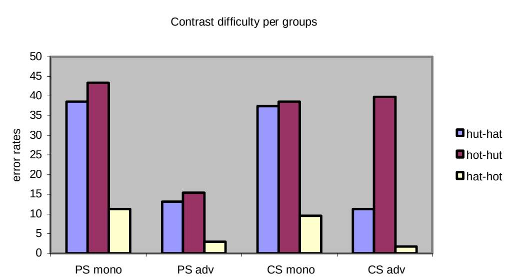
Table 1 – Mean error rates and standard deviations for each contrast and group

	/ɑ - ʌ/		/ʌ - æ/		/æ - ɑ/	
	error rate	*s.d.*	*error rate*	*s.d.*	*error rate*	*s.d.*
Peninsular monolinguals	43.3	(22.1)	38.6	(20)	11.3	(10.8)
Peninsular advanced	15.4	(12.2)	13.1	(12.6)	2.97	(5.2)
Cuban monolinguals	38.6	(19.8)	37.4	(14)	9.5	(8.2)
Cuban advanced	39.8	(24.6)	11.3	(10)	1.7	(3.2)

5.1. Contrast difficulty

Figure 1 shows that error rates and standard deviations were higher for the /ɑ-ʌ/ contrast and decreased for front-back vowel contrasts. The same error pattern was observed for all four groups (Peninsular and Cuban, monolinguals and advanced).

Figure 1. Error rates with /ɑ-ʌ/, /ʌ-æ/ and /æ-ɑ/ for 4 groups



5.2. L2 experience

Experience influenced performance with L2 contrasts as, overall, advanced learners from Spain and Cuba pooled had lower error rates than non-learners, as shown in Figure 1. More variability is observed for monolinguals for the /ʌ-æ/ and /æ-ɑ/ contrasts. A series of t-tests for independent samples yielded significant effects of experience for /ʌ-æ/ (p=.00013), and /æ-ɑ/ (p=.0158) and not significant for /ɑ-ʌ/. The analysis of error rates within the Cuban group produced an unexpected result: advanced learners have a higher error rate than monolinguals for the /ɑ-ʌ/ contrast (Table 1).

Table 2. L2 experience – error rates and standard deviations per contrast and experience groups

	/ɑ - ʌ/		/ʌ - æ/		/æ - ɑ/	
	errors	s.d.	errors	s.d.	errors	s.d.
mono (PS+CS)	41.1	(20.3)	38.1	(17.5)	10.4	(10)
adv (PS+CS)	21.3	(22.3)	14	(11)	2.3	(4.2)

5.3. Native dialect effects

The total number of errors for 'different' pairs was comparable for the dialects investigated (PS – 210, CS – 234) and the error rates were also similar for the front-back vowel contrasts. However, theare are cross-dialectal differences for the /ɑ-ʌ/ contrast, as shown in Table 3, with Cubans having a higher error rate (39.3%) overall as compared to Peninsular listeners (29.4%). A t-test for independent samples for the /ɑ-ʌ/ contrast returned a not statistically significant difference between means. However, there are statistically significant differences between dialectal subgroups of advanced PS and CS listeners (t=-2.349, p=.044)

Table 3. Native dialect - error rates per contrasts and dialectal groups

	/ɑ - ʌ/		/ʌ - æ/		/æ - ɑ/	
	errors	s.d.	errors	s.d.	errors	s.d.
PS	29.4	(22.5)	25.6	(21.5)	7.1	(9.7)
CS	39.3	(21.5)	24.7	(17.7)	5.6	(7.6)

6. DISCUSSION

This study investigated non-native perception in relation to three factors: the perceptual difficulty of the contrasts, the L2 experience and the effects of the native dialect. The hierarchy of difficulty predicted for the contrasts investigated here was confirmed. The overall discrimination accuracy was above chance, however the same pattern of errors emerged for all groups. Namely, the /ɑ-ʌ/ contrast was the most 'difficult', followed by /ʌ-æ/ and /æ-ɑ/. The mid back vowel /ʌ/ appeared in the problematic contrasts indicating that it is easily confused by L2 listeners. The hierarchy of difficulty obtained for L2 matches the pattern of difficulty reported for L1 (Lindblom, 1986, Syrdal and Gopal, 1986) and supports the finding that contrasts among back vowels are inherently more difficult to perceive as compared to contrasts among front vowels or front-back contrasts.

Experience with L2 contrasts had an effect on the listeners' performance as, overall, advanced learners discriminated the contrasts better than the non-learners. The contrast difficulty was correlated with higher error rates in the monolingual groups as compared to the advanced groups. For the 'difficult' contrast /ɑ-ʌ/ the error rate dropped 48% from advanced to monolinguals, 63% for /ʌ-æ/ and 77% for the 'easy' contrast /æ-ɑ/, which 70% of all listeners discriminated correctly. However, a closer look at the Cuban subgroups (Table 1) indicate that L2 experience had a negative effect as non-learners performed slightly better than learners. A similar effect is reported by Levy and Strange (2008) who showed that perception of the L2 French contrast /u-y/ improved with experience but unevenly as, occasionally, inexperienced listeners performed better than experienced listeners in particular consonantal contexts.

Native dialect had an effect for advanced L2 learners but not on monolinguals in the discrimination of the 'difficult' contrast /ɑ-ʌ/. Advanced Cubans had significantly more errors than advanced Peninsulars with this contrast, but performed comparably the same for the other two contrasts /ʌ-æ/ and /æ-ɑ/. It is known that perception is sharper at category boundaries and less accurate far from these boundaries (Strange, 1995). In other words, between-category contrasts are better discriminated than within-category contrasts. The fact that the advanced Cuban group obtained a higher error rate with the L2 /ɑ-ʌ/ suggests that it represents a within-category contrast, which is farther from the interlanguage boundary they may have for this contrast (L1 /a/-/o/ boundary). On the other hand, the good performance that the Peninsular learner group shows with this contrast indicates that their interlanguage boundary of the /ɑ-ʌ/ contrast is closer to the L1 /a/-/o/ boundary or it may even be the case that /ɑ-ʌ/ is a between-category contrast. Thus, different error rates with the /ɑ-ʌ/ contrast indicate that learners from distinct native dialects have different mappings of the L2 vowels. Additionally, greater variability in discrimination accuracy points to a fuzzier boundary for the contrast for the Cuban advanced group. Learners adopt the same perceptual strategy, that of shifting the L1 boundary between the L1 vowels /a/-/o/, however, the extent and the direction of this shift differs cross-dialectally. An alternate interpretation using Best's (1995) Perceptual Assimilation Model, accounts for dialectal differences in terms of different assimilation patterns, a category-goodness assimilation for Cubans and a two-category assimilation for Peninsular learners.

7. REFERENCES

Best, C. 1995. A direct realist view of cross-language speech perception. *Speech Perception and Linguistic Experience in Cross language Research*. Strange W. (ed.) York Press Timonium. 171-204

Escudero, P., Boersma P. 2004. Bridging the gap between L2 speech perception research and phonological theory. *SSLA* 26. 551-585

Flege, J.E. 1995. Speech Learning Theory, Findings, and Problems. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, Baltimore: York Press. 279–304.

Fox, R.A., Flege J.E., and Munro M.J. 1995. The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis. In *JASA* 97 (4) 2539-2551

Guitart, J. M. 1996. On the source of variance among Hispanophones with regard to the way they perceive certain American English vowels. *1996 AATSP Biennial Northeast Regional Meeting*, University of Amherst

Guitart, J. M. 1985. The resolution of phonological ambiguity in a simulated English-Spanish borrowing situation. *Selected Papers from the XIIth linguistic symposium on Romance languages*. King, L.D. and C. A. Malley (eds.) University of North Carolina, Chapel Hill, John Benjamins Publishing Co., Amsterdam/Philadelphia Márquez, Pan Am. Univ. at Brownsville

Holden K.T. and T.M. Nearey (1986) A preliminary report on three Russian dialects: Vowel perception and production. *Russian Language Journal*. 40. 3-21

Levy, E. S., Strange W. 2008. Perception of French vowels by American English adults with and without French language experience. *Journal of Phonetics* 36. 141-157

Lindblom, B. 1986. Phonetic universals in vowel systems. Ohala J., Jaeger J.J. (eds.) *Experimental Phonology*. Academic Press. Orlando. 13-44

Miller, J.L. and Grosjean F. 1997. Dialect effects in Vowel Perception: The Role of Temporal Information in French. *Language and Speech* 40 (3). 277-288

Morrison, G.S. 2008. Perception of synthetic vowels by monolingual Canadian-English, Mexican-Spanish, and Peninsular-Spanish listeners. *Canadian Acoustics* 36 (4). 17-23

Polka, L. 1995. Linguistic influences in adult perception of non-native vowel contrasts. *JASA* 97 (2). 1286-1296

Strange, W. 1995. Cross-language Studies of Speech Perception in *Speech Perception and Linguistic Experience in Cross language Research*. Strange W. (ed.) York Press Timonium. 3-45

Syrdal, A., Gopal, H. (1986) A perceptual model of vowel recognition based on the auditory representation of American English vowels. In *JASA* 79. 1086-1100

Wagner, A., Ernestus M. 2008. Identification of Phonemes: Differences between Phoneme Classes and the Effect of Class Size. *Phonetica* 65. 106–127

Werker, J., Logan, J. 1985. Cross-language evidence for three factors in speech perception. *Perception and Psychophysics* 37 (1), 35-44

Characterising the internal structure of learner intonation and its development over time

Ineke Mennen, Aoju Chen** & Fredrik Karlsson****

*ESRC Centre for Research on Bilingualism, Bangor University **Max Planck Institute for Psycholinguistics, Nijmegen ***Department of Language Studies, Umeå University
i.mennen@bangor.ac.uk; aoju.chen@mpi.nl; fredrik.karlsson@ling.umu.se

ABSTRACT

This paper describes the internal structure and longitudinal development of L2 learner intonation in two groups of three speakers from structurally different source languages (Punjabi and Italian) learning the same target language (English). The speakers were recently arrived immigrants in their host country and were followed over a period of thirty months. During this period, their intonation was analysed at two longitudinal points along three dimensions of the intonation system: (i) the inventory of structural elements (pitch accents and boundary tones); (ii) the way these are realised; and (iii) their distribution. Some attention is also paid to the functional aspect of learner intonation. Results show that there are many similarities across speakers in most dimensions of intonation. It seems that learners start out with the same inventory of structural elements from the start of their learning, regardless of the source language of the learners, although there are some differences in the frequency of use of certain patterns and the rate at which the system develops.

Keywords: intonation, learner variety, longitudinal development.

1. INTRODUCTION

Research into L2 acquisition of intonation is mainly concerned with comparisons of the L2 learner's production with that of native speakers of the target language (TL). In this approach, any observed differences are seen as 'imperfect, deficient imitations' of the TL standard (Klein and Perdue 1997: 306). While this target deviation approach certainly has its merits (particularly its methodological and pedagogical attractiveness), by focusing on 'deviations' it only provides a partial picture of the learner's system and fails to shed light on our understanding of the principles which underlie the acquisitional process. The intonational aspects that have been investigated in this way include amongst others whether L2 learners produce L2-specific pitch contours (Willems 1982), how they are realised (e.g. Jilka 2000; Mennen 2004, Kim, Curtis and Carmichael 2001) and how intonation is used in certain functions such as focus marking, turn-taking as well as paralinguistic functions (McGory 1997; Wennerstrom 2001; Chen 2009, 2010). Intriguingly, many similarities of errors (cf. Mennen 2007) were found in these target deviation studies, leading to assumptions about whether there are universal patterns in acquiring the system of a L2.

There is, however, very little empirical evidence for the assumption of universality. Firstly, the majority of these studies are restricted to investigations of L2 learners from various language backgrounds who acquire the same L1, namely English. It is therefore not clear whether the similarities observed among L2 learners reflect universal aspects in acquisition of intonation or the idiosyncrasies of the English intonational system. Research examining L2 learners acquiring different L2s is needed to shed light on this issue. Secondly, although findings from target deviation studies enrich our knowledge on learners' errors and imperfections in the production of L2 intonation, they tell us little about how a learner intonation variety is internally organised at a given time and whether it evolves over time. To our knowledge, no study has considered a learner intonation variety in its entirety. Yet, focusing on one aspect of learner intonation may lead to the erroneous interpretation of what certain 'deviations' imply and what has or has not been acquired. Finally, none of the above mentioned studies have investigated whether and how intonation develops longitudinally in adult L2 learners. Yet, in order to be able to fully understand the acquisitional process and to address issues such as ultimate attainment and fossilization, it is important to follow the development of learners (or the lack of it) over time.

In this paper we move away from a characterisation of the process of L2 acquisition of intonation in terms of errors and deviations, but rather describe it in terms of the systematicity which it exhibits. We will make a first attempt to describe the internal structure of intonation L2 learners start out with and how it evolves longitudinally.

2. THE LEARNER VARIETY APPROACH

This study took a 'learner variety' approach (Klein and Perdue 1997). To our knowledge this approach has never been applied to the analysis of L2 intonation.

2.1. The principles underlying the learner variety approach

The central assumption of this approach is that a learner variety is not an impoverished and distorted version of the TL but rather a system in its own right (Perdue 1993: 3). Adult L2 learners are somehow able to communicate quite early on in their L2 learning process, despite having only a few TL words at their disposal. Klein and Perdue (1997) demonstrated that these L2 utterances, which consist of only a few constituents, actually share a similar structure across learners, even when these learners have different source languages (SL) and are learning different TLs. They refer to this relatively stable and similar system as the 'basic variety', which can be characterised in terms of 'its lexical repertoire, [and] the principles according to which utterances are structured, and temporality and spatiality expressed' (ibid: 302).

2.2. Characterising the internal structure of learner intonation

In order to apply the learner variety approach to intonation, we need to investigate the internal structure of each learner variety in its own right. Ladd (1996: 119) recognises four areas which together make up an intonational variety and along which varieties can vary:

- the inventory of boundary tones and pitch accents ('systemic dimension');
- the phonetic implementation of these structural elements ('realisational dimension');
- the distribution of boundary tones and pitch accents ('distributional dimension');
- functionality ('semantic dimension').

The first three areas may be considered the basic structural aspects of the intonation system of a language, and it is these three aspects which we will focus on in the current paper. By analysing each learner variety along these different dimensions of intonation, we can ensure that the learner intonation variety is analysed in its entirety, rather than providing just a partial description of certain aspects of intonation. In this way, we can establish whether learners show regularities in their intonation system, in any or all dimensions of intonation. Furthermore, by following learners longitudinally, we are able to establish whether learner intonation varieties evolve over time in each of their structural dimensions, and whether there are regularities in the longitudinal development of the systems, regardless of inevitable differences between learners who acquire the TL in a naturalistic setting.

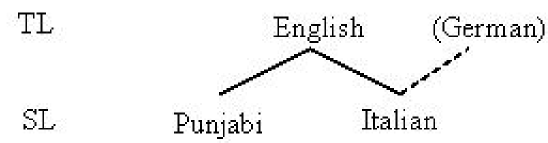
2.3. Our corpus

We used a balanced subset of the European Sciences Foundation (ESF) L2 Database to construct our corpus. The ESF L2 Database has recordings of various verbal tasks performed by untutored (not in classroom setting) L2 learners with different combinations of SLs and TLs over a period of 30 months after the recent arrival of the learners in their new language environment. Klein and Perdue (1997) argue that it is important to investigate untutored learners since classroom learning does not reflect the natural acquisition process.

The subset we used to construct our corpus covers the longitudinal data of two groups of speakers from a structurally different SL (Punjabi and Italian) learning the same TL (English), and a group of Italian learners learning a different TL (German), as exemplified in Figure 1. This way we will be able to control for both SL and TL influences. However, these results concerning the TL influences will not be presented in this paper. The two SLs in our corpus, Punjabi and Italian, are structurally different in that the former is a tone language (Baart 2004) and the latter is an intonation language (Avesani 1990). In total, we have data from 6 learners:

three Italian learners of English (Lavinia, Andrea, and Vito), and three Punjabi learners of English (Ravinder, Jarnail, and Mandan). We selected a total of 1279 utterances (768 statements and 511 questions) produced in free conversation between the learners and their interlocutors (native speakers of English). The utterances were selected from two longitudinal moments: during the 1st and 3rd 10 month cycle of recordings for all speakers except for Mandan for whom we used the 2nd and 3rd 10 month cycle, as recordings for the 1st cycle were not available. We will refer to these longitudinal moments as the first and last cycle.

Figure 1: The target languages (TL) and source languages (SL) in the corpus. The brackets and dashed line indicate that in these learners are not presented in this paper.



As we used an existing database to construct our corpus, we had no control over possible differences between learners such as their sex (we have five males and one female learner) or their region of origin or differences between learners after their arrival in the host country (such as their degree of exposure to the language of their new social environment, their amount of L1/L2 language use, or their motivation to learn). In fact, in the original ESF L2 Database a deliberate choice was made to not control for these factors with the aim to allow researchers "to discern (through the inevitable variability encountered while studying real-life learners who acquire at their own pace) the shared structural characteristics of their progress…" (Klein and Perdue 1997: 309). In other words, the focus was not on uncovering differences, but rather on finding similarities despite unavoidable individual variation between learners. Our study had a similar objective.

3. THE INTERNAL STRUCTURE OF LEARNER INTONATION

3.1. The systemic and distributional dimension of learner intonation

3.1.1. Pitch accents and boundary tones

Our results show that all learners, regardless of their TL, used the following pitch accents from the outset in their conversations: L*, H*, L*H, H*L, and !H*L. No complex tones (such as H*LH or L*HL) were used by any of the learners. Figure 2 shows that the distribution of these pitch accents is similar across the various SLs, with H*L being the most prevalent one. This is also the most prevalent pitch accent in statements and questions produced by native speakers of the TL variety, i.e. London English (Grabe 2004). Over time, we see a noticeable decrease in the use of the L*H pitch accent, both by the Italian as by the Punjabi learners. This may be a development towards the TL variety where the use of L*H (albeit in read speech) is reported to be around 10% (Grabe 2004).

We also found that all speakers use all three types of final boundary tones in their conversations: high (H%), low (L%) and level (%). Figure 3 shows the distribution of these boundary tones across the two SLs. It can be seen that the Punjabi learners of English use the level boundary tones more often than the Italian learners do. In terms of longitudinal development, we can see that there is a similar pattern across groups with an increase in the use of % and a decrease in H% boundary tones.

3.1.2. Intonation contours

Figure 4 shows the inventory and distribution of the five most commonly used intonation patterns used overall in the first and last cycle by both groups. The learners in each group use rather simple intonation patterns during the 1st cycle of recordings, consisting of no more than one pitch accent followed by a boundary tone. All learners use more falling than rising intonation patterns. This is true for both cycles, although even more so in the last cycle (when – as we have seen in section 3.1.1 – the use of the L*H pitch accent diminishes). As the majority of utterances in the corpus are statements, this is as expected. There are however some differences across the SL groups in the frequency of use of some contours, with more instances of the H*L % contour in Punjabi learners than in the Italian learners of English, whose most common contour is the L*H H%. Over time, there is a decrease of the use of the L*H H%. During the last cycle, most contours still contain just a single pitch accent. However, the fifth most common intonation pattern used during the last cycle by the Italian learners is the H*L !H*L %, consisting of two pitch accents followed by a boundary tone. This was used on average in 4.4% of intonational phrases (IPs) by the Italian learners (mostly Andrea and Lavinia). The Punjabi learners used very few double-accent tunes.

Figure 2: Distribution of pitch accents across source language.

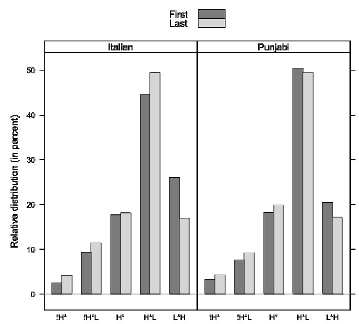


Figure 3: Distribution of boundary tones across source language.

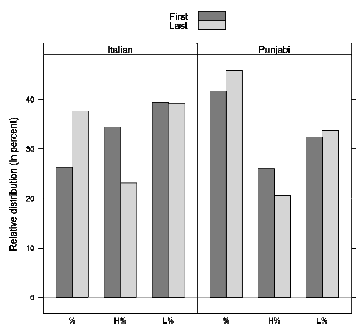
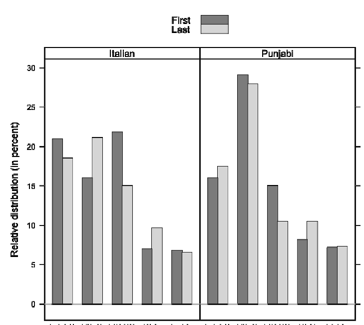


Figure 4: Intonation contours.



3.2. The realisational dimension of intonation

3.2.1. Intonational phrases

Intonational phrases (IPs) were very short for each learner, particularly during the first cycle. The IPs of the Italian learners consisted of 2.3 words on average. Those of the Punjabi learners were slightly shorter with an average of 2 words per IP. Over time, the length of IP increased in each group: from 2.3 to 2.9 words in the Italian learners and from 2.0 to 2.4 in the Punjabi learners. Alongside this increase in IP length, utterance length also increased, although more so for the Italian than the Punjabi learners. Utterance length increased from 3.1 words per utterance to 4.6 in the Italian learners, and from 2.9 to 3.4 in the Punjabi learners of English, and the number of IPs per utterance also increased for the Italians (from 1.4 to 1.6) but not the Punjabi speakers. These results show that although the learners produced longer utterances over time, they were broken up slightly longer (and in case of the Italians slightly more) IPs. This shows that the learners follow a similar developmental path (albeit slightly slower for the Punjabi than the Italian learners of English).

3.2.2. Number of pitch accents

The number of pitch accents per IP the learners produce during the first cycle of recordings is relatively small, as shown in figure 5, which gives each learner's median pitch accents per IP as well as the variability in the data. Both the Italian and the Punjabi learners produce on average 1.3 pitch accents per intonation phrase during the first cycle. This increases during the last cycle of recordings to 1.5 and 1.4, respectively. However, there are some individual differences. Lavinia, one of the Italian learners, shows an increase from a median of 1 to 2 pitch accents per IP. On the contrary, Jarnail (one of the Punjabi learners of English) shows no development at all. All other learners show an increase in variability of the number of pitch accents used.

Figure 4: The number of pitch accents per IP for the Italian (top row) and Punjabi learners of English (bottom row). Filled circles represent the median, unfilled circles represent outliers. The top and bottom of the squares (in this figure only tops are visible) present the 1st and 3rd quantile f the total distribution.

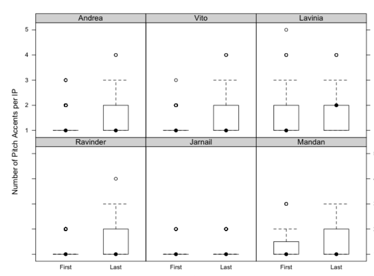
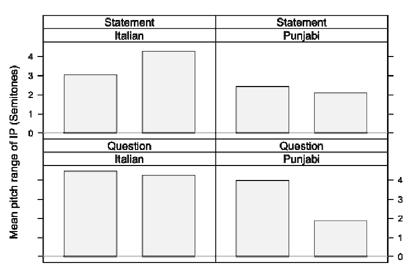


Figure 5: The overall pitch range of an IP. Box plot of pitch range of the IP, divided by source language.



3.2.3. Pitch range

Another aspect of the realisational dimension of intonation is the overall pitch range of utterances or IPs. High pitch is often used more frequently in questions than in statements. This high pitch can manifest itself in, amongst others, a more frequent use of final rises, higher peaks, a wider overall pitch range or a higher register (e.g. Ohala 1983, Haan 2001). We report how pitch range (measured as the 80% quantal range in semitones) is used by our learners to signal function, in this case to signal statements versus questions. Figure 5 shows that both the Punjabi and the Italian learners of English utilise this particular cue to mark interrogativity during the first cycle, with a wider overall pitch range in IPs of questions (4.5 ST for the Italian and 4 for the Punjabi learners) than in statements (with 3.0 ST and 2.4 ST respectively). During the last cycle, however, this pitch range difference between statements and questions has disappeared (possibly because the learners are relying on other cues to signal interrogativity, such as an increase in final rises). Finally, it can be seen that Punjabi learners have an overall narrower pitch range than the Italian learners (3.2 versus 3.8 respectively in the first cycle, and 2 versus 4.3 in the last cycle).

4. CONCLUDING REMARKS

The aim of this paper was to characterise the internal structure of learner intonation shortly after the learners' arrival in their host country, and to demonstrate how it evolves during the first 30 months of their stay. Our results show that from the outset (i.e. during their first 10 months in the host country) our learners appear to share the same inventory, which consists of all three final boundary tones (high, low and level) and five pitch accents (L*, H*, L*H, H*L, and !H*L). Learners combine these pitch accents and boundary tones into a set of basic intonation contours, consisting of just one pitch accent and a boundary tone. At this stage of learning, learners share the same set of most prevalent intonation contours, use more falls than rises, and all have the H*L as the their most prevalent pitch accent. They all break up their utterances into rather short IPs, containing two words on average, and have a wider pitch range in questions than in statements.

Over time, there is evidence of some development. Development appears to be rather slow and restricted to an increase in the length of IPs, an increase in the number of pitch accents they contain, and a change in the prevalence of certain pitch accents or contours. Learners appear to follow a similar developmental process by adopting more TL-appropriate patterns (i.e. less use of L*H, more falling intonation contours), although learners differ as to the rate of acquisition. No development in the inventory of pitch accents was observed between the first and the last cycle, with no evidence of complex contours being used by the learners (even though they are observed in London English, particularly in questions, Grabe 2004).

As mentioned earlier, in future studies we intend to also include a group of learners with the same SL (Italian) but learning a different TL (English and German), so that we can study similarities and differences in learners' intonational system due to acquiring different TLs.

5. REFERENCES

Avesani, C. 1990. A contribution to the synthesis of Italian intonation. *Proceedings of the International Conference on Spoken Language Processing*. 833–836.

Baart, J.L.G. 2004. Tone and song in Kalam Kohistani (Pakistan). In: Quené, H., van Heuven, V.J. (eds.), *On Speech and Language: Studies for Sieb G. Nooteboom*. Utrecht: Netherlands Graduate School of Linguistics. 5–16.

Chen, A. 2010. Intonation and reference maintenance in Turkish learners of Dutch: A first insight. *AILE (Acquisition et Interaction en Langue étrangère)* 1. 33–57

Chen, A. 2009. Perception of paralinguistic intonational meaning in a second language. *Language Learning 59 (2)*, 367–409.

Grabe, E. 2004. Intonational variation in urban dialects of English spoken in the British Isles. In: Gilles, P., Peters, J. (eds.) *Regional Variation in Intonation*. Linguistische Arbeiten, Tuebingen, Niemeyer. 9–31.

Haan, J. 2001. *Speaking of questions: an exploration of Dutch question intonation*. Utrecht: LOT.

Jilka, M. 2000. The contribution of intonation to the perception of foreign accent. PhD dissertation, University of Stuttgart.

Kim, S., Curtis, E., Carmichael, L. 2001. Emphatic Koreans and neutral Americans? *Journal of the Acoustical Society of America* 109 (5). 2474–2475.

Klein, W., Perdue, C. 1997. The basic variety (or: Couldn't natural languages be much simpler?). *Second Language Research* 13 (4). 301–347.

Ladd, D. R. 1996. *Intonational phonology*. Cambridge: Cambridge University Press.

McGory, J. T. 1997. Acquisition of intonational prominence in English by Seoul Korean and Mandarin Chinese speakers. PhD dissertation, Ohio State University.

Mennen, I. 2004. Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics* 32. 543–563

Mennen, I. 1997. Phonological and phonetic influences in non-native intonation. In: Trouvain, J., Gut, U. (eds.), *Non-native Prosody: Phonetic Descriptions and Teaching Practice (Nicht-muttersprachliche Prosodie: phonetische Beschreibungen und didaktische Praxis)*. Mouton De Gruyter.

Ohala, J. (1983). Cross-language use of pitch: An ethological view. *Phonetica* 40. 1–18.

Perdue, C. 1993. Adult language acquisition: cross-linguistic perspectives 2. Cambridge: Cambridge University Press.

Wennerstrom, A. 2001. *The music of everyday speech prosody and discourse analysis*. New York: Oxford University Press.

Willems, N. J. 1982. *English intonation from a Dutch point of view*. Dordrecht: Foris Publications.

Why are the Catalan contrasts between /e/-/ɛ/ and /o/-/ɔ/ so difficult for even early Spanish-Catalan bilinguals to perceive?

Joan C. Mora[1], James L. Keidel[2], James E. Flege[3]

[1]Universitat de Barcelona, [2]University of Manchester, [3]University of Alabama at Birmingham
mora@ub.edu, keidel@gmail.com, flegejames@yahoo.com

ABSTRACT

Previous research has shown that Catalan mid-vowel contrasts remain difficult for native Spanish adults to perceive, even those who began using Catalan as an L2 as young children. Here we tested 82 Spanish-Catalan bilinguals first exposed to Catalan by school age who differed in self-reported use of Catalan. All participated in categorization and AXB discrimination tasks using 10-step vowel continua. The endpoints of 2 continua (/i/-/e/, /u/-/o/) were vowels that contrast in both Spanish and Catalan whereas those in the remaining 2 continua were mid vowels (/e/-/ɛ/, /o/-/ɔ/) that do not contrast phonemically in Spanish. We hypothesized that Catalan mid-vowels contrasts are difficult not only because of cross-language phonological differences between Spanish and Catalan, but also because these phonemic contrasts are phonetically "weak" as a result of several factors. All participants perceived the high-mid contrasts less categorically than the mid-vowel contrasts but the magnitude of the difference seemed to depend on frequency of Catalan use. While not excluding possible effects of even brief delays in exposure to vowel contrasts found in an L2, these findings suggested the importance of variations in L1/L2 experience on the eventual perceptual performance of bilingual adults.

Keywords: Vowel perception, experience, maturational constraints, Catalan, Spanish.

1. INTRODUCTION

Age of onset of L2 learning is thought to affect L2 speech production and perception. According to some, age effects are caused by a loss of cerebral plasticity during the course of normal maturation (DeKeyser 2000). In fact, age of learning differences yield strong effects on overall degree of perceived foreign accent (see Piske et al. 2001). By now, a great deal of research has shown that early learners outperform late learners in specific aspects of L2 speech perception and production. Previous research has also reported robust effects of experience, normally operationalized as amount and type of L1/L2 input received, on bilinguals' speech production and perception between L1 and L2 phones (Flege and Liu 2001; Bohn and Flege 1990; Flege et al. 1997). One possibility, at least for individuals who learn an L2 upon immigration to an L2-speaking country, is that increased experience using the L2 may gradually augment perceptual sensitivity to L1-L2 phonetic differences, or to differences between L2 sounds that are not contrastive in the L1. This might increase the likelihood of new categories being established for sounds found in the L2 but not the L1 which, in turn may result in increased L2 production accuracy (Flege 1995, Flege 2002, 2007).

A series of studies carried out in Barcelona have demonstrated that even "early" learners of an L2 may not be fully native-like. This research has focused on contrasts between Catalan mid vowels (/e/-/ɛ/ and /o/-/ɔ/). These contrasts are perceptually difficult for native speakers of Spanish, presumably due to "single-category assimilation" patterns (Best and Tyler 2007). For example, Catalan has two vowels (/e/, /ɛ/) in the portion of vowel space where Spanish has just one, /e/ (Bosch et al. 2000). Single-category assimilation operates on Spanish-Catalan inter-phonology, often leading to Spanish learners' functional "deafness" to the contrast. Research using a variety of techniques has shown that Catalan mid-vowel contrasts are difficult even for native speakers of Spanish who began learning Catalan as young children (Sebastián-Gallés and Soto-Faraco 1999; Sebastián-Gallés and Bosch 2005, Pallier et al. 1997; Sebastián-Gallés et al. 2005). One might hypothesize that the L1 inevitably influences performance in a language learned later in life, even if the L2 is learned in early childhood.

The present study investigated another possible source of early Spanish-Catalan bilinguals' perceptual difficulty with the Catalan mid-vowel contrasts, one pertaining to the structure of Catalan rather than to Spanish-Catalan differences. There is reason to think that Catalan mid-vowel contrasts are less robust (Badia Margarit 1969, 1970; Recasens 1993) than contrasts between other pairs of Catalan vowels. Mid-vowel contrasts are not consistently realized in a relatively large number of words in the dialect of Catalan spoken in Barcelona , where the research just cited was carried out. Variation exists in how mid vowels are realized in dialects where mid vowels do contrast consistently (e.g. Eastern Catalan /'bɛwrə/ *beure* 'to drink' vs. Western Catalan /'bewrə/ (Carrera-Sabaté and Fernández-Planas 2005; Recasens and Espinosa 2006). Moreover, the implementation of mid vowels may vary across lexical items in a single dialect (*res* 'nothing' /res/-/rɛs/). Also, there is a general tendency in Catalonia for young people to neutralize mid vowel contrasts, for example, to pronounce /ɛ/ with an [e]-quality vowel (Recasens 1993: 86).

Other factors may also contribute to a reduction of robustness in Catalan mid-vowel contrasts. One is low functional load. There are relatively few Catalan minimal pairs involving the contrast /e/-/ɛ/. Vowel reduction processes in unstressed syllables may neutralize /e/-/ɛ/ into /ə/, and there may be inconsistencies in the articulation of root vowel (e.g. pes /pɛs/ 'weight' vs. peso /'pezu/ 'I weigh', cf. pesar /pə'za/ 'to weigh'). Finally, native Spanish speakers of Catalan often fail to produce an effective contrast between Catalan mid vowels (Badia Margarit 1970). Foreign-accented input may further obscure the phonetic nature of Catalan mid vowel contrasts for those learning Catalan in Barcelona.

Surprisingly little research has been carried out to test the hypothesis that a lack of robustness in Catalan mid vowel contrasts may help explain perceptual difficulties evidenced by early native Spanish learners of Catalan. A recent study by Mora and Nadeu (2009) compared two groups of Spanish-Catalan bilinguals. The members of both groups spoke Catalan from birth and were dominant in Catalan. The two groups differed only in terms of self-reported Catalan use. Neither group showed a fully categorical perception of /e/-/ɛ/. Interestingly, the less frequent users of Catalan showed significantly higher response latencies when discriminating /e/-/ɛ/, suggesting experience-related effects on processing speed.

These preliminary findings are consistent with recent research on the impact of phonological variation on phoneme perception investigating the robustness of the French vowel contrast /e/-/ɛ/, which is in the process of merging in Northern France. Brunellière et al. (2009) tested French-speaking participants from Switzerland on the perception of the unstable /e/-/ɛ/ contrast vs. the stable /ø/-/y/ contrast by means of a same-different task using behavioural and electrophysiological (ERP) measures. Their ERP results show that the two contrasts were processed differently. Specifically, the stable /ø/-/y/ contrast was discriminated faster and better than the unstable /e/-/ɛ/ contrast. The ERP results were consistent with the behavioural data, which showed higher error rates and slower responses for /e/-/ɛ/.

Taken together, findings from this and previous research on the Catalan mid-vowel contrasts call for further research on the degree of robustness of the mid-vowel contrasts and the effect of L1/L2 experience on bilingual vowel perception.

2. METHOD

This study assessed degree of perceptual "robustness" of the Catalan mid-vowel contrasts for Spanish-Catalan bilinguals who were exposed to their L2 (Spanish or Catalan) as young children and reported using both of their languages on a daily basis. It evaluated the effect of L1/L2 experience on the perception two Catalan mid vowel contrasts and two other Catalan vowel contrasts which served as controls.

It was impossible to recruit a group of Catalan monolinguals in Barcelona under the age of 50 years owing to the pervasiveness of Spanish. However, careful participant screening and selection procedures yielded groups of bilinguals who differed according to their frequency of Catalan use. We assumed that the bilinguals who used Spanish infrequently would most closely approximate a monolingual Catalan group, had it been possible to recruit such a group. That being the case, the "mostly Catalan" group would show the most categorical perception of Catalan mid vowel contrasts. All of the bilinguals participated in categorization and discrimination (AXB) perception tasks based on acoustic spectral continua and a production reading-aloud task. Owing to space limitations, just the perception results will be reported here.

2.1. Participants

A total of 82 Spanish-Catalan bilinguals were selected from the 719 Spanish-Catalan bilinguals who responded to ads placed in public libraries located throughout Barcelona. A short telephone interview and a written linguistic background questionnaire provided the information for subsequent screening. All the selected participants were first exposed to Catalan before they went to school, or at the latest when they started school at the age of 5 or 6 years. All participants were born and raised in Barcelona. All reported being able to speak and understand both Catalan and Spanish.

As shown in Table 1, participants were assigned to one of four groups based on self-reported percent of use of Catalan in several contexts (at home, at work, on social occasions, with relatives, with friends, and overall). To facilitate discussion, these bilingual groups will be referred to as “mostly Spanish” (< 25% Catalan use), “S/C” (30-50% Catalan use), “C/S” (50-70% Catalan use), and “mostly Catalan” (> 75%).

Table 1: Characteristics of the 4 groups; “S” and “C” refer to Spanish and Catalan. SDs in parentheses.

	Mostly S (*n*=15)	S/C (*n*=22)	C/S (*n*=26)	Mostly C (*n*=19)
Self-reported % C use	11 (8)	40 (7)	63 (6)	86 (8)
Chronological age at test (years)	30 (10)	32 (7)	32 (7)	35 (8)
Years of residence in Barcelona	30 (9)	31 (9)	32 (7)	34 (7)

2.2. Method

The stimuli used in both identification and discrimination tasks were drawn from 4 synthetic vowel continua. In each, F1, F2 and F3 frequencies varied in 10 steps whereas vowel duration and F4 were held constant (289 ms, 3570 Hz). For example, in the /e/-/ɛ/ continuum, F1 ranged from 450-580 Hz, F2 from 1840-1700 Hz, and F3 from 2570-2430 Hz. The endpoints of two continua were both mid vowels, /e/-/ɛ/ and /o/-/ɔ/. In the other two continua (/i/-/e/ and /u/-/o/), one endpoint was a mid vowel, the other was a high vowel.

A total of 120 stimuli were presented for forced-choice in the identification task (4 continua x 10 stimuli x 3 repetitions; ITI=2 sec.). The AXB discrimination task was designed with a 2-step resolution and contained 256 stimuli (8 triads x 4 orders x 4 continua x 2 repetitions; ISI=1 sec, ITI=1.5 sec). DmDx display software was used to run both experiments in a quiet room in a session lasting about 45 min. The participants heard the stimuli over headphones and selected response alternatives shown on a computer screen. These stimuli in both tasks were presented in fully randomized blocks (one per continuum). The order in which the continua were tested was counterbalanced across the participants in each group.

2.3. Analysis

Several measures were derived from responses to the identification (ID) and discrimination (DIS) tasks for each continuum. The ID measures included the estimated location of the 50% crossover from one response category to the other (i.e., the “boundary”), the width of the category boundaries, the overall amount of change from one response category to the other (spectral effect scores), percent correct ID, and difference scores computed for pairs of stimuli that did/did not straddle the boundary. The DIS measures included overall percent correct, and percent correct for pairs of stimuli that did/did not straddle category boundaries. Correlations were computed to explore effects of language use as well as the relation between identification and discrimination.

An exploratory factor analysis (PCA) was conducted to explore the interrelationship between language use and other variables. A multiple regression analysis was used to compare the contribution of language use and L1 independent variables to explaining the ID and DIS scores. Finally, a mixed-design ANOVA was used to test for a possible interaction between Group (4 levels: Mostly S, S/C, C/S, Mostly C) and Vowel Continuum (4 levels, 2 representing contrasts found in Spanish as well as Catalan).

3. RESULTS

We began by exploring whether %C use was correlated with the various outcome measures obtained for high-mid contrasts. Of the 28 tests performed (14 variables x 2 continua), only 2 reached significance ($p < 0.05$). When the same tests were carried out for the two mid vowel contrasts, 8 correlations reached significance at the .05 level, 4 at the 0.01 level. These mostly involved boundary width and spectral effect scores measures for the /e/-/ɛ/ contrast, suggesting that accuracy in the perception of this contrast is related to amount of use of Catalan.

In general significant positive correlations showed that the subjects' ID and DIS scores (ID scores mainly) for the /e/-/ɛ/ contrast (as opposed to /i/-/e/, /u/-/o/ and /o/-/ɔ/) varied as a function of %C use. Because the /i/-/e/ and /u/-/o/ contrasts exist both in Catalan and Spanish and are more robust than the mid vowel contrasts, little variation in performance (ID and DIS tasks) was found among participants as a function of L1/L2 experience. Significant correlations between ID and DIS scores were found in particular for the /e/-/ɛ/ contrast. In general, and for all vowel contrasts, significant correlations between ID and DIS measures were relatively weak (r <.5). For the /e/-/ɛ/ contrast, ID measures capturing the degree of categoriality in the perception of the contrast generally correlated significantly with discrimination scores, suggesting that speakers who perceived the /e/-/ɛ/ contrast more categorically also discriminated the contrast more accurately (particularly in across-boundary pairs drawn from the continuum).

3.1. Identification

Category boundaries were generally located between steps 5 and 6 for the /i/-/e/ and /u/-/o/ continua, and between steps 4-5 for the /o/-/ɔ/ continuum. It occurred between steps 2 and 3 (*mean* = 2.7) for Mostly-S, indicating substantially more /e/ than /ɛ/ responses. The boundary width (the range of stimuli between the 0.25 and 0.75 proportion of identification of the closer vowel in each continua) was narrower, indicating a more rapid change from one category to another (and thus more nearly "categorical" perception), for the two high-mid vowel contrasts (1-1.5 stimuli) than for the two mid-vowel contrasts (3.3-4 stimuli). The boundaries were also narrower for /o/-/ɔ/ (2.3-3.2 stimuli) than for /e/-/ɛ/ (3.4-4 stimuli). A tendency was observed for category width to become narrower as use of Catalan increased, particularly for the mid-vowel contrasts. Finally, lower slope coefficients (obtained through curve fitting by means of logistic regression) indicated steeper ID functions (more categorical perception) for high vowels than for mid vowels and for back vowels than for front vowels. Identification scores were submitted to mixed between-within ANOVAs with Contrast (4 levels: /i/-/e/, /e/-/ɛ/, /u/-/o/ and /o/-/ɔ/) as the within subjects factor and %C groups (4 levels: mostly-S, S/C, C/S, mostly-C) as the between-subjects factor. Main effects for Contrast did not reach significance for the boundary location measure ($F(3, 70)=2.54$, $p=0.63$), but were found to be significant for the boundary width ($F(3, 76)=54.2$, $p<0.01$; no significant Contrast x %Cgroup interaction, $p=.937$) and the identification function slope coefficients ($F(3, 74)=12.3$, $p<0.01$; the Contrast x %Cgroup interaction was non-significant, $p=.519$). Further Bonferroni-adjusted pairwise comparisons revealed that for both the boundary width and the slope measures, all vowel contrasts, except /i/-/e/ vs. /u/-/o/, differed significantly at the $p<.01$ level. This indicated that Spanish-Catalan bilinguals perceived high-mid and mid-vowel contrasts in a significantly different way, mid-vowel contrasts being perceived less categorically than high-mid vowel contrasts. The ANOVAs failed to reveal a significant main effect of %Cgroup-related differences observed in boundary location, boundary width and slope ($p=.418$, $p=.463$, $p=.265$, respectively).

An examination of spectral effect scores (SES, the difference in number of /i/, /e/, /u/ and /o/ responses for the first two and the last two stimuli in the /i/-/e/, /e/-/ɛ/, /u/-/o/ and /o/-/ɔ/ continua) revealed a similar pattern of results. The SES scores were found to be much lower for mid-mid than for the high-mid vowel contrasts, indicating much shallower ID function curves (Table 2). The SES scores varied systematically as a function of subject group for both *mid-mid* vowel contrasts, participants who used Catalan showing a more substantial shift in judgments as the result of spectral changes (larger SES scores and thus steeper curves, indicating more categorical perception of the contrast) than those who used Catalan relatively seldom. A mixed ANOVA on SESs yielded a significant main effect for Contrast ($F(3, 76)=28.4$, $p<.01$, the Contrast x

%Cgroup interaction and the main effect of %Cgroup effect were non-significant, p=.243 and p=.214, respectively.

Table 2: Spectral effect score (steps 1-2 minus step 9-10). SDs in parentheses.

SES	high-mid				mid-mid			
	Mostly S	S/C	C/S	Mostly C	Mostly S	S/C	C/S	Mostly C
FRONT Vs	100 (<1)	98 (6)	96 (14)	99 (4)	16 (50)	42 (63)	57 (55)	61 (50)
BACK Vs	99 (4)	100 (0)	99 (3)	100 (0)	44 (72)	58 (50)	63 (54)	69 (52)

In general, the ID scores obtained were consistent with a difference in the degree of categoriality in perception between high-mid and mid-mid vowel contrasts. The effect of L1/L2 experience on the perception of the mid-vowel contrasts did not reach significance for these ID measures, despite the consistent tendency observed of higher %C use Spanish-Catalan bilinguals perceiving the mid-vowel contrasts more categorically.

3.2. Discrimination

Three main vowel DIS measures were used : Mean percent correct discrimination across all vowel pairs in the continuum (DIS_1), mean percent correct discrimination in across-category pairs (stimuli pairs 4-6 and 6-7) where a discrimination peak is predicted on the basis of the category boundary location (DIS_2), and a mean percent correct discrimination in within-category pairs (DIS_3). In general the DIS scores reflect the ID data and follow a very similar pattern in that high vowel contrasts were discriminated more accurately than mid vowel contrasts. Back vowel pairs were discriminated more accurately than front vowel pairs (Table 3).

Table 3: Discrimination scores. *SD*s in parentheses.

DIS		high-mid				mid-mid			
		Mostly S	S/C	C/S	Mostly C	Mostly S	S/C	C/S	Mostly C
1	FRONT	64,4 (7)	67,1 (10)	68,5 (9)	70,4 (10)	58,5 (10)	64,4 (8)	62,8 (9)	63,5 (13)
	BACK	72,6 (10)	72,7 (9)	72,4 (12)	75,4 (8)	67,7 (8)	68,1 (10)	72,1 (12)	73,7 (10)
2	FRONT	67,9 (19)	70,1 (15)	65,8 (17)	72,7 (14)	53,6 (16)	63,7 (14)	63,4 (13)	64,2 (20)
	BACK	79,1 (13)	79,5 (13)	78,8 (14)	81,9 (10)	75,0 (13)	71,2 (16)	77,2 (20)	79,9 (15)
3	FRONT	63,3 (6)	66,1 (11)	69,3 (9)	69,7 (11)	60,1 (11)	64,6 (9)	62,6 (9)	63,2 (12)
	BACK	70,4 (11)	70,4 (10)	70,2 (12)	73,2 (8)	65,2 (8)	67,1 (10)	70,3 (10)	71,7 (10)

DIS scores were submitted to mixed between-within ANOVAs with Contrast (4 levels: /i/-/e/, /e/-/ɛ/, /u/-/o/ and /o/-/ɔ/) as the within subjects factor and %C groups (4 levels: mostly-S, S/C, C/S, mostly-C) as the between-subjects factor. Main effects for Contrast were found to be significant for the three discrimination measures (F(3, 76)=36.3, p<0.01 for DIS_1; F(3, 76)=26.8, p<0.01 for DIS_2 and F(3, 76)=7.77, p<0.01 for DIS_3), no significant Contrast x %Cgroup interactions or main effects for %Cgroup were significant. Bonferroni-adjusted pairwise comparisons revealed that for the DIS measures, differences in percent correct discrimination between /u/-/o/ and /o/-/ɔ/ were signficant at the p<.05 level, and at the p<.01 level for differences between /i/-/e/ and /e/-/ɛ/. This suggests, in line with the ID data, that Spanish-Catalan bilinguals discriminated high-mid and mid-vowel contrasts significantly differently, mid-vowel contrasts (/e/-/ɛ/ in particular) being discriminated at a lower percent correct rates.

4. CONCLUSION

This study set out to investigate a possible source for the often reported difficulty of early Spanish-Catalan bilinguals' perceptual difficulty with the mid-vowel contrasts other than those attributable to differences in early vs. late exposure to Catalan, or the interaction of the L1/L2 phonological subsystems of Catalan/Spanish bilinguals, all of which have already been extensively researched (e.g. Sebastián-Gallés and Bosch 2005). We hypothesized, on the basis of linguistic evidence in previous research on the instability of the /e/-/ɛ/ contrast, that the Catalan mid-vowel contrasts /e/-/ɛ/ and /o/-/ɔ/ are phonetically "weak", when compared to the high-mid vowel contrasts /i/-/e/ and /u/-/o/. We tested this hypothesis by having 4 groups of Spanish-Catalan speakers differing in the amount of Catalan spoken daily (Mostly-S, S/CB, C/SB, Mostly-C) participate in categorical identification and discrimination tasks based on mid (/e/-/ɛ/ and /o/-/ɔ/) and high (/i/-/e/ and /u/-/o/) vowel continua. The results show that all Spanish-Catalan bilinguals perceived the high-mid vowel contrasts more categorically than the mid-vowel contrasts, and the size of the difference seemed to be consistently affected, in the case of the mid-vowel contrasts, by how frequently Catalan was used. Spanish-Catalan bilinguals using Catalan more often perceived the "weak" mid vowel contrasts more categorically.

5. REFERENCES

Badia Margarit, A. M. 1969. Algunes mostres de les igualacions ɛ=e i Ɔ=o en el català parlat de Barcelona. In: Badia Margarit, A. M. (ed.) *Sons i fonemes de la llengua catalana*. Barcelona: Publicacions de la Universitat de Barcelona. 97-103.

Badia Margarit, A. M. 1970. Les vocals tòniques e i o en el català de Barcelona. In: Badia Margarit, A. M. (ed.) *Sons i fonemes de la llengua catalana.* Barcelona: Publicacions de la Universitat de Barcelona. 141-190.

Best, C., M. Tyler. 2007. Non-native and second language speech perception. In: Bohn, O.-S., Munro, M. J. (eds.) *Language experience in second language speech learning.* Amsterdam: John Benjamins. 15-34.

Bohn, O.-S., Flege, J. E. 1990. Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics* 11. 303-328.

Bosch, L., Costa, A., Sebastián-Gallés, N. (2000). First and second language vowel perception in early bilinguals. *European Journal of Cognitive Psychology* 12(2). 189-221.

Brunellière, A., Dufour, S., Nguyen, N., Frauenfelder, U. H. 2009. Behavioral and electrophysiological evidence for the impact of regional variation on phoneme perception. *Cognition* 111. 390-396.

Carrera-Sabaté, J. Fernández-Planas, A. 2005. *Vocals mitjanes tòniques del català. Estudi contrastiu interdialectal.* Barcelona: Hosori.

DeKeyser, R. 2000. The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition* 22. 499-534.

Flege, J. E. 1995. Second language speech learning: Theory, findings and problems. In: Strange, W. (ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press. 233-277.

Flege, J. E. 2002. Interactions between the native and second-language phonetic systems. In P. Burmeister, T. Piske & A. Rohde (Eds) An Integrated View of Language Development: Papers in Honor of Henning Wode. Trier: Wissenschaftlicher Verlag, pp. 217-244.

Flege, J. E. 2007. Language contact in bilingualism: phonetic system interactions. In: J. Cole, J., Hualde, J. (eds.) *Laboratory phonology 9*. Berlin: Mouton de Gruyter. 353-382.

Flege, J. E., Liu, S. 2001. The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition* 23. 527-552.

Flege, J. E., Bohn, O.-S., Jang, S. 1997. Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics* 25. 437-470.

Mora, J. C., Nadeu, M. 2009. Experience effects on the categorization of a native vowel contrast in highly proficient Catalan-Spanish bilinguals. *Journal of the Acoustical Society of America* 125(4, Pt. 2). 2775.

Pallier, C., L. Bosch, Sebastián-Gallés, N. 1997. A limit on behavioral plasticity in speech perception. *Cognition* 64. B9-B17.

Piske, T., MacKay, I. R. A., Flege, J. E. 2001. Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29. 191-215.

Recasens, D. 1993. *Fonètica i fonologia*. Barcelona: Enciclopèdia Catalana.

Recasens, D., Espinosa, A. 2006. Dispersion and variability of Catalan vowels. *Speech Communication* 48(6). 645-666.

Sebastián-Gallés, N., Bosch, L. (2005) Phonology and bilingualism. In: Kroll, J. F. de Groot, A. M. B. (eds.) *Handbook of bilingualism: psycholinguistic approaches.* 68-87.

Sebastián-Gallés, N., Soto-Faraco, S. 1999. Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition* 72. 111-123.

Sebastián-Gallés, N., Echeverría, S., Bosch, L. 2005. The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language* 52. 240-255.

Phonological and phonetic realisation of different types of focus in L2 speech

Mary O'Brien & Ulrike Gut

University of Calgary & University of Augsburg
mgobrien@ucalgary.ca & ulrike.gut@phil.uni-augsburg.de

ABSTRACT

The aim of this study is to investigate the strategies utilized by advanced learners of English with L1 German in the phonetic realisation and phonological distribution of three types of focus: narrow, broad and contrastive focus. Previous studies have shown that many characteristics of L2 intonation lie in the phonetic realisation of phonological categories rather than a divergent phonology (Grosser 1997; Gut 2009; Mennen 2007). Ten female native speakers of German were recorded performing the following tasks: responding to questions about a series of pictures to elicit narrow, broad and contrastive focus and reading a series of all-new sentences in both German and English. Auditory analyses were carried out to determine pitch accent placement in the different conditions, and acoustic measurements were taken of pitch alignment and the extent of rises. Subjects showed phonetic differences in the realisation of broad, narrow and contrastive focus. Moreover, differing task effects were found between a highly monitored sentence-reading task a semi-spontaneous task.

Keywords: focus, intonation, information structure, segmental anchoring, pitch alignment

1. INTRODUCTION

This study is concerned with the phonological and phonetic focussing strategies of German learners of English in their L2 as compared to those utilized in their L1. In particular, the placement of pitch accents and their phonetic realisation are investigated. Previous research on German learners of English has shown that both the placement of focus and the phonetic realisation of the relevant pitch accent differ from native speaker productions. In contrast to earlier studies that have mainly investigated reading passage style, this study also analyses semi-spontaneous speech. The paper is structured in the following way: Section 2 introduces the terms focus and information structure and discusses their phonological and phonetic realisation. Section 3 reports on previous findings on L2 focussing strategies. The methodology and results of our study are presented in section 4. We provide a discussion and conclusions in section 5.

2. FOCUS, INFORMATION STRUCTURE AND PHONOLOGY

Focus refers to that part of a sentence which is prominent in terms of content and which forms a contrast to the background information of the sentence. Focus can be differentiated in terms of scope: Broad focus falls on entire sentence constituents or even entire sentences, whereas narrow focus applies to individual lexical items (Ladd 1980). The interaction between focus and intonation has been the subject of many studies and has proven to be notoriously difficult to describe (e.g. Lambrecht 1994). Typically, it seems that focus is marked by a pitch accent. If intonation is employed to mark focus, it can have two functions (Dik 1997): Either it is used for the purpose of contrast or it marks the information structure of an utterance (e.g. new vs. given information). The term information structure was introduced by Halliday (1967). He defines given information as recoverable from the preceding discourse, whereas new information is focal either in the sense that the speaker presents it as not recoverable or in the sense that it is contrary to some predicted or stated alternative. Chafe (1976) postulates three degrees of activation status of discourse elements: A given element is already active in the listener's mind at the time of the utterance. Accessible elements are activated from a previously semi-active state. New information is activated from a previously inactive state.

Lambrecht (1994) further divides accessible information into textually accessible, situationally accessible and inferentially accessible.

Several researchers have analysed the relationship between information status and pitch accents. Brown (1983) found that 87% of all new concepts and 79% of the accessible concepts are accented whereas only 4% of given information, which had already been mentioned in the text, have accents. Pierrehumbert and Hirschberg (1990) state that in English new information is assigned an H* pitch accent and that given information is either not accented or receives an L* accent. Listeners tend to expect new information to be accented and given information to be deaccented. In an analysis of a retold story, van Donzel and Koopmans-van Beinum (1995) found that 52% of all new items in a discourse are perceived as accented, but only 13% of all given items. Based on an investigation of the difference between given and accessible information in German reading passage style, Baumann and Hadelich (2003) report that listeners prefer given information to be deaccented and new information to be accented. Important to note, however, is that languages like German and English show gradience in emphasis, such that it may prove difficult for listeners to accurately perceive which element is focused (Gussenhoven 2004, p. 52). Baumann (2006) reports on the results of perceptual studies that have shown that the categories of given and new are indeed not polar.

It is very likely that the marking of the given/new distinction depends on speaking style. This was noted by van Donzel and Koopmans-van Beinum (1995), who were not able to replicate findings based on sentences that were read aloud in story retellings. Ito, Speer and Beckman (2004) found that in spontaneous dialogues given information is only deaccented if it occurs in the current discourse segment but not if it is given in terms of the entire discourse. They further report that the word's functional role has a large effect on accent distribution.

3. THE REALIZATION OF FOCUS IN L2 SPEECH

Various studies have found that L2 speakers display different focussing strategies than native speakers. Gut's (2009) analysis of the LeaP corpus showed systematic differences between native and non-native English in terms of nucleus placement. In non-native English, some pronouns receive nuclear accents that never do so in native English, and given or accessible information is not always deaccented. Both native and non-native speakers of English produce distinct pitch movements on pre-nuclear words more often on content words than on function words, but these pitch movements are usually rises in native English whereas they are falls in non-native English. Ramirez Verdugo's (2002) analysis of Spanish learners of English showed that they do not mark the intonational distinction between new and given information in the way native speakers do. She found that they often place the nucleus on given information. Equally, the nucleus is often placed on circumstantial information such as "at the moment" instead of on the last new lexical item. In addition, in tag questions, non-native speakers produce the main accent on the pronoun instead of on the auxiliary. In general, there is a tendency to locate the nucleus on the last word of an utterance notwithstanding its information status or word class. Ramirez Verdugo (2002) further observed that English native speakers mark new information with a fall and given information with a low rise. By contrast, the Spanish speakers of English do not mark the difference in information status intonationally and produce a fall in both cases. Furthermore, the pitch movement on the nucleus is smaller for the non-native speakers.

Previous research on German learners of English has further shown that the phonetic realisation of pre-nuclear rising pitch accents differs from native English in a sentence reading task (Atterer and Ladd 2004). While in native English, the lowest point of pitch is aligned with the beginning of the first consonant in the onset of the stressed syllable and the pitch peak is reached at the end of the onset, in their subjects' L2 English the lowest point in pitch is aligned with the beginning of the stressed vowel and the pitch peak is reached in the following unstressed syllable, similar to the pattern found in native German. By the same token, Trofimovich and Baker (2006) found a delayed pitch peak in L2 English produced by Korean native speakers in a sentence repetition task, which again mirrored the pitch alignment patterns of their L1. For Korean immigrants who had arrived in an English-speaking country in childhood, it was found that pitch alignment reached native values after a certain length of residence (Trofimovich and Baker 2007). The advanced German learners of English in Atterer and Ladd's study, by contrast, did not achieve native-like values of pitch alignment. Interestingly, the native speakers of Dutch in Mennen's (2004) study failed to produce native-like alignment patterns both in their L2 Greek and in their L1 Dutch. Thus, the question remains as to whether tonal alignment patterns are transferred.

The main goal of this study is to investigate the phonetic realization and phonological distribution of three types of focus: narrow, broad and contrastive focus produced by advanced learners of English with L1 German in both their L1 and L2. In particular, we examine whether these learners exhibit L1 transfer effects in their phonetic realisation of pre-nuclear rising pitch accents in their L2 English. Previous studies have shown that many characteristics of L2 intonation lie in the phonetic realisation of phonological categories rather than a divergent phonology (Grosser 1997; Gut 2009; Mennen 2007).

4. METHOD

4.1 Participants

Ten female native speakers of German, all of whom grew up and were living in or around Augsburg participated in the study. Their mean age was 25.3 years, ranging from 23 to 29 years. They had been speaking English for approximately 14.6 years, and all had intermediate to advanced proficiency in English, as was shown by the results of the Oxford Online Placement Test. Subjects were placed at the B2 (N= 4), C1 (N = 2) or C2 (N = 5) levels according to the Common European Framework of Reference for Languages. Seven of the subjects had spent time in an English-speaking country, on average 0.75 years in Canada (N = 2), England (N = 3) or Australia (N = 2). All were university students, and only two of them reported speaking dialect in their daily lives.

4.2 Data collection

All participants were recorded performing the following tasks: responding to questions about a series of pictures to elicit narrow, broad and contrastive focus based on the *Questionnaire on Information Structure* (Skopeteas et al. 2006) and reading a series of all-new sentences (Atterer and Ladd 2004). Unlike previous studies, we relied on a variety of task types in the L1 and the L2 as a means of investigating both transfer and task effects.

4.3 Analyses

The following two analyses were carried out:

1. an auditory analysis of nucleus placement and the marking of new and given information in utterances with different focus conditions;
2. acoustic analysis of pre-nuclear rises in broad, narrow and contrastive focus condition in both read speech and elicited speech.

For the first analysis, four utterances each produced by the speakers in their L1 German and L2 English were selected. Two of them were replies to the single-subject questions "*What is the man pushing*" and "*Who is hitting the ball*" and contained the given information on either the utterance-final object ("*The man is pushing **a table***") or the utterance-initial subject ("***The man** is hitting the ball*"). The other two elicited utterances were replies to the double-subject questions "*What is the man drinking and what is the woman drinking*" and "*Who is eating the apple and who is eating the banana*" that trigger given information either on the IP-final objects ("*The woman is eating **the apple** and the man is eating **the banana***") or the IP-initial subjects ("***The woman** is eating the apple and **the man** is eating the banana*"). Both authors determined the nucleus placement independently. Agreement was 100% for the native German data and 95% (38 out of 40) for the L2 data. In addition, three students were asked to analyse the L2 data, but their results did not increase the inter-rater agreement.

The analysis of the phonetic realisation of pre-nuclear rises was carried out for both read speech and elicited speech. The read speech consisted of 13 sentences with broad focus ("all-new") in German and 15 sentences with broad focus in English (Atterer and Ladd 2004). The elicited speech consisted of six utterances produced by the speakers in each German and English. Four of them (utterances 1 to 4) were produced in broad focus condition as responses to the question "*What's happening*". Utterance 5 was produced in a narrow focus condition as a reply to the question "*Who is hitting the ball*", while utterance 6 was produced in contrastive focus condition as a reply to the question "*Who is eating the apple and who is eating the banana?*".

1) A **house** is burning.
2) **A baby** is sleeping.
3) **A mother** is reading to her child.
4) **A cat** is swimming in the lake.
5) **The man** is hitting the ball.
6) The **woman** is eating the apple and the **man** is eating the banana.

In each utterance, it was determined whether a pre-nuclear rise was produced on the sentence element given in bold. If a rise was present, pitch alignment and pitch height was measured following Atterer and Ladd (2004): the beginning of the onset of the stressed syllable was marked C0, the beginning of the vowel was marked V0, the beginning of the coda was marked C1 and the beginning of the vowel in the following unstressed syllable was marked V1. The lowest pitch and highest pitch were measured in relation to these points.

4.4 Results

4.4.1 Marking of new and given information in L1 German and L2 English

Table 1 shows the nucleus placement of the 10 speakers in the four German and the four English utterances. In their L1 German, all speakers consistently produce the nucleus on the new information if it occurs utterance-finally, both in the single-subject and the double-object condition. Likewise, for double-subjects, new information is marked by the nucleus when it occurs utterance-initially. The single-subject question "*Wer wirft den Ball*" (*Who is throwing the ball*), however, triggers replies with the nucleus on the utterance-final *Ball* for two speakers (18 and 28). The results for L2 English are very similar. While the focus is exclusively on new information in both double-subject conditions and the single-subject utterance-final condition, in the single-subject utterance-initial condition two speakers (18 and 38) produce the nucleus on *ball* in their reply to the question "*Who is hitting the ball*". Two further speakers (28 and 34) produce equal focus on both *man* and *ball* in their elicited responses. As far as the focussing strategies in their two languages are concerned, only two speakers (34 and 38) show cross-linguistic differences: they produce no (clearly perceptible) focus on new information in their L2 where they had done so in their L1.

Table 1. Marking of new/given information with the nucleus by the 10 speakers in their L1 German and L2 English. * Listeners did not agree on nucleus placement in two utterances.

	Single initial	Single final	Double initial	Double final
L1 German	8/2	10/0	10/0	10/0
L2 English	6/2*	10/0	10/0	10/0

4.4.2 Phonetic realisation of broad and narrow focus in L1 German and L2 English

Table 2 shows the alignment of the lowest and highest pitch as well as the phonetic extent of the rise that the 10 speakers produce in their L1 German and L2 English in broad, narrow and contrastive focus condition, with three outlying values removed per language. The first major difference between the L1 and the L2 are the number of rises (L+H* or L*+H) that are produced in these conditions. While 40 out of the 70 utterances produced in German have a rise, only 23 out the 70 produced in English do so. Especially narrow focus is marked by the speakers with a fall rather than a rise in their L2 English. Further differences are observable in tonal alignment: in German in the broad focus condition, the pitch peak is reached, on average, 4.3 ms before the beginning of the vowel of the following unstressed syllable. In their L2 English, the speakers on average reach the pitch peak 27.7 ms after the beginning of the vowel. For narrow focus, the pitch peak seems to be reached earlier in L2 English than in L1 German, but numbers are too small to be interpreted meaningfully. In L1 German, narrow focus – in contrast to broad and contrastive focus – is realised with a pitch trough that occurs, on average, 23.9 ms before the onset of the stressed vowel. Contrastive focus is realised by the speakers with a much higher rise than the other two focus conditions in both their L1 and L2. Clear differences between the speakers' L1 and L2 can be seen in tonal alignment, where the pitch trough occurs on average 51.3 ms before the onset of the stressed vowel in L2 English but 3.5 ms after it in L1 German. Moreover, the pitch peak is reached before the onset of the following unstressed vowel in German, but after it in English.

Table 2. Tonal alignment and extent of rises produced on pre-nuclear rises in L1 German and L2 English in broad, narrow and contrastive focus condition in the semi-spontaneous data.

		L in relation to V0 (ms)	**H in relation to V1 (ms)**	**Rise (Hz)**	***Rises produced/total number of utterances***
Broad focus	L1 German	10.8	-4.3	47.7	*22/40*
	L2 English	1.9	27.7	55.1	*10/40*
Narrow focus	L1 German	-23.9	-50.4	31	*8/10*
	L2 English	55.4	5.1	28.3	*2/10*
Contrastive focus	L1 German	3.5	-20.6	87.7	*10/20*
	L2 English	-51.3	19.6	72.4	*11/20*

Table 3 presents the results for tonal alignment and the phonetic realisation of the rises produced by the speakers in their L1 German and L2 English in the read sentences with broad focus. It shows that tonal alignment in the two tasks differs systematically: while the pitch trough occurs slightly after the onset of the stressed vowel in the semi-spontaneous data, it occurs well before this (on average 42.8 ms in German and on average 49.8 ms in English) in the read data. Additionally, the pitch peak produced by the speakers in the read speech occurs well after (37.4 ms) the beginning of the following unstressed syllable and not before it as in the semi-spontaneous German speech. Furthermore, the extent of the rise is much greater in the German read speech. Although they are Southern speakers of German, the ten speakers align the pitch trough in relation to the beginning of the stressed vowel (-42.8 ms on average) like the Northern speakers analysed by Atterer and Ladd (2004) do (-39.4 ms). The pitch peak alignment is similar to that found by Atter and Ladd for the Southern German speakers (34 ms).

Table 3. Tonal alignment and extent of rises produced on prenuclear rises in L1 German and L2 English in broad focus condition in the read data.

		L in relation to V0 (ms)	**H in relation to V1 (ms)**	**Rise (Hz)**	***n***
Broad focus	L1 German	-42.8	37.4	80.6	*133*
	L2 English	-49.8	55.1	52.9	*120*

5. DISCUSSION AND CONCLUSIONS

The results show that L1 German speakers of English use systematically different focussing strategies in broad, narrow and contrastive focus conditions, in particular as far as choice of pitch accent, tonal alignment and the phonetic extent of rises are concerned. Likewise, they use systematically different tonal alignment and higher rises in read speech than in semi-spontaneous speech. Evidence for or against L1 influence, however, is contradictory: Like the non-native speakers analysed by Gut (2009), our speakers produced falls rather than rises on pre-nuclear pitch accents. In contrast to the findings reported by Ramirez Verdugo (2002) and Gut (2009), however, only two out of ten learners whose speech was analysed here produce focus on given information. This might either be due to most subjects' fairly high competence in the L2 or to the fact that the utterances that were elicited were very short and usually required focussing that coincided with the strategy of placing the nucleus on the last content word. Indeed, it was only in those utterances that required focus on sentence-initial new information where learners produced focus on given information. The observed difference in tonal alignment between the speakers' two languages cannot be explained based upon L1 influence.

The experiments further showed how difficult it is to reproduce findings for read speech in semi-spontaneous speech. Firstly, the less controlled nature of the elicited speech means that, unlike in the experiments carried out by Braun (2006) and Atterer and Ladd (2004), responses are more variable and utterances contain voiceless consonants that make phonetic measurements more difficult. Moreover, unlike the participants in Braun's (2006: 462) study, the speakers analysed here did not exclusively produce rises on focussed elements, not even in their L1, possibly because the semi-spontaneous utterances were much shorter than the reading material used in the experiment

6. REFERENCES

Atterer, M., Ladd, D. R. 2004. On the phonetics and phonology of 'segmental anchoring' of F0: Evidence from German. *Journal of Phonetics* 32. 177–197.

Baumann, S., Hadelich, K. 2003. Accent type and givenness: an experiment with auditory and visual priming. *Proceedings of the 15th International Conference of Phonetic Sciences, Barcelona, Spain*. 1811–1814.

Baumann, S. 2006. Information structure and prosody: Lingusitics categories for spoken language annotation. In: Sudhoff, S., Lenertová, D., Meer, R., Pappert, S., Augurzky, P., Mleinek, I., Richter, N., Schließer, J. *Methods in Empirical Prosody Research*. Berlin: Walter de Gruyter. 153–180.

Braun, B. 2006. Phonetics and phonology of thematic contrast in German. *Language and Speech* 49(4). 451–493.

Brown, G. 1983. Prosodic structure and the given/new distinction. In: Cutler, A., Ladd, D. R. (eds.), *Prosody: Models and Measurements*. Berlin: Springer Verlag. 67–77.

Chafe, W. 1976. Givenness, contrastiveness, definiteness, subjects, and topics. In: Li, C. (ed.), *Subject and Topic*. New York: Academic Press. 27–55.

Dik, S. 1997. *The Theory of Functional Grammar*. Berlin: Mouton de Gruyter.

Grosser, W. 1997. On the acquisition of tonal and accentual features of English by Austrian learners. In: James, A., Leather, J. (eds.), *Second language speech: Structure and process*. Berlin: Mouton de Gruyter. 211–228.

Gussenhoven, C. 2004. *The phonology of tone and intonation*. Cambridge: Cambridge University Press.

Gut, U. 2009. *Non-native speech: A corpus-based analysis of phonological and phonetic properties of L2 English and German*. Frankfurt: Peter Lang.

Halliday, M. 1967. *Intonation and Grammar in British English*. The Hague: Mouton.

Ito, K., Speer, S., Beckman, M. 2004. Informational status and pitch accent distribution in spontaneous dialogues in English. *Proceedings of Speech Prosody 2004*, *Nara, Japan*. 279–282.

Ladd, D. R. 1980. *The structure of intonational meaning: Evidence from English*. Bloomington: Indiana Press.

Lambrecht, K. 1994. *Information structure and sentence form: Topic, focus and the mental representations of discourse Referents*. Cambridge: Cambridge University Press.

Mennen, I. 2004. Bi-directional interference in the intonation of Dutch speakers of Greek." *Journal of Phonetics* 32. 543–563.

Mennen, I. 2007. Phonological and phonetic influences in non-native intonation. In: Trouvain, J., Gut, U. (eds.), *Non-native prosody: Phonetic description and teaching practice*. Berlin: Mouton de Gruyter. 53–76.

Pierrehumbert, J., Hirschberg, J. 1990. The meaning of intonational contours in discourse. In: Cohen, P., Morgan, J., Pollack, M. (eds.), *Intentions in Communication*. Cambridge, Mass.: MIT Press. 271–311.

Ramirez Verdugo, D. 2002. Non-native interlanguage intonation systems: a study based on a computerized corpus of Spanish learners of English. *ICAME Journal* 26. 115–132.

Skopeteas, S., Fiedler, I., Hellmut, S., Schwarz, A., Stoel, R., Fanselow, G., Féry, C., Krifka, M. 2006. *Questionnaire on Information Structure*. Potsdam: Universitätsverlag Potsdam. < http://www.sfb632.uni-potsdam.de/~d2/ > Last access 12 Feb 2010.

Trofimovich, P., Baker, W. 2006. Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition* 28. 1–30.

Trofimovich, P., Baker, W. 2007. Learning prosody and fluency characteristics of second language speech: The effect of experience on child learners' acquisition of five suprasegmentals. *Applied Psycholinguistics* 28. 251–276.

van Donzel, M. E., Koopmans-van Beinum, F. 1995. Prominence judgements and textual structure in discourse. *Proceedings 19 Institute of Phonetic Sciences*, Amsterdam, Netherlands. 11–23.

L2 Acquisition of Sentential Stress: Implications for UG*

Öner Özçelik

McGill University
oner.ozcelik@mail.mcgill.ca

ABSTRACT

This paper investigates the adult second language (L2) acquisition of sentential stress, and aims to provide insight into the role of UG in L2 acquisition. Based on the results of an experimental study with advanced L1 English learners of L2 Turkish, it is concluded that UG is available in adult L2 acquisition, and that interlanguage grammars are constrained by the options provided by UG. These conclusions are reached based on the finding that the subjects were able to use target-like prosodic structures in representing Turkish stress, and that they were able to do so in the correct context, despite the fact that L1 and L2 differ in this regard and that the subjects were previously taught an incorrect sentential stress rule that makes sense pedagogically but is impossible linguistically. L2 learners can, then, go beyond instruction, and, where relevant, they can even get rid of its negative effects, for their grammars prohibit mental representations that are not constrained by UG. L2 learners' success in this study cannot be explained based on input alone, for input data are conflicting in this domain. Negative evidence, likewise, seems not to be available. All in all, the study provides strong evidence for UG in L2 acquisition.

Keywords: L2 acquisition of stress, UG in SLA, L2 acquisition of prosody, prosody-syntax interface

1. INTRODUCTION

Turkish sentential stress is generally assumed, by grammarians and educators, to fall on the word immediately preceding the predicate, (see (1)), and second language (L2) learners of Turkish are explicitly taught this:

(1) <u>Taught rule</u>: Stress the element immediately preceding the predicate, as in the following sentence:

Ben	**çocúk**	gör-dü-m.
I	child	see-PAST-1st sing

"I saw a child."

Though this rule can capture most cases of sentential stress in Turkey, it is wrong and - crucially - linguistically impossible. Therefore, if L2 learners are guided by UG (e.g. White 1989), we expect them to figure this out even if the input is not sufficient to lead them to the correct analysis, supporting previous studies such as Belikova 2008, *forthcoming*. If, however, only domain-general problem-solving skills are operative in L2 acquisition (e.g. Bley-Vroman 1990), L2 learners could hypothesize unnatural grammars that make sense pedagogically. Whether L2 learners can successfully get rid of (1) or not, then, has important implications for L2 theory.

2. TURKISH PROSODY

The generalization in (1) captures most sentential stress cases in Turkish, because this position is also the focus position in this language (Inkelas & Orgun 2003; Kornfilt 1997); therefore, stressing this position will never be incorrect (on a focused interpretation), thereby making it a pedagogically reasonable generalization. Moreover, sentential stress sometimes happens to fall on this position, even when the pre-predicate constituent is not focused. However, this rule is, by no means, linguistically correct.

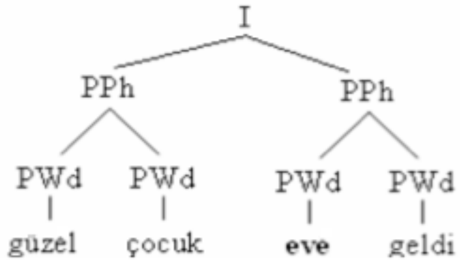
Turkish stress, in fact, falls on the first prosodic word (PWd) in a phonological phrase (PPh) (Kabak & Vogel 2001) (indicated in boldface in (2) and the rest of this paper), and on the last PPh in an intonational phrase (I) (Özçelik & Nagai 2009, 2010) (the head of which is underlined in (3) and the rest of this paper). Word-level stress, which will not be the focus of this paper, almost always falls on the last syllable of a word, except for some irregular cases (see e.g. Kabak& Vogel 2001; Özçelik 2010), and is indicated with an acute accent:

(2) a. [**güzél** çocúk]PPh b. [**ó** adám]PPh
beautiful kid that man
"beautiful kid" "that man"

(3) a. [[**güzél** çocúk]PPh [**ev-é** gel-dí]PPh]I b. [[**ó**]PPh [**adám**]PPh]I
beautiful kid home-Dat come-PAST that man
"The beautiful kid came home." "That is a man."

That is, in Turkish, the head of a PPh is the leftmost PWd, and the head of an I is the rightmost PPh, a case much more complex than is depicted by the rule in (1). Examine (4), where this is illustrated with tree level representations:

(4) a. [**güzél** çocúk]PPh b. [[**güzél** çocúk]PPh [**ev-é** gel-dí]PPh]I



3. INTERPRETATION OF BARE NOUNS AND PROSODY

Bare nouns in Turkish are ambiguous between a definite and an indefinite interpretation (see e.g. Göksel & Kerslake 2005; Kornfilt 1997; Özçelik & Nagai 2009, 2010). A sentence like *Man arrived* could, therefore, have two different readings, given below in (5a) and (5b):

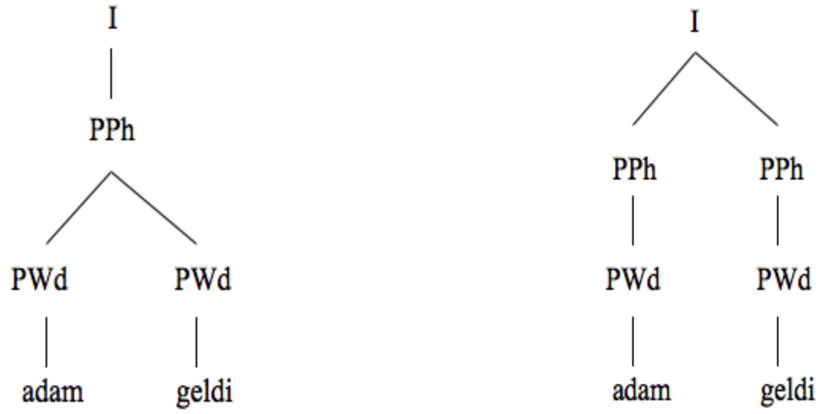
(5) a. adam gel-di b. adam gel-di
man arrive-PAST man arrive-PAST
"A man arrived." "The man arrived."

Though (5a) and (5b) look exactly the same on the surface, they differ in terms of prosodic structure. Whereas (5a) is composed of only one PPh, (5b) is composed of two PPhs, illustrated below in (6a) vs. (6b) (see also (2b) vs. (3b)) (examples from Özçelik & Nagai 2009, 2010):

(6) a. [[**Adám** gel-dí]PPh]I b. [[**Adám**]PPh [**gel-dí**]PPh]I
man arrive-PAST man arrive-PAST
"A man arrived." "The man arrived."

Tree representations for these two sentences are respectively given below in (7a) and (7b):

(7) a. b.



Özçelik & Nagai (2009, 2010) argue, based on these facts, that the indefinite *adam* 'man' in (6a) (or (5a)) remains within the same syntactic projection as the verb (and thus sharing the same phonological phrase with it, whereas the definite *adam* 'man' in (6b) (or (5b)) is external to the root-VP in syntax (thus creating its own phonological phrase domain.

Crucially, this means that sentential stress will fall on the first word in (6a) (=(5a)), and the last word in (6b) (=5b)), meaning that the rule in (1) will fail for sentences like (6b). Since the prosody in (6b) is also used for topicalization constructions in Turkish, input data will also not be sufficient to lead learners to the correct prosodic representations.

4. CURRENT STUDY

4.1. Hypotheses

Given L1 English and L2 Turkish, in accordance with the Full Transfer/Full Access (FTFA) (Schwartz & Sprouse 1996), we hypothesize that L2 learners of Turkish should fail, at advanced levels, to internalize linguistically-misleading classroom generalizations such as (1), and thus be able to acquire the difference between (7a) and (7b).

Success is also predicted by the Prosodic Transfer Hypothesis (PTH) (Goad, White & Steele 2003; Goad & White 2004), which allows full access to UG in syntax but access through existing L1 prosodic representations in phonology (though see Goad & White 2008). Accordingly, though English stresses the subject of unaccusative sentences with both definite and indefinite subjects (see e.g. Selkirk 1984, Zubizarreta 1998), using a structure like (7a), L1 English learners of Turkish should correctly be able to use (7b), too, either by combining two PPhs, a strategy that exists in L1 or by using the prosody of unergative constructions available in L1, which is similar to (7b), though for a different reason: in English unergatives, the rightmost PWd is the head of a PPh, as well as the rightmost PPh being the head of an I (i.e. all heads are rightmost unlike Turkish).

Being able to produce Turkish sentences with either prosody should not, then, be a problem for L1 English speakers, given the PTH. And given the FTFA, they should be able to use the correct prosody in the correct context, despite the L1-L2 differences in terms of the usage of these prosodic structures, differences in parameters in assigning head status to several prosodic constituents, and, of course, despite the rule (1).

4.2. Experiment

In order to test these hypotheses, several elicited production tasks have been conducted.

4.2.1. Subjects

The subjects were two advanced L1 English learners of L2 Turkish (confirmed independently by a cloze test). They got instructed according to the pedagogically reasonable/linguistically incorrect rule (1), as has been confirmed by a detailed examination of textbooks and an interview with the teacher.

4.2.2. Analysis

Each subject produced a total of about 2000 utterances during the interviews. There were three interviews with subject 1 and two interviews with subject 2. We focused on one comparison, that between sentences like (6a) vs. (6b), repeated below as (8a) vs. (8b):

(8) (=(7)) a. [[**Adám** gel-dí]PPh]I vs. b. [[**Adám**]PPh [**gel-dí**]PPh]I
man arrive-Past man arrive-Past
"A man arrived." "The man arrived."

Sentences like (9) or (10), on the other hand, were not included in the analysis:

(9) [[**Adám**]PPh [**ev-é** gel-dí]PPh]I → *vague between Rule (1) and the*
"The man arrived at home." *correct interpretation*

(10) [[**Iyí** adám gel-dí]PPh]I → *solves the vagueness of (8a), but*
"A good man arrive *very few examples in data…*

Forms like (9) were not included, because they are ambiguous with respect to what we are testing: if learners can produce such sentences with the correct stress pattern, it is not entirely clear whether this has been done because they have used the correct prosodic structures (i.e. based on the parameters PPh-head=leftmost and I-head=rightmost), or because they have simply employed the rule in (1). As for (10), even though this would help resolve a similar case of ambiguity caused by sentences with indefinite subjects, such as (8a), this was not included in the analysis, either, for there were very few examples of such sentences in learners' production. This could be due to the fact that when the noun is modified with an element like an adjective, there is a higher possibility for that noun to be definite than indefinite semantically.

4.3. Results

The results confirm our hypotheses. Both subjects stressed the correct constituent more than 95% of the time in both (6a) and (6b) type of sentences, simple sentences composed of a predicate and an indefinite or a definite subject. The results from both subjects are summarized below in Table 1 and 2 below:

Table 1: Results for Subject 1.

Subject 1	Sentences with an indefinite subject	Sentences with a definite subject
Correct	52	83
Incorrect	1	4
Percentage correct	98.11%	95.40%

Table 2: Results for Subject 2.

Subject 1	Sentences with an indefinite subject	Sentences with a definite subject
Correct	45	68
Incorrect	2	2
Percentage correct	95.74%	97.14%

5. DISCUSSION AND CONCLUSIONS

The purpose of this study was to see if L2 learners can acquire linguistic representations constricted by UG despite linguistically-misleading classroom generalizations such as (1), and given the lack of sufficient input to lead them to the correct analysis. We have argued that if UG is available to L2 learners of Turkish, they should be able to unlearn a rule like (1), which is impossible linguistically. In other words, if interlanguage grammars are restricted by the options provided by UG, there is no reason for them to assume that a UG-unrestricted rule such as (1) could hold true, no matter how pedagogically reasonable it is.

The results of the current study confirm our hypothesis: Our subjects correctly placed sentential stress on the first PWd within the last PPh in the I (as per the PPh- and I-level stress/prominence rules of Turkish, see Özçelik & Nagai 2009, 2010), and they did so more than 95% of the time, irrespective of this strategy stressing the predicate as in (6b) (a sentence with a definite subject) or the word (or phrase) immediately preceding the predicate as in (6a) (a sentence with an indefinite subject).

Notice that using rule (1) cannot help the learner reach the correct stress pattern for sentences like (6b), for this rule fails to account for the *non*-pre-predicate stress pattern observed in such sentences. Moreover, a simple strategy that says, "Stress the predicate in sentences with a definite subject, but stress the subject in those with an indefinite subject" would not work, either, for it is not always the case that sentential stress/prominence falls on the predicate when a sentence has a definite subject, as illustrated by examples like (4b). Here, sentential prominence happens to fall on the constituent immediately preceding the predicate, as would also be predicted by rule (1), for this constituent is the first PWd within the last PPh in the I. Our subjects' production of these sentences were also target-like, though the formal analysis presented here did not focus on this issue. As such, such a strategy, which makes use of definiteness and overlooks the facts of the syntax-prosody interface or the prosodic parameters of Turkish, could not have been used in the process.

The findings are in line with the FTFA and the PTH. As expected by the FTFA, though L1 and L2 differ with respect to the relevant prosodic and syntactic parameters, L2 learners were able to reach target-like representations, and this is possible only with access to UG, for neither teaching nor input is helpful to learners in the process. The results are also in line with the Full Access without Transfer (e.g. Flynn & Martohardjono 1994) (following White's 2003 usage of the term), however, since no lower-level learners of Turkish have been tested. As such, we do not know whether there would have been initial L1 effects if such learners had also been tested. This is an issue for further research to investigate. Likewise, as expected by the PTH, the learners were able to use both prosodic representations in (6)(=(7)), for both could be reached based on L1 prosodic structures, although the two languages employ different prosodic parameters. In sum, then, learners were not only able to access all the relevant prosodic structures as would be predicted by the PTH, but also they used them correctly in the correct context, as expected by UG-based approaches such as the FTFA.

Finally, the findings are clearly at odds with the Fundamental Difference Hypothesis (e.g. Bley-Vroman 1990) or other approaches that attribute no place to UG in adult second language acquisition, such as Clahsen & Hong 1995 and Beck 1998. According to these approaches, only domain-general problem-solving skills are operative in L2 acquisition, and adult L2 learners cannot access UG. We have, however, found, in this study, that L1 English learners of L2 Turkish were able to converge on the grammars of native Turkish speakers. That is, parameter resetting was also possible, contra 'no parameter resetting' approaches such as Hawkins & Chan 1997.

In fact, learners of Turkish were able to reset their parameters despite all the challenges of the task: First of all, instruction is not helpful to them in the process; in fact, instruction leads them to an incorrect analysis, as we have detailed above. Second, input data is not very helpful to get rid of the effects of wrong instruction: the prosodic structure in (6b), the one not captured by rule (1), is used also for topicalization constructions in Turkish. So a learner faced with such a prosodic structure will probably not feel the need to revise his or her knowledge gained from (1), but will instead assume that this is something different, something that involves topicalization, which doesn't pose a challenge to (1). Second, sentential stress happens to fall on the pre-predicate position in many cases (see e.g. (4)), thereby confirming the learner's initial incorrect assumption about the target language.

Likewise, negative evidence is also probably not available in this domain: A learner always *producing* sentences consistent with (1) will not be wrong (though problems might arise in *interpreting)*, for such sentences will be correct at least on a focused interpretation of the pre-predicate constituent. Therefore, an L2 learner of Turkish who consistently chooses the wrong stress pattern, one in line with (1), will not be corrected.

In conclusion, the fact that learners of Turkish were able to notice that rule (1) is not indeed possible could only have been doable under a view that interlanguage grammars are constrained by the options made available by UG. The learners of Turkish tested in this paper were not only able to notice that rule (1) does not hold true, but also they were able to attain the target-like representations even though (a) L1 and L2 differ with respect to the relevant parameters, (b) input is not sufficient, and (c) instruction points to a different analysis, implicating, again, the availability of UG in second language acquisition.

6. REFERENCES

Beck, M.-L. (1998). L2 acquisition and obligatory head movement: English-speaking learners of German and the local impairment hypothesis. *Studies in Second Language Acquisition, 20*: 311-348.

Belikova, A. Forthcoming. Getting L2 reciprocal and reflexive verbs right. Ph.D. dissertation, McGill University.

Belikova, A. 2008. Explicit instruction vs. linguistic competence in adult L2-acquisition. In H. Chan, E. Kapia & H. Jacob, eds., Proceedings of BUCLD 32, 48-59. Somerville, MA: Cascadilla Press.

Bley-Vroman, R. 1990. The logical problem of foreign language learning. *Linguistic Analysis* 20: 3-49.

Clahsen, H. & Hong, U. (1995). Agreement and null subjects in German L2 development: New evidence from reaction-time experiments. *Second Language Research*, *11*: 57-87.

Flynn, S. & Martohardjono, G. (1994). Mapping from the initial state to the final state: the separation of universal principles and language-specific principles. In B. Lust, M. Suner and J. Whitman (eds.), *Syntactic theory and first language acquisition: crosslinguistic perspectives.* Vol. 1: *Heads, projections and learnability* (pp. 319-335). Hillsdale, NJ: Lawrence Erlbaum.

Goad H, White L, Steele J. 2003. Missing inflection in L2 acquisition: Defective syntax or L1-constrained prosodic representations? *Canadian Journal of Linguistics* 48: 243-263.

Goad, H., & White, L. 2004. Ultimate attainment of L2 inflection: Effects of L1 prosodic structure. In S. Foster-Cohen, M. Sharwood Smith, A. Sorace & M. Ota, eds., *EUROSLA Yearbook* 4, 119-145. Amsterdam: John Benjamins.

Goad, H. & White, L. 2008. Prosodic structure and the representation of L2 functional morphology: A nativist approach. *Lingua* 118: 577-594.

Göksel, Aslı, and Celia Kerslake. 2005. *Turkish: A Comprehensive Grammar*. New York: Routledge.

Hawkins, R. & Chan, Y.-H. C. (1997) The partial availability of Universal Grammar in second language acquisition: the 'failed functional features hypothesis'. *Second Language Research, 13*: 187-226.

Inkelas, S. & Orgun, C. O., 2003. Turkish stress: a review. *Phonology* 20:139-161. 2003.

Kabak, B., & Vogel, I. 2001. The phonological Word and Stress Assignment in Turkish. *Phonology 18*, 315-360.

Kornfilt, J. 1997. *Turkish Grammar*. London: Routledge.

Nespor M., & Vogel, I. 1986. *Prosodic phonology*. Dordrecht: Foris Publications.

Özçelik, Ö. 2010. Exceptions in stress assignment: Feet in input. In the *Proceedings of the 40th Annual Meeting of the North East Linguistic Society (NELS 40)*. MA: GLSA.

Özçelik, Ö. & Nagai, M. 2009. Possible syntactic subject positions in Turkish: Evidence from phonology. Paper presented at the *Sixth Workshop on Altaic Formal Linguistics (WAFL 6)*, Nagoya University, Nagoya, Japan.

Özçelik, Ö. & Nagai, M. 2010. Multiple subject positions: A case of perfect match between syntax and prosody. Paper presented at the *28th Western Conference on Formal Linguistics (WCCFL 28)*. USC. Los Angeles, CA, USA

Schwartz, B. D. & Sprouse, R. 1994. Word order and nominative case in nonnative language acquisition: a longitudinal study of (L1 Turkish) German interlanguage. In T. Hoekstra and B. D. Schwartz (eds.), *Language acquisition studies in generative grammar* (pp. 317-368). Amsterdam: John Benjamins.

Schwartz, B. D. & R. Sprouse. 1996. L2 cognitive states and the full transfer/full access model. *Second Language Research* 12: 40-72.

Selkirk, E. 1995. Sentence prosody: intonation, stress and phrasing. In *The Handbook of Phonological Theory*, ed. John Goldsmith. London: Blackwell.

White, L. 1989. Universal grammar and second language acquisition. Amsterdam; John Benjamins.

White, L. (2003). *Second language acquisition and universal grammar*. Cambridge University Press: Cambridge, UK

Zubizarreta, M. L. 1998. *Prosody, focus, and word order.* Cambridge, MA: MIT Press.

NOTES

· Support for the dissemination of this research was provided by the Eileen Peters Fellowship awarded to the author. Thanks also to Heather Goad and Alyona Belikova for helpful comments!

Contributions of experimental phonetics to the didactics of the pronunciation of the French as a Foreign language: stage 1: reflection around the establishment of a speaking materials

Claire Pillot-Loiseau, Angélique Amelot, Florentina Fredet

Institut de Linguistique et de Phonétique Générales et Appliquées (ILPGA), and Laboratoire de Phonétique et Phonologie UMR 7018 CNRS : Université Paris 3 Sorbonne Nouvelle
claire.pillot@univ-paris3.fr

ABSTRACT

Our investigation of the pronunciation of learners of French as a Foreign Language (FFL) involves (i) selecting criteria for defining types of FFL learners; (ii) a preliminary reflection on the choice of materials to be recorded and the nature of the recordings; (iii) a feasibility test for the speaking materials; (iv) acoustic and physiological recordings of learners and of native speaker French subjects at the Phonetic Laboratory. Our work so far has allowed for the refining of the speaking materials, and will allow for the creation of a database containing acoustic and physiological measurements. Several implications for teaching will be studied later: analyses of errors of voicing and nasalization by the learners; assessment grids before and after specific pronunciation classes; and appropriate pronunciation exercises.

Keywords: Experimental phonetics, French as a Foreign Language, speaking materials, acoustico-physiological recordings, noninvasive instrumentation

1. INTRODUCTION

At present, the oral component of foreign languages is more essential than ever; the foreign learner must master the articulatory and prosodic habits of the target language (L2) if he wishes to exchange successfully in natural situations of communication with native speakers. For the evaluation and the teaching of French *pronunciation* of the tested learners, we used three non-invasive instruments commonly used in experimental phonetics. Much of this section is devoted to the first step in the acquisition of these data: the selection of a set of speaking materials adapted to the capacities and needs of foreign learners, and establishing the nature and conditions of the recording of the learners. On the basis of this study, we can build systematic analysis grids and evaluation of phonetic abilities of learners, and compare these productions with a control population. Following these collections of recordings and analysis of learners' voicing and nasality gaps, a database of foreign learners and native speakers will also be built.

2. OBJECTIVE

This study is the starting-point of a major project to establish a database of control subjects and foreign learners from speaking materials designed to explore their segmental and suprasegmental production in French (L2). Initially, recordings were performed on 10 speakers to verify the adequacy of the speaking material, and the tradeoff between experimental efficiency (sufficiently large number of phrases and phonetic contexts explored) and its possible implementation by learners with a different level of learning L2. This is to avoid (i) fatigue (speaking material too long), (ii) artificial propagation of errors (use of phrases with a lexicon or a syntax too complex or remote expressions of oral daily). Another goal of this study is to validate the recording conditions and equipment used. At a second stage, we hope these data will provide some preliminary comparisons between controls and foreign learners, draw a grid of evaluation, and establish a methodology for processing the data and the organization of the database.

3. DATA ABOUT SUBJECTS

3.1. Selection criteria for the presentation of the subjects

A table for the self-introduction of foreign subjects is useful to all teachers of FFL. The questionnaire at the top of Table 1 is supplemented by questions asked to the learner about the / **his** pronunciation of French. This is intended to provide a better understanding of each topic and ultimately to integrate essential information in the database. We assume that these criteria affect the ranking of the speaker in a "level group", and the results obtained: a person engaged in the study of French for a long time will probably have a better pronunciation than those who only recently began to practise recently, but this factor can be balanced by other elements (type, duration of training in L2 pronunciation, area where they learned L2, negative perception of sounds of L2 : see Table 2).

Table 1: Criteria for the presentation of the French Foreign Language learners.

Coordinates, language, nationality, occupation	- Name - Age - Native language - Address in France - Job in your country - Email - Nationality - Phone in France - Job in France
Living in France	- Since when are you in France? - Have you ever made a living in France or in other countries? - When? How long?
Level of education	- Itemize your studies in your country: degrees, years in which school, speciality ... - Have you received degrees in French? If yes, where?
Training in French and other languages; aware of the difficulties	- List the schools and French classes you took (in your country and France) - Since when do you speak French? - Where did you learn French? ☐ In your country of origin ☐ in France ☐ Elsewhere. Specify: - Have you ever taken French lessons to improve your pronunciation? ☐ yes ☐ no - Do you have difficulties pronouncing the French? If yes, why? - Other language spoken than French or your language? If yes, which? how long?
Subjective opinions on the pronunciation of French	Give **your** frank opinion on the pronunciation of French. - When I started learning French, **for my teachers**, the pronunciation was a matter: not at all important / somewhat important / very important - When I started learning French, **for me**, the pronunciation was a matter: not at all important / somewhat important / very important I found that: - **The music of the French language** is: beautiful / more beautiful / normal / rather ugly / very ugly - **The vowels of French** are very beautiful / more beautiful / normal / rather ugly / very ugly - **The consonants of the French** are very beautiful / more beautiful / normal / rather ugly / very ugly - When I hear **my name** pronounced by a (e) French (e) is: ridiculous / normal / fun - When I hear the French (es) speak, I think it's ... - When a French makes a remark **about my accent**, I think ... - When a French makes a remark **about a grammar mistake**, I think ... - To work on the grammar, I must ... - To work on the pronunciation, I must ... - I have ideas for learning (book, games, songs): never / sometimes / often - I sing / French songs? never / sometimes / often - I speak French **only in my head**? never / sometimes / often - I speak French **only with loud voice**? never / sometimes / often - When I speak French, my mouth is ... - When I speak French, my energy to utter a sound ... - When I speak with the least possible foreign accent to a French, that is: unnecessary / superfluous /a way to be more effective / gift - Today, for me, the French pronunciation is: great suffering / a little uncomfortable / normal / nice / great pleasure

3.2. Example of presentation of foreign learners

This information appears in Table 2: these subjects have a very good understanding of spoken and written French (levels B2 and C1).

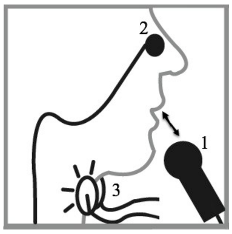
Table 2: Presentation of two of the four foreign learners from the preliminary study.

Subject (sex)	Age	Job	Level of education	Native language	In France Since …	French spoken since …	French courses, type location	French learned …
MS (M)	55	Unemployment; Schoolteacher	Master of General Linguistics	Bambara	1973	1956	School and tutoring	Benin, France
DK (F)	24	Student, Child Care	Master of Philosophy	Polish	2006	2006	Sorbonne University; particular courses	Poland

4. INSTRUMENTATION USED FOR THE RECORDINGS

For each subject (controls and learners), a noninvasive recording with three simultaneous instrumentations has been achieved: microphone recordings, piezoelectric, and electroglottographie (Figure 1). These facilities (Vincent et al., forthcoming) (i) do not interfere with the speakers in the emission of natural speech; (ii) capture the acoustic output at different points in the vocal tract.

Figure 1: Instrumentation used for the recordings.



1. Headset (constant distance between the mouth of the speaker and microphone throughout the recording).
2. Piezoelectric Accelerometer (capture vibrations from the nose).
3. Electroglottographic electrodes (capture laryngeal movements).

For recording microphone (buccal), each subject was fitted with a helmet (*Beyerdynamique*) supporting a hypercardioid (HS 4VXS) microphone (*Schoeps*) so that the distance between the mouth and the microphone remains constant (10 cm) throughout the recordings, a constraint required for future perceptual assessments of learners productions by expert judges in FFL.
The accelerometer is a piezoelectric contact microphone with two metal one-cm diameter sensors (Brand: *K & K Sound*), placed on either side of the nasal side: this non-invasive device records vibrations through the nasal cavity nasal, and allows to record the acoustic vibrations traveling through the nasal wall (Horii 1980: 254). Nasality indices will be determined by measuring the ratio between the amplitude of the acoustic signal of the oral and nasal signal (Figure 2). These results are particularly relevant in the field of FFL (nasal vowel produced as an oral one by a FFL learner, or oral vowel nasalized in contact with a nasal consonant, Racine et al. 2009).
The electroglottograph indirectly records laryngeal behavior by measuring the change in electrical impedance across the throat during phonation (Frokjaer-Jensen and Thorvaldsen 1968:2; Henrich 2001:87). The signal obtained, modulated in function of the contact between the vocal folds, allows optimal detection of the fundamental frequency and phonemes voicing (Figure 2) : in future studies, it will allow us to see differences in voicing (voiceless and voicing consonants).

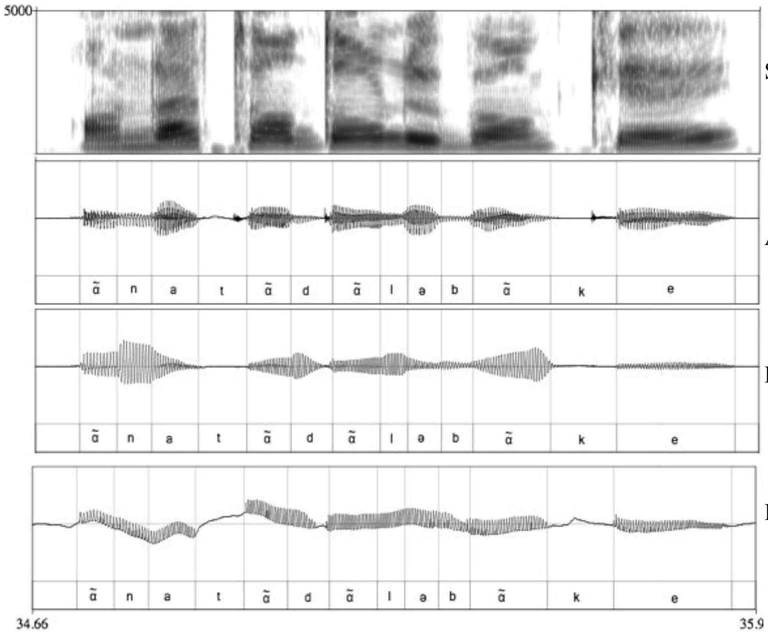
5. CALIBRATION AND RECORDING PROCEEDINGS

After installing the device and the calibration of different instruments (Table 3), 20 minutes of recorded speech per speaker were obtained.

Table 3: calibration at the beginning of the recordings, and goal (total duration of the experiment per subject: one hour).

Type of calibration : production of…	Goal
First sentences of the speaking material	Calibrate the overall level of registration to avoid saturation of the various signals
a) two sustained [a] and [m] b) "Il a dit tatatat trois fois. Il a dit titìtit trois fois. Il a dit mamamam trois fois. Il a dit mimimim trois fois." ("He said tatatat/tititit/mamamam/mimimim three times. ")	The presence of occlusive and great intensity of [a] avoid saturation of the microphone signal ; the production of [m] avoid saturation of the piezoelectric signal.

Figure 2: Plots obtained for the French phrase "... en attendant le banquet ..." spoken by a male control subject. From top to bottom: spectrogram; acoustic signal; signal from the piezoelectric accelerometer; and electroglottographic signal.



6. SPEAKING MATERIALS: COMPOSITION AND ADVANTAGES

This speaking material has been made in two stages. A first material is an adaptation of a phonetic database produced under a contract with the Francophone Agency for Higher Education and Research (Vaissière, 1998: 3). Six control subjects (five women and one man) and four foreign students (two women and two men) were recorded (mean age of controls: 36.2 years, SD: 7.1; average of foreign learners: 36.2 years, SD: 14). The first students have however shown a significant fatigue probably due to a speaking material inaccessible to them (such as: sentences containing words whose meaning was unknown to them).

We have therefore undertaken to rehabilitate a second speaking material comprising four tasks, in accordance with techniques normally used in teaching (Charliac et al. 2003, 2004, 2006; Fredet and Patel 2009: 11; Lauret 2007: 37; Leon 2003: 5 ; Mabilat and Martins 2004: 4; Martinie and Wachs 2007: 7): 1) repetition, 2) prepared and unprepared reading, 3) spontaneous speech (narration, dialog), 4) reproduction of extracts stored. In most of the second protocol, we avoid the use of written material to avoid influence from the spelling. The perception and speech production may undergo various treatments implying different brain areas and activities by these four tasks: from an auditory processing to a linguistic processing and orthoepic one, these various treatments affect the performance of phonetic speaker (Lauret 2007: 26).

In order that this speaking material is accessible to the learners, it prevents their fatigue and does not bias the formation of these data, we eliminated the multiple repetitions of each sentence that are common in the acquisition of such data. The statistical treatment may be satisfactory nonetheless when a large number of speakers will be recorded.

This speaking materials consists of the elements described in Table 3: first, vowels in various contexts (repetition) allows us to study all the French vowels according to the initial consonant environment while keeping the same intonation pattern. We kept the open syllables for vowel easily pronounceable in this context, and closed syllables for those who usually pronounce it.

Second, the achievement of French consonants in intervocalic position with [a] and [i] allows the speaker to focus on the production of consonants because these vowels are common to many languages (Maddieson 1984: 56).

Thirdly, the repetition at the beginning of this recording, then the unprepared reading at the end of it concerns the production of French in the natural context for an internationally known text, *Le Petit Prince*, in two contexts of narrative: dialogue and prose. It also allows the analysis of the prosody of the speaker.

Fourthly, the context sentence "He said" <vowel> "slowly", translated into the native language of the learner is then recorded. "<vowel>" represents each vowel in this native language. It is thus possible to know if the learner's mistake is influenced by the vowel system of their native language.

Fifthly, the reading of ambiguous sentences in any order assessing prosody demarcation of their meaning (rising intonation between the groups in a sentence, and lengthening of the last syllable of each group). Between them are introduced other phrases acting as distractors (eg, "I do not think"). Sixth, through the repetition and the production of a sentence on several intonations, the various modalities of the same intonational phrase in French are analyzed.

Seventh, spontaneous speech (two minutes) is produced with two methods: 1. Telling about the activity of the weekend 2. Learner asks questions to the experimenter. Finally, Prévert's poem "*Pour toi mon Amour*" memorized in advance by the learner, is chosen for its simplicity of language and mnemonic convenience (due to its repetitive nature).

Table 4: Speaking materials: left column: items; right column : description of the French speaking materials

<table>
<tr><td>Vowels in various contexts (in repetition)</td><td>1. Il a dit "V_f" de pV_fpV_f, de tV_ftV_f, de kV_fkV_f, de bV_fbV_f, de dV_fdV_f, de gV_fgV_f, de fV_ffV_f, de sV_fsV_f, de chV_fchV_f, de vV_fvV_f, de zV_fzV_f, de jV_fjV_f, de lV_flV_f, de mV_fmV_f, de nV_fnV_f, de rV_frV_f. (Translation : He said "V_f" of ...)
V_f : vowel or "a" or nasal vowels, pronounced in an open syllable.
V_f = "i", "é", "u", "eu", "ou", "o", "a", "in", "an", "on"
2. Il a dit "V_o" de pV_orpV_or, de tV_ortV_or, de kV_orkV_or, de bV_orbV_or, de dV_ordV_or, de gV_orgV_or, de fV_orfV_or, de sV_orsV_or de chV_orchV_or, de vV_orvV_or, de zV_orzV_or, de jV_orjV_or, de lV_olV_o, de mV_omV_o, de nV_onV_o, de rV_orV_or.
V_o: open vowel pronounced in one closed syllable; V_o = "oe", open "o".</td></tr>
<tr><td>Intervocalic consonants ([a] and [i], repetition)</td><td>C'est aCa qui l'a dit; c'est iCi qui l'a dit
Translation : "This is aCa who said it" ; C : consonant in the following list:
C = "p", "t", "k", "b", "d", "g", "f", "s", "ch", "v", "z", "j", "l", "m", "n", "r", "gn".</td></tr>
<tr><td>French text: Le petit Prince:

Two extracts

- Repetition (at the beginning of the recording)
- Not prepared reading, at the end of it</td><td>C'est alors qu'apparut le renard : - Bonjour, dit le renard.
- Bonjour, répondit poliment le petit prince, qui se retourna mais ne vit rien.
- Je suis là, dit la voix, sous le pommier…
- Qui es-tu ? dit le petit prince. Tu es bien joli…
- Je suis un renard, dit le renard.
- Viens jouer avec moi, lui proposa le petit prince. Je suis tellement triste…
- Je ne puis pas jouer avec toi, dit le renard. Je ne suis pas apprivoisé.
- Ah pardon, fit le petit prince. Mais après réflexion, il ajouta :
- Qu'est-ce que signifie « apprivoiser » ?
- Tu n'es pas d'ici, dit le renard, que cherches-tu ? »
- Je cherche les hommes, dit le petit prince. Qu'est-ce que signifie « apprivoiser » ?
- Les hommes, dit le renard, ils ont des fusils et ils chassent. C'est bien gênant ! Ils élèvent aussi des poules. C'est leur seul intérêt. Tu cherches aussi des poules ?
- Non, dit le petit prince. Je cherche des amis. Qu'est-ce que signifie « apprivoiser » ?
<hr>On ne connaît que les choses que l'on apprivoise, dit le renard. Les hommes n'ont plus le temps de rien connaître. Ils achètent des choses toutes faites chez les marchands. Mais comme il n'existe point de marchands d'amis, les hommes n'ont plus d'amis. Si tu veux un ami, apprivoise-moi! Adieu, dit le renard. Voici mon secret. Il est très simple: on ne voit bien qu'avec le cœur. L'essentiel est invisible pour les yeux.</td></tr>
<tr><td>Sentence context</td><td>« Il a dit „<voyelle>" lentement », pronounced in the native language, production.
Translation : He said slowly „<vowel>"</td></tr>
<tr><td>Ambiguous sentences (unprepared reading)</td><td><table><tr><td>Jean lève son verre.
Tu n'imagines pas !
Les gares sont dessinées.
Tu parais très soucieux.
J'enlève son verre.</td><td>Jean porte un journal.
<u>La belle ferme</u> le voile.
Le rapace la noie.
C'est bien d'accord.
Une tasse, deux chocolats.</td><td>J'emporte un journal.
Tu paraîtrais soucieux.
Le rat passe la noix.
Les garçons dessinaient.</td><td>Une tasse de chocolat.
Je viendrai demain.
<u>La belle</u> ferme le voile.
Je ne pense pas.</td></tr></table></td></tr>
<tr><td>Sentence</td><td>Intonation, repetition, production: « Il fait beau » (It's nice) : question, statement, surprise, sadness, joy, anger, fatigue</td></tr>
<tr><td>Conversation</td><td>(spontaneous speech)</td></tr>
<tr><td>Part of text stored (Pour to mon Amour, Prévert)</td><td><table><tr><td>Je suis allé au marché aux oiseaux
Et j'ai acheté des oiseaux
Pour toi, mon Amour
Je suis allé au marché aux fleurs</td><td>Et j'ai acheté des fleurs
Pour toi, mon Amour
Je suis allé au marché à la ferraille
Et j'ai acheté des chaînes, de lourdes chaines</td><td>Pour toi, mon Amour
Et je suis allé au marché aux esclaves
Et je t'ai cherchée
Mais je ne t'ai pas trouvée, mon Amour</td></tr></table></td></tr>
</table>

7. CONSTITUTION OF THE DATABASE

The data has been integrated into a module adapted from the platform "PCF" ("Phonology of Contemporary French" 2010) version 2 (Durand et al. 2002: 94). This tool helps to organize the speaking materials, the annotation (metadata), perform queries, etc.. It also offers advanced integration with the signal analysis software Praat (Boersma and Weenink 2009) and allows back-and-forth between the data of the speakers and those on the signal.

8. CONCLUSION

This reflection on the selection of speaking materials that can conveniently be recorded by foreign learners is a prerequisite for a much larger project: that of constructing a large database of productions by learners of FFL. From the measurements made through non-invasive instruments, our analysis will allow us: (i) to present, in future, studies of their results (eg differences in voicing of certain consonants, oral vowels nasalized in contact with a nasal consonant); (ii) to develop relevant teaching methods to help students to improve their French pronunciation.

REFERENCES

Boersma, P., Weenink, D. 2010. Praat: doing phonetics by computer (V. 5.1.05). <www.praat.org>. Last access 01 Mar 2010.

Charliac, L., Morton A.C., Le Bougnec J.T., & Loreil, B. 2003. *Phonétique progressive du français : niveau : 1) débutant :2003 ; intermédiaire : 2004 ; avancé : 2006*. Paris : CLE International.

Phonologie du Français Contemporain. 2009. PFC Homepage. <http://www.projet-pfc.net>. Last access 01 Mar 2010.

Durand, J., Laks, B., Lyche, C. 2002. La phonologie du français contemporain: usages, variétés et structure. In: Pusch, C., Raible W. (eds.) *Romanistische Korpuslinguistik-Korpora und gesprochene Sprache/Romance Corpus Linguistics - Corpora and Spoken Language*. Tübingen: Gunter Narr Verlag. 93-106.

Fredet, F., Patéli, M. 2009. *Prononciation du français.* Athènes : Anoikto Vivlio, Ed.

Frokjaer-Jensen, B., Thorvaldsen, P. 1968. Construction of a Fabre Glottograph. ARIPUC. Copenhagen University. 3. 1-8.

Henrich, N. 2001. Etude de la source glottique en voix parlée et chantée: modélisation et estimation, mesures acoustiques et électroglottographiques, perception. PhD Dissertation, Paris 6 University. Also at <http://tel.archives-ouvertes.fr/tel-00123133>. Last access 01 Mar 2010.

Horii, Y. 1980. An Accelerometric approach to nasality measurement: A preliminary report. *Cleft Palate Journal 17*. 254-261.

Lauret, B. 2007. *Enseigner la prononciation du français : questions et outils*. Paris : Hachette.

Leon, M. 2003. *Exercices systématiques de prononciation française*. Paris: Hachette.

Maddieson, I. 1984. *Patterns of Sounds*. Cambridge: Cambridge University Press.

Mabilat, J.-J., Martins, C. 2004. *Sons et intonation*. Paris : Didier –FLE.

Martinie, B., Wachs, S. 2007. *Phonétique en dialogue*s, *niveau débutant.* Paris : CLE international.

Racine, I., Detey, S., Zay, F. & Kawaguchi, Y. 2009. Le projet « Interphonologie du français contemporain » : réflexions méthodologiques et premières données d'apprenants hispanophones et japonophones. *Colloque AFLS Langue française en contextes*, Université de Neuchâtel, 3-5 septembre 2009.

Vaissière, J. 1998. *Corpus élaboré dans le cadre d'un contrat sur la comparaison des langues entre l'Institut de Linguistique et de Phonétique Générales et Appliquées (ILPGA) et l'Agence Francophone pour l'Enseignement Supérieur et la Recherche (AUPELF – UREF).*

Vincent, C., Amelot, A., Crevier-Buchman, L., Fougeron, C., Honda, K., Lo Bue, G., Maeda, S., & Vaissière, J. A paraître. PEP2, la plateforme d'étude physiologique de la parole du Laboratoire de Phonétique et Phonologie : Présentation et retour d'expérience. *5ème colloque interdisciplinaire en instrumentation* - C2i 2010, Hermès, Paris, in press.

Investigating native and non-native vowels produced in conversational speech

Thorsten Piske[1], James E. Flege[2], Ian R. A. MacKay[3], Diane Meador[4]

[1]University of Education Schwäbisch Gmünd, Germany; [2]University of Alabama at Birmingham, USA; [3]University of Ottawa, Canada; [4]University of Alaska Southeast, USA
thorsten.piske@ph-gmuend.de, flegejames@yahoo.com, imackay@uOttawa.ca, Diane.Meador@uas.alaska.edu

ABSTRACT

The primary aim of this study was to determine if errors observed when fluent early bilinguals produced L2 vowels in non-words would also occur when the same bilinguals produced vowels in conversational speech. The subjects were native speakers of Italian who began learning English when they immigrated to Canada as children or adults ("early" vs. "late" bilinguals). The early bilinguals were subdivided into groups differing in amount of continued L1 use ("Early-low" vs. "Early-high"). Vowel production accuracy was assessed auditorily by native English-speaking listeners. The listeners knew beforehand the identity of the target vowels being evaluated. The experiment revealed that the early bilinguals' vowel errors in non-words were not apparent in conversational speech. The findings reported here support the view that L2 production data elicited in experiments involving the use of written materials may not always reveal how accurately bilinguals can produce the sounds of an L2 in conversational speech.

Keywords: L2 vowel production, conversational speech, spelling pronunciations

1. INTRODUCTION

The research reported here was motivated by the results of a study by Piske et al. (2002), who examined the production of 11 English vowels (i.e., /i ɪ eɪ ɛ æ u o ʌ ɒ ʊ ɚ/) by 54 native Italian subjects differing in terms of age of arrival to Canada (AOA) and/or in terms of their self-reported percentage use of Italian. Piske et al. (2002) elicited vowel production in two different ways: The subjects first repeated a series of four real English words that were modelled aurally and which also appeared on a written list. When the same four words were presented a second time, the subjects inserted the vowel common to all four (e.g, /i/ in *read*, *deed*, *heed*, *bead*) into a /b_do/ frame, creating a non-word (e.g., /bido/). 11 native speakers of English were recruited to auditorily evaluate the 1584 vowel stimuli that had been produced by three groups of Italian-English bilinguals and one group of native English (NE) speakers. Piske et al. (2002) found that whereas none of the vowels produced by a group of early bilinguals who seldom used their L1 Italian (i.e., the "Early-low" group) received significantly lower ratings than vowels produced by the NE subjects, some of the vowels produced by a group of early bilinguals who used their L1 frequently (i.e., the "Early-high" group) did receive lower ratings than vowels spoken by the NE subjects. Interestingly, most of the observed differences between the NE and Early-high groups were for vowels spoken in the non-word condition. In a second experiment included in the Piske et al. (2002) study, native English-speaking listeners used keywords in order to classify the same vowels that had been rated in the first experiment. The results of the classification experiment suggested that many of the errors produced by the Early-high group were due, at least in part, to the influence of orthography. The vowels that were produced less accurately in non-words than in words by the Early-high group included the two vowels /ɪ/ and /ʊ/, which – as far as we know – do not have a phonetic counterpart in any variety or dialect of Italian. Table 1 shows how the NE and Early-high groups' productions of these two vowels in words and non-words were classified in the classification experiment reported by Piske et al. (2002).

Table 1: Native English listeners' classifications of 2 English vowels spoken by the subjects in two groups (adapted from Piske et al. 2002, p. 64).

Group	Vowel	Words	Nonwords
NE	/ɪ/	ɪ(95) ɛ(5)	ɪ(95) ɛ(4)
Early-high	/ɪ/	ɪ(97) ɛ(3)	ɪ(35) i(57)
NE	/ʊ/	ʊ(97) ʌ(3)	ʊ(74) u(15) ʌ(7) o(3)
Early-high	/ʊ/	ʊ(94) ʌ(5)	o(34) ʊ(31) u(30) ɒ(3)

Only percentages greater than 2% are shown.

As summarized in Table 1, the /ɪ/s produced by the Early-high subjects in words were never classified as /i/, but the /ɪ/s they produced in non-words were classified as /i/ in 57% of instances. It appears that they pronounced the letter "i" in *rid*, *did*, *hid* and *bid* as it is pronounced in Italian (i.e., /i/) when asked to insert the vowel common to all four words into a /b_do/ frame. The errors involved in the production of /ʊ/ in non-words also appeared to be due, at least in part, to the influence of orthography. All four words with /ʊ/ were spelled with at least one "o" (*good*, *could*, *would*, *hood*), which is often pronounced /o/ in Italian. The Early-high group's productions of /ʊ/ in words were never classified as /o/, whereas the /ʊ/s they produced in non-words were heard as /o/ in 34% of instances.

The purpose of the experiment reported in this paper was to examine whether the spelling pronunciations observed for some vowels in the non-word condition would also occur in conversational speech samples. We hypothesized that errors observed in the Early-high subjects' productions of English vowels in non-words were an artifact of the elicitation procedure and that they did not typify how experienced Italian-English bilinguals produce English vowels when asked to answer questions in a relatively spontaneous way. If many of the Early-high subjects' errors for vowels in non-words were, in fact, an artifact of the elicitation procedure, two things should be true: First, the types of errors observed for the Early-high subjects' non-word productions of /ɪ/ and /ʊ/ should not occur in conversational speech samples. Second, their productions of other vowels in conversational speech should not differ from the NE subjects' productions of these vowels in conversational speech samples.

Most researchers would acknowledge that conversational speech *should* represent the most important criterion for success in acquiring L2 vowels, but surprisingly few studies have been undertaken (e.g., Wode 1981; Piske et al. 1999; Tsukada 2001, Bent et al. 2007). The most likely reason for this gap is the inherent difficulty in analyzing conversational speech under controlled conditions. A technique with satisfactory experimental control was used in this study to analyze conversational speech samples.

2. METHOD

2.1. Subjects

The same sample of subjects as those who participated in the Piske et al. (2002) study was used in the study presented here. The mean age of the 72 subjects who participated in the study was 48 years. Three groups of Italian-English bilinguals and a group of NE speakers (18 per group) were recruited in Ottawa, Ontario.[1] The native Italian subjects had been living in Canada for a minimum of 18 years at the time of testing (mean = 35 years). The subjects in a group referred to as the "Late-high" group had arrived in Canada later in life (mean = 19 years) than two groups of early bilinguals had (mean = 7 years for both). The early bilinguals in a group referred to as the "Early-low" group reported using Italian much less often (mean = 8%) than the early bilinguals in an "Early-high" group did (mean = 32%). An ANOVA examining the native Italian subjects' self-estimates of percentage L1 use was significant $F(2,51) = 18.4$, $p < 0.01$. A Tukey's test showed that, as intended by the design, the Early-high and Late-high groups used Italian more frequently than the Early-low group (p < 0.01), whereas the Early-high and Late-high groups did not differ significantly (p > 0.05).

2.2. Speech Samples

The following procedure was used to elicit conversational speech samples. The subjects heard a dialog between a husband and wife speaking English and Italian, respectively. The dialog was followed by alternating questions in English and Italian (four each). The NE and Italian subjects responded in English to the questions posed in English. The native Italian subjects also responded, in Italian, to the questions posed in Italian. The questions and the preceding dialog focused on issues like bilingualism and immigration. The questions were not designed to elicit the specific vowels of interest.

The subjects' responses to the English questions were recorded using a DAT tape recorder (Sony TCD-D8), and later digitized at 11.025 kHz. The digitized recordings were orthographically transcribed by an author who is a native speaker of English (i.e., the fourth author of the present paper). The transcriptions were then used to identify tokens of /ɪ/ and /ʊ/ as well as of four additional vowels (i.e., /ɚ i æ o/) produced by the subjects in words consisting of 1 to 3 syllables. An attempt was made to only select tokens occurring in content words. However, due to an insufficient number of content words with /ʊ/, tokens of /ʊ/ occurring in the modals *should*, *could* and *would* were also selected. All of the vowels included for analysis occurred in lexically stressed syllables.

Phonetic segments that are heard as intended in their original context may be misidentified when excerpted (e.g., Picket & Pollack 1963). We therefore preserved a short stretch of speech before and after each selected vowel token before storing it in a separate file. The preceding and following contexts both contained at least one syllable unless the syllable containing a target vowel was immediately preceded or followed by a pause. We avoided excerpting just parts of words. The contexts therefore included more than one syllable when this was necessary to preserve whole words. For example, in the excerpt "very d*i*fficult", the target vowel /ɪ/ was preceded by two syllables (the word *very*) and followed by two syllables (which completed the word *difficult*). An attempt was made to find three tokens per subject of all six vowels selected for analysis (/ɪ ʊ ɚ i æ o/). The average number of tokens available per vowel averaged 2.8 for the NE subjects, 2.4 for the Early-low subjects, 2.5 for the Early-high subjects, and 2.3 for the Late-high subjects. However, in some cases, no tokens of a particular vowel could be identified for certain subjects. This happened for 1 subject for the vowel /i/, 2 for /ɚ/, 4 for /o/, 7 for /æ/, and 18 for /ʊ/.

2.3. Auditory evaluation

Native speakers of Canadian English (2 males, 4 females living in Ottawa, Ontario), having a mean age of 34 years, were asked to phonetically transcribe the target vowels occurring in the short excerpts that had been edited out of the conversational speech samples. Although all of the listeners had previously received training in the use of phonetic symbols, they were nevertheless required to demonstrate an ability to reliably transcribe the Canadian vowels /i ɪ ɛ æ u o ʌ ɒ ʊ ɚ/ and the Italian vowels /a e ɔ/ prior to participating in the experiment reported here.

During the experiment, tokens of each target vowel were randomly presented over headphones (Koss TD 65) in separate counterbalanced blocks (2 in each of 3 sessions). Blocking on vowel was intended to ensure that the listeners knew the intended identity of the target vowels (which were often, but not always obvious from the surrounding context). The listeners were told that they would hear short excerpts of English speech that had been produced by native speakers of English and Italian. They were told the identity of the target vowel to be evaluated in each block before it began. Each block began with five trials that were included for practice and were not analyzed.

An orthographic representation of each excerpt appeared on the screen as it was presented auditorily. The location of each target vowel was marked by an asterisk in the orthography (e.g., "very d*fficult"). The listeners were told to focus their attention on the target vowel, and to identify it using one of the 15 response alternatives that appeared on the computer screen. For example, the responses offered for the target vowel /ɪ/ were: /i/, /e/, /ɛ/, /æ/, /a/, /u/, /ʊ/, /o/, /ɔ/, /ɒ/, /ʌ/, /ɚ/, "good /ɪ/", "slightly distorted /ɪ/", and "very distorted

/ɪ/". This array of response alternatives permitted the listeners to identify the target vowel as being an instance of some other category or, when a vowel token was heard as intended, to rate it for goodness.

3. RESULTS

In order to determine whether the types of errors observed when subjects in the Early-high group produced vowels such as /ɪ/ and /ʊ/ in non-words would also occur when the Italian-English produced conversational speech, the percentage of times that each subject's production of the target vowels was heard as intended was calculated.[2] The means for each subject's productions of /ɪ ʊ ɚ i æ o/ were based on a maximum of 18 judgments (6 listeners x 3 tokens). The mean scores obtained for the four groups are shown in Table 2. The vowels spoken by subjects in the NE, Early-low, and Early-high groups were identified correctly more often (98.4-99.1%) than were the vowels spoken by the Late-high group (mean = 84.6%).

Table 2: The mean percentage of times that English vowels spoken in conversational speech by subjects in four groups were heard as intended, and ANOVAs testing for between-group differences.

	1 Native English	2 Early-low	3 Early-high	4 Late-high	*ANOVA*	*a posteriori* t-tests
/ɪ/	100.0 (0.0)	99.7 (1.3)	99.7 (1.3)	77.3 (26.0)	F(3,68)= 22.0*	4 < 1, 2, 3
/ʊ/	100.0 (0.0)	99.1 (2.2)	100.0 (0.0)	90.1 (17.3)	F(3,50)= 5.2*	4 < 1, 2, 3
/ɚ/	99.4 (1.8)	99.7 (1.4)	99.4 (1.8)	85.8 (14.2)	F(3,65)= 18.5*	4 < 1, 2, 3
/i/	99.1 (2.9)	99.0 (2.2)	99.1 (2.1)	94.8 (5.5)	F(3,67)= 6.1*	4 < 1, 2, 3
/æ/	96.6 (4.3)	94.4 (8.1)	94.8 (6.9)	70.8 (18.1)	F(3,61)= 16.8*	4 < 1, 2, 3
/o/	99.4 (1.8)	98.6 (4.3)	99.0 (4.0)	89.7 (12.6)	F(3,64)= 10.0*	4 < 1, 2, 3
M	99.1 (1.0)	98.4 (2.1)	98.7 (1.2)	84.6 (9.2)		

An asterisk indicates significance at the 0.05 level. The Bonferroni-corrected level used for the *a posteriori* tests was 0.05.

The standard deviations associated with the Late-high groups' scores were larger than those for the other three groups, so the scores were submitted to an arcsine transformation (Kirk 1968) before being examined in a series of one-way ANOVAs. As summarized in Table 2, the effect of group was significant for all six target vowels ($p < 0.05$). A series of five *a posteriori* t-tests were carried out to test for between-group differences for each vowel (NE vs. the three native Italian groups; Early-low vs. Early-high; Early-high vs. Late-high). Significantly lower scores were obtained for all six vowels produced by the Late-high than NE group (Bonferroni $p < 0.05$). No other between-group difference reached significance, however. As regards the intelligibility of the vowels examined here, these results confirmed the prediction that, for vowels produced in conversational speech, the Early-high subjects' productions of neither /ɪ ʊ/ nor /ɚ i æ o/ would differ from the NE subjects' productions of these vowels.

Table 3: The mean percentage of classifications, by native English-speaking listeners, of six vowels that were spoken in conversational speech by the subjects in four groups.

Group	Target Vowel	Classified as	Target Vowel	Classified as
Native English	/ɪ/	ɪ(100)	/i/	i(99.1)
Early-low	/ɪ/	ɪ(99.7)	/i/	i(99.0)
Early-high	/ɪ/	ɪ(99.7)	/i/	i(99.1)
Late-high	/ɪ/	ɪ(79.2) i(19.5)	/i/	i(94.2) ɪ(4.8)
Native English	/ʊ/	ʊ(100)	/æ/	æ(96.6)
Early-low	/ʊ/	ʊ(98.6)	/æ/	æ(95.0) ɛ(3.1)
Early-high	/ʊ/	ʊ(100)	/æ/	æ(94.4) a(3.0) ɛ(2.6)
Late-high	/ʊ/	ʊ(88.9) u(11.1)	/æ/	æ(71.4) ɒ(11.4) a(11.4) ɛ(4.3)
Native English	/ɚ/	ɚ(99.4)	/o/	o(99.4)
Early-low	/ɚ/	ɚ(99.6)	/o/	o(99.2)
Early-high	/ɚ/	ɚ(99.4)	/o/	o(99.6)
Late-high	/ɚ/	ɚ(86.6) ʊ(6.3) ɒ(3.1)	/o/	o(90.0) ɔ(7.0) ʌ(2.1)

Percentages lower than 2.0% are not reported.

Table 3 summarizes the phonetic transcriptions obtained for the vowels spoken in conversational speech. The percentages shown here were calculated by dividing the number of times each phonetic symbol was used to identify a group's production of a particular vowel by the total number of available tokens. When the native Italian subjects' vowels were heard as instances of the intended category, they were usually heard as the neighboring vowel. This was expected from the results obtained by Piske et al. (2002) and in other studies (e.g., Flege et al. 1999). So, for example, intended /ʊ/s were sometimes heard as /u/, /ɚ/s as /ʊ/, /æ/s as /ɒ/ (or Italian /a/), and /o/s as Italian /ɔ/. The target vowel /ɪ/ was sometimes heard as /i/ and the target vowel /i/ was sometimes heard as /ɪ/.

In their classification experiment, Piske et al. (2002) had found that the Early-high subjects' productions of /ɪ/ in non-words were classified as /i/ in 57% of instances (see Table 1). The /ɪ/s produced in words by the same subjects were never classified as /i/, however. The /i/-for-/ɪ/ substitutions in non-words were interpreted as resulting from orthography. As shown in Table 3, neither group of early bilinguals produced /ɪ/s that were heard as /i/ in conversational speech. This confirms the hypothesis that the early bilinguals' /i/-for-/ɪ/ substitutions in non-words did not typify their production of /ɪ/. Similarly, the Early-high groups' productions of /ʊ/ in words were never classified as /o/ in the classification experiment conducted by Piske et al. (2002), whereas their /ʊ/s were classified as /o/ in non-words in 34% of instances. This was also interpreted to be a spelling pronunciation. In support of this, the Early-high groups' productions of /ʊ/ in conversational speech were never classified as /o/.

4. DISCUSSION

Taken together, the results of the Piske et al. (2002) study and the results reported here support the view that in certain elicitation conditions fluent early bilinguals may not produce all L2 vowels in a manner that is indistinguishable from L2 monolinguals. However, the errors produced by early bilinguals in specific elicitation conditions may not be apparent when they produce L2 vowels in real words or in conversational speech. In the Piske et al. (2002) study, none of the vowels produced by a group of early bilinguals who only seldom used their L1 Italian were found to differ from the vowels produced by a group of NE subjects. Some of the vowels produced by a group of early bilinguals who continued to use their L1 Italian frequently were, on the other hand, found to differ from the NE subjects' vowels. Most of the observed differences between the NE and the Early-high groups were for vowels spoken in a non-word condition. The results of the classification experiment conducted by Piske et al. (2002) suggested that some of these errors were due to the influence of orthography. The results of the study reported in this paper confirmed the prediction that the errors observed in the Early-high subjects' productions of English vowels in non-words were an artifact of the elicitation procedure and that they did not typify how experienced Italian-English bilinguals produce vowels in conversational speech, because the vowel errors that had been produced by the subjects in the Early-high group in non-words were not apparent in the conversational speech samples analyzed here.

The Late-high and Early-high groups examined in this study and in the Piske et al. (2002) study were roughly matched for amount of continued L1 use. As expected from previous work (e.g., Flege et al. 1999), the data presented by Piske et al. (2002) revealed an effect of AOA on Italian-English bilinguals' accuracy in producing English vowels. Seven of the 11 English vowels elicited in word and non-word conditions by Piske et al. (2002) were produced less accurately by the Late-high than by the Early-high group. In the study reported here, vowels produced by the Late-high group in conversational speech were identified correctly less often (mean = 84.6%) than were the vowels produced by the Early-high group (mean = 98.7%). However, in no case did the differences between the Early-high and the Late-high groups reach significance.

Although the subjects in the Early-high group examined here and in the Piske et al. (2002) study had already been living in a predominantly English-speaking environment for many years when they were tested, some or their errors in the production of vowels such as /ɪ/ and /ʊ/ in non-words still reflected the influence of orthographic input. The observation that these errors did not occur when the same subjects produced conversational speech supports the view (see, e.g., Piske et al. 2001) that it is generally problematic to include written materials in experiments designed to examine how accurately bilinguals can produce an L2 (recall that the subjects examined by Piske et al. (2002) produced non-words after repeating four real English words that were modeled aurally and which also appeared on a written list). Additional research is needed, however, to determine why the grapheme-phoneme correspondence rules of Italian appeared to influence the Early-high group's production of English vowels only in non-words, but not in words and under what specific circumstances the grapheme-phoneme correspondence rules of a bilingual's L1 continue to exert a long-term influence on her/his production of an L2.

5. ACKNOWLEDGMENTS

This research was supported by grant no. DC00257 from the National Institute for Deafness and Other Communication Disorders.

6. REFERENCES

Bent, T., Hargus Ferguson, S., Kewley-Port, D. 2007. Listener influences on vowel intelligibility. *Journal of the Acoustical Society of America* 122(5). 3016(A).

Flege, J. E., MacKay, I. R. A., Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America* 106(5). 2973-2987.

Kirk, R. 1968. *Experimental design: Procedures for the behavioral sciences*. Belmont: Brooks/Cole Publishing.

Picket, J. M., Pollack, I. 1963. Intelligibility of excerpts from fluent speech: Effects of rate of utterance and duration of excerpt. *Language and Speech* 6: 151-164.

Piske, T., Flege, J. E., Mackay, I. R. A., Meador, D. 1999. Non-natives' production of vowels in conversational speech. *Collected papers from the 137th Meeting of the Acoustical Society of America and the 2nd Convention of the European Acoustics Association: Forum Acusticum*. Paper no. 2ASCB_4.

Piske, T., MacKay, I. R. A., Flege, J. E. 2001. Factors affecting degree of foreign accent in an L2. A review. *Journal of Phonetics* 29. 191-215.

Piske, T., Flege J. E., MacKay, I. R. A., Meador, D. 2002. The production of English vowels by fluent early and late Italian-English bilinguals. *Phonetica* 59. 49–71.

Tsukada, K. 2001. Native vs. non-native production of English vowels in spontaneous speech. An acoustic study. *EUROSPEECH 2001 Scandinavia, 7th European conference on speech communication and technology, 2nd INTERSPEECH event*, Aalborg. 305-308.

Wode, H. 1981. *Learning a second language: An integrated view of language acquisition*. Tübingen: Narr.

NOTES

[1] Vowels spoken by 18 native speakers of Italian with a mean AOA of 14 years were also elicited and evaluated by the listeners. In the interest of economy, the results of this "Mid" group will not be reported here.

[2] This means that only the identification data but not the goodness ratings obtained from the native Canadian listeners were used in the analyses presented here.

First and second language phonological processes in clinical populations

Monika Połczyńska, Marta Marecka

Adam Mickiewicz University
plmonik@ifa.amu.edu.pl

ABSTRACT

In clinical populations phonological processes (speech errors) may be caused by problems with speech production or speech perception. The aim of this paper is to present L1 (Polish) and L2 (English) phonological processes found in two clinical groups: (1) dysarthric individuals with traumatic brain injury (TBI) and (2) individuals with sensorineural hearing loss (SNHL). The subjects represented different levels of L2 proficiency, however there were no highly proficient users with native-like fluency. The results indicate that each group used more types of processes in L2 than in L1. The TBI group applied a higher number of more radical processes in L2 (e.g. consonant deletions) than in L1. On the other hand, the two most common processes in the SNHL group – sibilant imprecision and vowel centralisation that constituted 56% of all the types of processes applied – were used with nearly the same frequency of occurrence in L1 and L2 (SD=0,5). Thus, it might be argued that impaired perception of speech sounds results in a more consistent application of phonological processes in L1 and L2 and distorted speech production leads to the application of more radical processes in L2.

Keywords: phonological processes, traumatic brain injury, sensorineural hearing loss, dysarhtria, second language

1. INTRODUCTION

Typically developing children apply natural phonological processes in first language acquisition (e.g. Stampe 1972; Grunwel 1985; Dressler et al. 1987; Barlow 2001). Phonological processes are speech errors that are applied to phonemes which cause articulatory difficulties. Numerous studies report that phonological processes also occur in disordered speech of children (e.g. Ingram 1990; Wyllie-Smith et al. 2006) as well as adults (e.g. Duffy 1995; Połczyńska-Fiszer and Pufal, 2006). According to Natural Phonology (NP) (e.g. Dressler et al. 1987; Dziubalska-Kołaczyk 2002) and Phonology as Human Behaviour (PHB) (e.g. Diver 1979; Tobin 1997) processes are classified into three major types: (1) substitutions (e.g. spirantisation, stopping, vowel centralisation), (2) assimilations (e.g. devocing, vowel spirantisation), and (3) changes in syllable structure (e.g. consonant deletion, unstressed vowel deletion). Based on her research on post-traumatic dysarthric speech, Połczyńska (2006; 2009) added two other types of phonological processes: (4) underarticulations and (5) changes in articulatory force. Underarticulations involve an articulatory undershoot of consonant phonemes. This type of processes is further subdivided into (a) Incomplete Consonant Closure (ICC) and Consonant Approximation (CA). In ICC the intended consonant is articulated imprecisely, yet it can still be discriminated by the listener. The processes is phonetic (rather than phonemic) in nature as it reflects insufficient muscle strength required to complete the articulatory gesture, e.g. the phoneme /k/ in a Polish word *koń* /koɲ/ "horse") is produced as /koɲ̆/. CA, on the other hand, involves only a slight approximation of articulators that appears in positions in whic̑h a consonant is expected. This process is common if individuals with severe dysarthtria and it is caused by inability to make more extensive articulatory gestures, e.g. *koła* /kowa/ ("wheels") produced as /ko̯˘a/. CA can only be observed with acoustic instrumentation. The fifth type of phonological processes, articulatory force changes, are divided into weak (whisper) articulation and strong or hyper-articulation.

Abundant research shows that phonological processes occur as the result of both disordered articulation and impaired hearing (e.g. Polczyńska-Fiszer 2006; Marecka and Połczyńska 2009). More specifically, in this paper we will analyse the speech of dysarthric individuals after Traumatic Brain Injury (TBI) and individuals with Sensorineural Hearing Loss (SNHL).

Dysarthria is a neuromotor speech disorder (Duffy 1995) that comprises about one third of all communication impairments in the TBI population (Sarno et al. 1986). TBI may vary extensively as to the extent of cerebral damage involving neural control of the subsystems of speech production: articulation, phonation, respiration and resonation (McHenry 1999, 2000; Theodoros et al. 1995). The speech disorder may impair the following aspects of articulatory movements: strength, range, accuracy and speed (Duffy 1995; Bartle 2006; Kuruvilla et al. 2008). Because different neural subsystems can be damaged, dysarthric TBI individuals apply different phonological processes. Obviously, the more severe dysarthria, the higher the number of types of processes applied (Połczyńska-Fiszer and Pufal 2006). However, the most commonly applied processes in the TBI dysarthria are ICC, spirantisation and vowel centralisation (Połczyńska-Fiszer 2006; Połczynska 2009). Furthermore, problems with producing stop consonants (ICC and spirantisation) causes the highest degree of difficulty among the TBI population.

SNHL individuals are characterised by limited perception of sound frequencies and poorer frequency resolution, and narrower intensity of dynamic range of sounds (Boothroyd 1984; Moore 1995; Jürgens et al. 2009). Having distorted aural perception of speech sounds, especially high frequency sounds, SNHL individuals who use oral language produce distorted sounds. SNHL speakers often rely on tactile cues which provide feedback during phoneme articulation (Nasir and Ostry 2008). Thus, speech segments that are most commonly affected by hearing loss are those with little tactile feedback, that is vowels (Lane et al. 2007; Ménard et al. 2007) and sibilants (the latter are imperceptible to many SNHL individuals as they are high frequency sounds) (Matthies et al. 1994; Pittman and Stelmachowicz 2000; Kosky and Boothroyd 2001; Halpern and Tobin 2008). Distortions of sibilants are particularly evident in languages with a high number of sibilants, such as Polish that has 12 sibilants articulated in three places of articulation (located in close proximity). More specifically, Polish has three voiceless sibilant fricatives: dental /s/, alveolar retroflex /ʂ/ and alveolo-palatal /ɕ/. Obviously, this contrast is considerably less salient in terms of tactile cues and frequency, than the English contrast /s/ – /ʃ/. According to the study carried out by Marecka and Połczyńska (2009), the most commonly applied process by Polish SNHL speakers is sibilant imprecision (the name denotes imprecise articulation of high frequency phonemes), whereas the second most frequently applied process is vowel centralisation.

The aim of this study is to compare phonological processes in first (L1, Polish) and second (L2, English) language applied by two different clinical groups: (1) dysarthric individuals with TBI and (2) SNHL speakers. The results for the TBI group are based on the study conducted by Połczyńska (2009) and the results for the SNHL group come from the study by Marecka and Połczyńska (in preparation). We hypothesise that there should be more phonological processes in L2 in both groups because all the subjects were much less proficient in this language than in their L1. This is because L2 is a more demanding linguistic context and speakers are expected to compromise the communicative value of their L2 utterances to a greater degree than in L1. At the same time we expect the most common processes characteristic of both groups (that is, ICA and spirantisation for the TBI group and sibilant imprecision and vowel centralisation for the SNHL group) to be applied with a comparable frequency in both languages due to a universal nature of the processes in the investigated populations.

2. METHODS

Nine subjects participated in the study. The TBI group comprised six subjects (one female and five males) and the SNHL group comprised three subjects (two females and a male). The TBI group did not have any type of hearing impairment, whereas the SNHL group was not diagnosed with paresis of their articulatory musculature. The average age of the participants was 26.3 (SD=5.58) in the TBI group and 24.6 years in the SNHL group (SD=3.39). Table 1 provides further biomedical information about the subjects, along with their L2 proficiency level that was verified with *The Clinical Test of Proficiency in English as the Second Language* (CTP ESL) (Połczyńska-Fiszer 2006; Połczyńska-Fiszer and Mazaux 2008).

All the participants performed the *Polish Dysarthria Test for TBI Patients* (Połczyńska-Fiszer and Pufal 2006) and the *English as the Second Language Test in Dysarthria* (Połczyńska-Fiszer 2006). The tests consist of sets of tasks designed to verify speech in different phonological settings, e.g. vowels and syllables produced in isolation, repetitions, reading and spontaneous speech.

The patients were recorded individually with an Audacity programme. The speech samples were then analysed with PRAAT (Boersma and Weenink 2005) and transcribed by the authors in order to identify phonological processes. The processes were counted as follows: (1) the total number of processes in all the subjects in each group in all the tasks, (2) the total number of processes in L1 and L2 separately in all the subjects in all the tasks. The same procedure was applied for each subject individually. We calculated the number of types of phonological processes (e.g. stopping, vowel centralisation) applied by the participants in their speech. We then counted the frequency of occurrence of a given process (x%) in relation to all the processes applied (100%). The investigation of occurrence of processes was carried out with the spectrographic analysis carried out in the PRAAT software and it was also based on the authors' perceptual analysis. Pronunciation errors that are typically made by Polish users of English (e.g. devoicing of obstruents in word-final position) were not classified as clinical phonological processes in the speech of the subjects.

Table 1: Biomedical data of six TBI and three SNHL subjects and the results of CTP ESL

Subjects	Sex	Age	Severity of dysarhtria	L2 proficiency	Type of injury	Coma duration	Time after awakening
T1	M	27	mild	intermediate	intracerebral haematoma of the right temporo-parietal lobe with a perforation of the ventrical system	6 weeks	1 year
T2	M	22	mild	upper- intermediate	extensive brain oedema	8 weeks	5 years
T3	M	25	mild	intermediate	a trace of blood in the right parietal horn; a small hypodense region in the semiovale centre	6 weeks	2 years 7 months
T4	M	23	moderate	intermediate	subdural haematoma of the right fronto-temporal lobe	3 weeks	1 year 4 months
T5	F	22	moderate	poor	subdural haematoma of the biletaral frontal lobe	3 weeks	1 year 6 months
T6	M	38	moderate	poor	haemorrhagic foci in the frontal regions and in the 3rd ventricle	7 weeks	1 year 5 months
			Severity of SHNL		Age of acquisition of SNHL	Communication mode	Heating aid
S1	F	26	moderate	pre-intermediate	7 months	speech	often
S2	M	20	profound	intermediate	congenital	speech	yes
S3	F	28	profound	poor	congenital	sign language, speech	yes

3. RESULTS

In total, the TBI group applied 46 types of phonological processes and the SNHL group applied 45 types of processes. In L1 the TBI group applied 34 types of processes, whereas the SNHL group used 30 types of processes. In L2 the TBI group applied 44 types of processes and the SNHL group – 43 types. Although the results are not statistically significant, it can clearly be observed that both groups used more processes in their L2: in the TBI group there was an increase by 29.4% and in the SNHL group there was an increase by 43.3%.

The most common processes in L1 in the TBI group was ICC (17.2% of all the processes applied) and the most common process in L2 was consonant deletion (13.6%). The most frequent process in L1 and L2 in the SNHL group was sibilant imprecision (29, 3% and 29% of all the processes applied). Among the remaining ten processes applied most commonly in L1 by the TBI group were: spirantisation (14.3%), consonant cluster reduction (10.7%), vowel centralisation (8.7%), consonant deletion (8.6%), devoicing (8.5%), unstressed vowel deletion (3.7%), deaffrication (3.2%), vowel nasalisation (3%), CA (2.7%) and stopping of fricatives (2.3%). In L2 the remaining ten most frequent processes were: consonant cluster reduction (13%), ICC (9.8%), spirantisation (9.6%), vowel centralisation (6.2%), vowel nasalisation (5.7%), unstressed vowel deletion (3.8%), devoicing (2.6%), backing, affrication of plosives and stressed vowel deletion (2.2%). Among the remaining ten most commonly applied processes in L1 by the SNHL group were: vowel centralisation (28.6%), devoicing (8%), voicing (5.9%), denasalisation (4.2%), deaffrication (3.5%), consonant epenthesis (2.6%), affrication (2.6%), vowel epenthesis (2.1%), vowel nasalisation (1.1%) and fronting (0.9%). The remaining ten most commonly applied processes in L2 by the SNHL group were: vowel centralisation (25.8%), voicing (6.3%), devoicing (4.4%), affrication of stops (3.8%), consonant epenthesis (3.1%), consonant cluster reduction (2.3%), stopping of fricatives (1.9%) consonant deletion (1.8%), fronting (1.8%) and affrication of fricatives (1.7%).

We found that in the SNHL group the two subjects with lower L2 proficiency level (S1 and S3) applied significantly more processes in L2 than in L2. On the other hand, S2 (the subject with a higher L2 proficiency level) applied more processes in L1 than in L2 (this result is also statistically significant). The picture is much more blurred in the TBI group where the only subject in who the difference between the number of applied phonological processes in L1 and L2 reached statistical significance was T6 (low L2 proficiency level, moderate dysarthria). The remaining five subjects also had a higher number of processes in L2, yet the difference was not statistically significant, even in the subjects whose L2 proficiency level was higher and the degree of dysarthria was less severe.

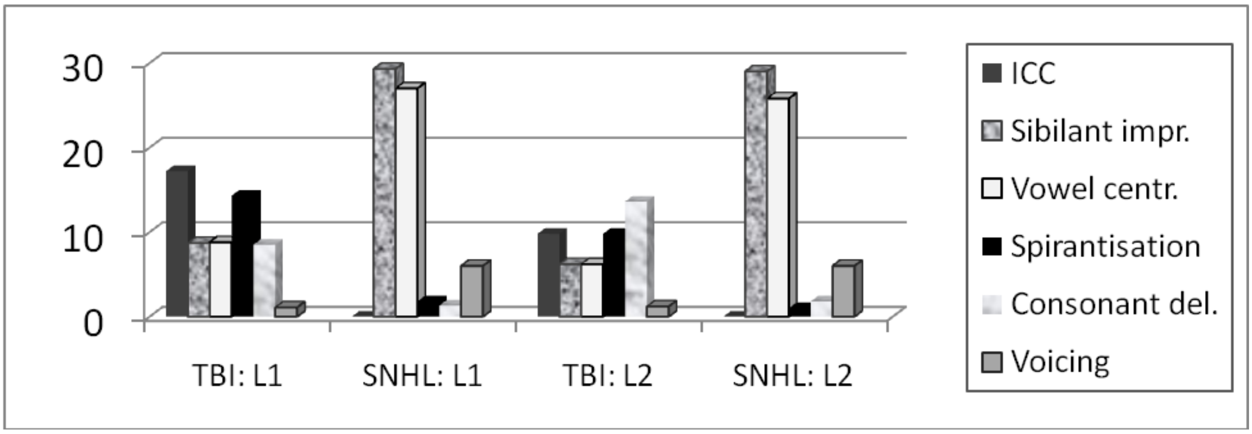
4. DISCUSSION

As indicated in the previous section, the TBI group and the SNHL group applied a similar number of processes in all the tasks in both languages. The groups both exhibited a higher number of phonological processes in L2. This implies that some processes were idiosyncratic in the speakers' L1 and they appeared only in L2. It appears that L2 is a more demanding phonetic environment – none of the participants had a high proficiency in their L2. Hence, we may assume that L2 phonology of the subjects was not completely acquired. L2 articulatory patterns were not sufficiently exercised and for this reason they caused more problems (Połczyńska-Fiszer 2006; Połczyńska 2009). The subject S2 was the only one to have more processes in L1. This might indicate that in case of individuals with SNHL who have higher L2 proficiency level language practice and conscious effort can facilitate the production of phonemes in their L2. Despite being qualified as an intermediate L2 speaker in the CTP ESL, S2 had indeed a high command of L2. It was mainly his general oral comprehension problems stemming from his hearing loss that resulted in the lower result in the CTP ESL. Alternatively, we suggest that English phonetics might be universally easier than Polish phonetics for SNHL individuals because the number of sibilant sounds in English is considerably lower than in Polish. Although this hypothesis needs verification, we postulate that English as L2 might be an easier phonetic environment for SNHL individuals with higher L2 proficiency.

Although the number was higher in L2 in both groups, the TBI group and the SNHL group clearly differed with regard to the type of processes applied in L1 and L2. Fig. 1 presents six most commonly used processes by the two groups. The three most frequently applied processes by the TBI group were: ICC, spirantisation and consonant deletion. The three processes which occurred the most commonly in the SNHL group were: sibilant imprecision, vowel centralisation and voicing.

It can be observed that the SNHL group was very consistent in the application of processes in their L1 and L2. It was mentioned that SNHL individuals exhibit problems with perceiving sibilants and vowels (e.g. McGarr et al. 2004, Ménard et al. 2007). These problems clearly translate into speech production. Even though L2 is a more demanding phonetic environment, the sounds which are difficult to perceive turn out to be produced with errors (i.e. they are universal). Thus, sibilant imprecision and vowel centralisation appeared similarly frequently: 58% of all types of processes in L1 and 54.9% in L2. All the remaining processes were used more rarely (the result is statistically significant).

Figure 2: The frequency of occurrence of the six most common processes in the TBI group and the SNHL group.



At the same time, we can see that processes were distributed more evenly in the TBI group and that more radical processes were used in L2. In particular, the two most frequent processes in L1 were ICC and spirantisation, whereas the most two commonly applied processes in L2 were consonant deletions (consonant deletion and consonant cluster reduction). Obviously, phoneme deletions are viewed as more radical processes because they leave no trace of the sound that was supposed to be articulated, while processes such as ICC serve to mildly modify a phoneme that is difficult to produce so that it lacks the problematic feature.

The phonetic and acoustic analyses imply that the phonological processes in the TBI group were caused by the weakness of active articulators: the tongue, lower jaw, lips, soft palate and vocal folds (usually in combination). In consequence, the TBI subjects had problems with forming a full stricture of the vocal tract that is required to perform more extensive gestures of the lips, the tongue and the lower jaw, to shut off the nasal passage during the articulation of oral phonemes and to excite the vocal folds properly (cf. Duffy 1995. Kuruvilla 2008; Połczyńska 2009). In the case of SNHL individuals, the processes applied stemmed from the perceptual difficulties that led to the lack of acoustic feedback and possibly underdeveloped sound inventory.

Both groups applied proccesses which commonly occur in L1 acquisition, as well as processes which are idiosynchratic in typically developed children. Moreover, the idiosynchratics processes, such as sibilant imprecision, vowel centralisation and ICC, were among the most frequently applied processes. Idiosynchratic processes in the two clinical populations under investigation are caused by an organic impairment of articulatory musculature or by distorted perception of sounds. The two types of impairments result in a specific set of processes which are applied both in L1 and L2. Hence, processes in the two clinical groups are quite regular and they are relatively easy to predict.

5. CONCLUSION

The aim of this study was to investigate phonological processes which occur in L1 and L2 of two clinical groups: individuals with TBI and SNHL. The results indicate that individuals from both groups applied a similar number of processes in all the tasks in both languages. The subjects applied a higher number of phonological processes in L2. At the same time, the groups clearly differed with regard to the type of processes applied in L1 and L2. The TBI group applied a higher number of more radical processes in L2 than in L1. On the other hand, the two most common processes in the SNHL group, sibilant imprecision and vowel centralisation that constituted over a half of all the types of processes applied, were used with nearly the same frequency of occurrence in L1 and L2. Thus, it might be argued that impaired perception of speech sounds results in a more consistent application of phonological processes in L1 and L2, while distorted speech production leads to the application of more radical processes in L2.

6. REFERENCES

Bartle, C.J, Goozée, J.V., Scott, D., Murdoch, B.E, Kuruvilla, M. 2006. EMA assessment of tongue-jaw co-ordination during speech in dysarthria following brain injury. *Brain Injury* 20(5). 529-545.

Barlow, J.A. 2001. A preliminary typology of initial clusters in acquisition. *Clinical Linguistics and Phonetics* 15. 9-13.

Boersma, P., Weenink, D. 2005. *Praat: Doing phonetics by computer* (Version 4.3.22) [Computer programme]. <http://www.praat.org>. Last access 8 September 2005.

Boothroyd, A. 1984. Auditory perception of speech contrasts by subjects with sensorineural hearing loss. *Journal of Speech, Language, and Hearing Research* 27. 133-144.

Grunwell, P. 1985. *PACS: Phonological Assessment of Child Speech*. Winsor: NFER-Nelson.

Diver, W. 1979. Phonology as human behavior. In: Aaronson, D., Reiber, R. (eds.), *Psycholinguistic research: Implications and applications*. Hillside NJ: Lawrence Erlbaum. 161–186.

Dressler, W. U., Mayertaler, W., Panagl, O., Wurzel, W. (eds). 1987. *Leitmotifs in natural phonology*. Amsterdam & Philadelphia: John Benjamins.

Duffy, J.R. 1995. *Motor speech disorders. Substrates, differential diagnosis and management*. St Louis: Mosby.

Dziubalska-Kołaczyk, K. 2002. *Beats-and-binding phonology*. Frankfurt am Main: Peter Lang.

Grunwell, P. 1987. *Clinical phonology*. Kent: CroomHelm.

Ingram, D. 1990. *Phonological disability in children.* London: Whurr.

Kosky, C. Boothroyd, A. 2001. Perception and production of sibilants by children with hearing loss: A Training Study. *Volta Review* 103(2). 71-98.

Kuruvilla, M., Murdoch, B., Goozée, J. 2008. Electropalatographic (EPG) assessment of tongue-to-palate contacts in dysarthric speakers following TBI. *Clinical Linguistics and Phonetics* 22(9). 703-725.

Lane, H., Matthies, M., Guenther, F.H, Denny, M., Perkell, J.S, Stockmann, E., Tiede, M., Vick J., Zandipour, M. 2007. Effects of short- and long-term changes in auditory feedback on vowel and sibilant contrasts. *Journal of Speech, Language, and Hearing Research* 50. 913-927.

Halpern, O., Tobin, Y. 2008. 'Non-Vocalisation'–A phonological error process in the speech of severely and profoundly hearing impaired adults according to the theory of phonology as human behaviour. *International Journal of Clinical Phonetics and Linguistics* 22(10-11). 824-834.

Jürgens, T., Brand, T., Kollmeier, B. 2009. Predicting consonant recognition in quiet for listeners with normal hearing and hearing impairment using an auditory model. *The Journal of the Acoustical Society of America* 125(4). 2533.

Marecka M, Połczyńska M. 2009. Phonological processes in Polish speakers with sensorineural hearing loss. In: Marrero, V., Pineda, I. (eds.), *Proceedings of the II International Conference on Clinical Linguistics*. Madrid: Euphonia Ediciones. 54-62.

Matthies, M.L, Svirsky, M.A., Lane, H.L, Perkell, J.S. 1994. A preliminary study of the effects of cochlear implants on the production of sibilants. *The Journal of the Acoustical Society of America* 96. 1367-1373.

McGarr, N.S, Raphel, L.J., Kolia, B., Vorperian, H.K, Harris, K. 2004. Sibilant production in speakers who have hearing loss: Electropalatographic and Perceptual Evidence. *The Volta Review* 104(3). 119-140.

McHenry, M. A. 1999. Aerodynamic, acoustic, and perceptual measures of nasality following traumatic brain injury. *Brain Injury* 13. 281–290.

McHenry, M. A. 2000. Acoustic characteristics of voice after severe traumatic brain injury. *Laryngoscope* 110. 1157–1161.

Ménard, L., Polak, M, Denny, M., Burton, E., Lane, H., Matthies, M.L., Marrone, N., Perkell, J.S., Tiede, M., Vick, J. 2007. Interactions of speaking condition and auditory feedback on vowel production in postlingually deaf adults with cochlear implants. *Journal of the Acoustical Society of America* 121(6). 3790-3801.

Moore, B.C. J. 1995. *Perceptual consequences of cochlear damage.* Oxford: Oxford Medical Publications.

Pittman, A., Stelmachowicz, P. 2000. Perception of voiceless fricatives by normal-hearing and hearing impaired children and adults. *Journal of Speech, Language,and Hearing Research* 43. 1389–1401.

Połczynska-Fiszer, M. 2006. *First and second language dysarthria in TBI patients after prolonged coma.* Unpublished doctoral dissertation, Adam Mickiewicz University.

Połczyńska-Fiszer, M., Pufal, A. 2006. Classification of dysarthria in Polish TBI patients using acoustic analysis. *Acta Neuropsychologia* 4(4). 257–285.

Połczyńska-Fiszer, M., Mazaux, J.-M. 2008. *Second language acquisition after traumatic brain injury: A case study*. Disability and Rehabilitation 30. 1397-1407.

Połczyńska, M. 2009. *Dysarthric processes in first and second language used by patients with traumatic brain injury*. *The Asia Pacific Journal of Speech, Language and Hearing* 12(2). 137-156.

Sarno, M.T., Buonaguro, A., Levita, E. 1986. Characteristics of verbal impairment in closed head injury patients. *Archives of Physical medicine and Rehabilitation* 67. 400-405.

Stampe, D. 1972. *A dissertation on natural phonology*. New York, Garland.

Theodoros, D. G., Murdoch, B. E., Stokes, P. D., 1995. A physiological analysis of articulatory dysfunction in dysarthric speakers following severe closed-head injury. *Brain Injury* 9. 237–254.

Tobin, Y. 1997. *Phonology as human behavior: Theoretical implications and clinical applications*. Durham, NC & London: Duke University Press.

Wyllie-Smith, L., McLeod, S., Ball, M.J. 2006. Typically developing children and speech-impaired children’s adherence to the sonority hypothesis. *Clinical Linguistics and Phonetics* 20. 271-291.

Prosodic Transfer: an event-related potentials approach

Adèle-Elise Prévost

McGill University
adele-elise.prevost@mail.mcgill.ca

ABSTRACT

This study investigates the possible electrophysiological evidence of the influence of L1 prosodic structure on a speaker's second language, specifically in the context of the Prosodic Transfer Hypothesis of Goad & White (2004, 2009), with Turkish as the L1 and English as the L2. Turkish prosodic structure differs from English in its treatment of articles in ways that suggest that Turkish articles are affixal clitics whereas English articles are free clitics. Crucially, it follows that a correct English article-adjective-noun sequence violates Turkish prosody, since adjectives cannot intervene between articles and noun heads in Turkish, and therefore that Turkish speakers will be unable to correctly prosodify the sequence. Behavioural production evidence – in which Turkish speakers delete, substitute, or stress the English article in asymmetrical ways predictable by prosodic structure – robustly supports this claim. The current experiment uses ERP recording to elucidate the online processing of Turkish speakers hearing English sentences that either do or do not violate Turkish prosodic structure, with the aim of demonstrating real-time neural responses to L1-L2 prosodic mismatch.

Keywords: Prosodic Transfer, ERP, articles, Turkish.

1. INTRODUCTION

A currently fertile area of SLA research is the possible impact of a speaker's first language (L1) on learning a second language (L2). Recent studies debate whether the structural organization of L1 prosody may influence the prosody of the L2 (e.g. Goad, White and Steele 2003, Trenkic 2007, Kupisch & Snape 2009). There have so far however only been behavioural studies on the topic, in spite of the presence of L1 transfer effects in other subdomains of neurolinguistics such as morphosyntax (e.g. Sabourin & Stowe, 2008). A processing study demonstrating the relevance of L1 prosodic cues in guiding L2 comprehension will provide robust support for a theory where L1 prosody influences L2 acquisition in systematic ways, such as the Prosodic Transfer Hypothesis (PTH) (Goad, White and Steele 2003; Goad and White 2004), arguing against predominantly syntactic models of L2 variability (e.g. Hawkins 2003, Trenkic, 2007). Moreover, an investigation of the neurocognitive effects of prosodic violations will lay the groundwork for subsequent important research in second language acquisition.

2. CONTEXT

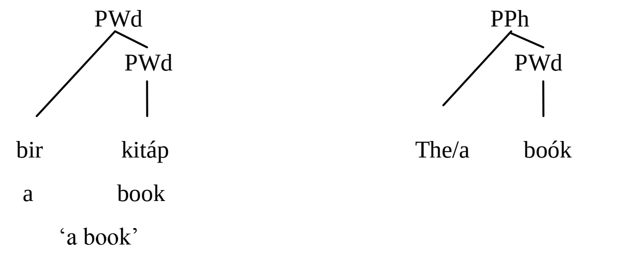
2.1. Difference in prosodic structure of Turkish and English articles

One well-studied case from the linguistic literature is that of L1 speakers of Turkish learning English as their L2. The prosodic structure of Turkish DPs differs from that of English in well-delineated ways. Turkish has no definite article, only an indefinite article, *bir* (Kornfilt, 1997). This article only ever surfaces without stress. Other Turkish determiners, such as numerals, require stress – among them *bir*, 'one', which although homophonous with the indefinite article observes a strictly different distribution (Kornfilt 1997; Goad and White 2004). The unstressed indefinite article is adjoined to the prosodic word (PWd) of its host as an 'affixal clitic', as shown in (1).

Unlike Turkish, English has two articles, definite and indefinite. Moreover, English articles link directly to the phonological phrase (PPh) as 'free clitics' as in (2); Turkish is argued to lack this possible representation (Selkirk 1996; Kornfilt 1997; Goad and White 2004). (All examples are adapted from Goad and White 2009)

(1) **Article structure**

a. Turkish article: Affixal clitic　　b. English article: Free clitic

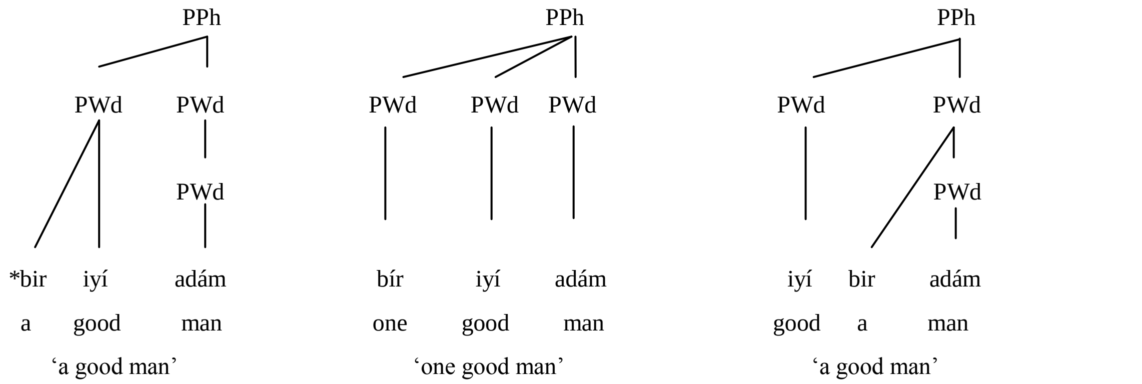


2.2. Predictions for L2 acquisition

Crucially, according to the PTH, the difference in clitic type above predicts a systematic asymmetry in the production of L2 English by L1 Turkish speakers. Affixal clitics must immediately precede the head noun if they are to attach to it, and this is only possible in a subset of English sentences. Thus, DPs without adjectives may surface as target-like, but production of DPs with adjectives will be impaired since English word order places the adjective between the article and its head noun, which prohibits cliticization as seen in (3) below (Goad and White 2004). Comparable Turkish constructions appear in (4): a numeral always precedes the adjective as an independent PWd, bearing stress, whereas an indefinite article must follow the adjective, and never bears stress.

(2) **Word order**

a. English: article-adjective-noun　b. Turkish: Numeral-adjective-noun　c. Turkish: Adjective-article-noun



2.3. Behavioural evidence

The asymmetry between constructions with and without adjectives has been robustly attested in behavioural production studies (Goad & White 2004, 2009). Turkish subjects rely on repair strategies such as substitution or stressing far more in DPs with adjectives, indicating that the determiner has been prosodified as an independent PWd. According to the PTH, the reason for this asymmetry is that English constructions of this type require a prosodic representation that is unavailable in the L1: since Turkish does not allow free clitics, the prosodic structure required in (2a) is impossible for subjects to build (Goad & White, 2009).

2.4. ERP context and precedent

The experiment proposed herein presents itself as a preliminary neurolinguistic investigation of the prosodic structure violations hypothesized above. There exists some prior electrophysiological evidence that neurocognitive responses to aspects of the L1 can transfer into the L2 grammar, for instance in terms of feature agreement (Sabourin and Stowe, 2008) and word order (Drury, Bourguignon, Lin and Steinhauer, in progress). Prosodic interference itself, however, remains an important gap. In order to test this, we are undertaking experiments wherein L1 Turkish/L2 English speakers are presented in the auditory modality with English sentences, the key conditions of which violate the Turkish prosodic structure seen (2a) above. ERPs act as a measure of potential differences on the scalp corresponding to neural activation time-locked to cognitive events, as recorded by electroencephalography. Electrophysiological correlates of online sentence processing are investigated in order to determine whether any adherence to L1 prosodic structure is in play during L2 sentence comprehension. The excellent temporal resolution of ERP recordings serves to adjudicate on possible transfer effects that may not be visible in an offline task, by revealing the real-time electrophysiology involved in processing L2 prosody. Moreover, the sensitivity and reliability of this method over behavioural data will be instrumental in confirming that the effects predicted in terms of production by the PTH are equally attested in comprehension.

3. METHODOLOGY

3.1. Participants

The test population consists of 15 right-handed speakers of L1 Turkish/L2 English of intermediate English proficiency. The control population is comprised of two groups of 15, English monolinguals and L1 French/L2 English learners. Since French allows the English prosodic structure which is hypothesized to be unavailable to Turkish speakers, the purpose of the French group is to ensure that any possible transfer effects recorded for the Turkish group truly reflect a dependency on L1 structure, rather than simply being L2 effects. All participants are healthy adults between the ages of 18 and 40, screened for no known history of neurological or speech/language disorders, and with normal hearing and normal or corrected-to-normal vision.

3.2. Procedure

Participants are fitted with a standard electrode cap according to the international 10-20 system, with reference electrodes at the mastoid bones and bipolar electro-oculograms recorded to control voltage differences due to eye movement; trials with EOG interference above 70 µV will be excluded from analysis. The experiment is run in a sound-attenuated booth. Participants are seated comfortably and follow the instructions displayed on a computer monitor while listening to recorded stimuli on headphones. After a short practice session, EEG data is recorded continuously from 20 electrode sites. Electrode impedance will be kept between 1 and 5 kΩ and a relevant bandpass filter will be applied post-recording in order to reduce noise. For the purposes of presenting a consistent baseline, each sentence pair is identical up to the point of violation, and the ERPs are time-locked to the critical word in the violation condition and its corresponding word in the control condition.

3.3. Stimuli

Stimuli consist of grammatical and ungrammatical sentences of English. There are three sentence structures and two stress patterns present in the experimental stimuli as shown in the table below. Grammatical acceptability in Turkish, English and French are also noted, as well as an example of each type. (Boldface indicates stress on the article.)

Table 1: Sentence types and grammaticality

condition list	WORD ORDER	stressed article	GRAMMATICALITY			EXAMPLE SENTENCE
			Turkish	English	French	
C1	S V art ADJ O	N	N	Y	N	Kristin fought a wild bear.
C2		Y	Y	N	N	Kristin fought **a** wild bear.
C3	S V art O	N	Y	Y	Y	Kristin fought a bear.
C4		Y	Y	N	N	Kristin fought **a** bear.

The sentences were recorded by a native speaker and controlled for length and syntax. Article stress was determined according to intensity and pitch, the cues most relevant to stress in Turkish (Levi, 2005). Measurements using Praat (Boersma and Weenick 2010) confirmed discrete pitch and intensity ranges for the stressed and unstressed article. Each participant was presented with 160 test sentences in pseudorandom order. The English sentences are constructed according to grammaticality relations in Turkish, always in pairs of conditions, one where the article is unstressed and one where it is stressed. Conditions 1 and 2 correspond to the prosodic structure in (2a). In English this construction is prosodically licit when the article is unstressed (C1) but illicit when it is unstressed (C2), at least in non-focus contexts, as is the case for all sentences in this study. The inverse exists in Turkish, where C2 is allowed but prosodic wellformedness constraints block C1. In the second pair of conditions, the adjective is omitted, corresponding to the prosodic structure in (1). The crucial discinction between these pairs is that in Turkish C1 corresponds to an ungrammatical sentence but C2 is grammatical, whereas both C3 and C4 are grammatical in Turkish, albeit with differing interpretations (indefinite article and numeral, respectively).

4. DISCUSSION

4.1. Predictions

There have not yet been perception studies on how speakers integrate morphemes into the prosodic structure of their L2. Goad & White do not make the assumption that prosodic transfer would influence L2 perception (2004: 120); moreover, no electrophysiological study to date has investigated this aspect of prosodic perception. The precise nature of the ERP response expected therefore has yet to be determined. However, if the results follow what has been seen in behavioural data, and under the assumption that there is a correspondence between production symmetries and perception, then a number of predictions can be made. We expect a measurable reaction on the EEG in the Turkish population for C1 (L1-L2 prosodic mismatch) but not C2, one that should not appear in the EEG data of the controls. Recall that C1 sentences correspond to the licit English prosodic structure hypothesized as unavailable to Turkish learners, as evidenced by their significant behavioural difficulties in article suppliance for this specific sentence type. Conversely, C2 violates non-focus English prosody but should fit the Turkish structure. In the table below, a violation where one but not the other of the conditions paired presents licit L2 prosody that would constitute an ungrammatical sentence in the L1, is labelled as ‘L1-L2 prosodic mismatch’. This case is contrasted with condition pairs where the two L2 sentences would simply represent a difference in meaning in the L1. Results of testing are currently preliminary and do not appear to contradict the expected pattern to date.

Table 2: ERP predictions by condition pairs

Condition	TURKISH		ENGLISH		FRENCH	
PAIRS	L1-L2 prosodic mismatch	meaning difference	L1-L2 prosodic mismatch	meaning difference	L1-L2 prosodic mismatch	meaning difference
C1/C2	√	N/A	N/A	√	X	N/A
C3/C4	X	√	N/A	N/A	X	N/A

4.2. Direction of ongoing work

4.2.1. Additional contrast in stimuli

An identifiable ERP response for C1 that is not present for C2 would suggest that prosodic transfer is at work in Turkish learners’ real-time processing of English sentences. This would support a prosodic transfer account of the Turkish behavioural asymmetry mentioned above. However, the data predicted in 4.1 present a second interpretative possibility. The prosodic structure in (2a) not only violates Turkish prosodic wellformedness, but necessarily Turkish word order as well. In order to rule out the possibility that any observed C1-C2 contrast will represent a word-order violation rather than a prosodic one, an additional condition will be interspersed in the test conditions, which contains a word order violation but which allows cliticization of the article onto the head noun. These sentences are presented in Table 3.

Table 3: Additional test conditions

condition list	WORD ORDER	stressed article	GRAMMATICALITY			EXAMPLE SENTENCE
			Turkish	English	French	
C5	S V art O ADJ	N	N	N	Y	Kristin fought a bear wild.
C6		Y	N	N	N	Kristin fought **a** bear wild.

4.2.2. Experiment 2: Turkish

Furthermore, in order to claim that any of the results from the experiment detailed above truly represent prosodic transfer from Turkish into English, it should be the case that an equivalent pattern of response is observable for Turkish subjects tested in Turkish. Two versions of the experiment are therefore planned. While Experiment 1 tests Turkish participants and French and English controls on English sentences, Experiment 2 tests exclusively the Turkish population, on Turkish sentences. The experimental procedure of this follow-up is identical to that of the English experiment and sentences are adapted from the English stimuli. The condition of interest is now the contrast between C7 and C8, which are Turkish equivalents to C1 and C2. The table below details the revised predictions expected with the addition of the material above.

Table 4: Predictions

Condition PAIRS	TURKISH			ENGLISH			FRENCH		
	L1-L2 prosodic mismatch	word order violation	meaning difference	L1-L2 prosodic mismatch	word order violation	meaning difference	L1-L2 prosodic mismatch	word order violation	meaning difference
C1/C2	√	√	N/A	N/A	N/A	N/A	X	√	N/A
C3/C4	X	X	√	N/A	N/A	N/A	X	√	N/A
C5/C6	X	√	N/A	N/A	√	N/A	X	X	√
C7/C8	√	√	N/A						

4.3. Conclusion

If L1 prosodic structure is responsible for differences in L2 article supplicance, as claimed in the PTH, and if perception is affected in the same way as producton, then evidence of this dependence on L1 strategies should be visible on the time-locked event-related potential data. By contrasting a ‘prosodic mismatch’ condition pair (C1/C2) to prosodically acceptable pairs where similar variation results only in a meaning difference (C3/C4), or a word-order violation (C5/C6), this study seeks to isolate a neurocognitive pattern of response that coresponds to prosodc structure violation. Confirmation that Turkish speakers of L2 English evidence an electrophysiological correlate of dependence on L1 prosody in their processing of L2 DPs containing adjectives would strongly support the Prosodic Transfer Hypothesis, particularly if the effect is not visible in either control group but presents similarities to the Turkish L1 response on equivalent violations. The current study will also have helped to determine that prosodic transfer is as relevant to comprehension as it is to production. Furthermore, a first mapping of the electrophysiological response to violations of prosodic structure will have been undertaken, complementing previous work on intonational patterns (e.g. Steinhauer, Alter & Friederici, 1999).

5. REFERENCES

Boersma, P., Weenink, D. 2010. Praat: doing phonetics by computer (Version 5.1.30) [Computer program]. Retrieved February 21, 2010, from http://www.praat.org/

Drury, J.E., Bourguignon, L., Lin, T., Steinhauer, K. 2009. ERP evidence of native language influence on the processing of late acquired second language. Poster at the Neurobiology of Language Meeting, Chicago, Oct.

Goad, H., White, L., Steele, J. 2003. Missing inflection in L2 acquisition: Defective syntax or L1-constrained prosodic representations? Canadian Journal of Linguistics 48. 243-263.

Goad, H. & White, L. 2004. Ultimate attainment of L2 inflection: Effects of prosodic structure. In: Foster-Cohen, S., et al. (eds.), EUROSLA Yearbook 4. Amsterdam: John Benjamins. 119-145.

Goad, H., & White, L. 2009. Prosodic transfer and determiners in Turkish-English interlanguage. In: Snape, N., et al. (eds.), Representational deficits in L2 acquisition. Amsterdam: John Benjamins. 1-26.

Hawkins, R. 2003. 'Representational deficit' theories of (adult) SLA: Evidence, counterevidence and implications. Plenary paper presented at EUROSLA, Edinburgh, Sept.

Kornfilt, J. 1997. Turkish. London: Routledge.

Kupisch, T., Snape, N. In revision. Ultimate attainment of L2 articles: A case-study of an endstate L2 Turkish-English speaker.

Levi, S.V. 2005. Acoustic correlates of lexical accent in Turkish. Journal of the International Phonetic Association 35. 73-97.

Sabourin, L., Stowe, L.A. 2008. Second language processing: when are first and second languages processed similarly? Second Language Research 24(3). 397-430.

Selkirk, E.O. 1996. The prosodic structure of function words. In: Morgan, J.L. and K. Demuth (eds.), Signal to Syntax. Mahwah NJ: Lawrence Erlbaum. 187-213.

Steinhauer, K., Alter, K., Friederici, A. 1999. Brain potentials indicate immediate use of prosodic cues in natural speech processing. Nature neuroscience 2(2).191-196.

Trenkic, D. 2007. Variability in L2 article production -beyond the representational deficit vs. processing constraints debate. Second Language Research 23. 289-327.

The production of French nasal vowels by advanced Japanese and Spanish learners of French: a corpus-based evaluation study

Isabelle Racine[1], Sylvain Detey[2], Nathalie Bühler[1], Sandra Schwab[1], Françoise Zay[1] and Yuji Kawaguchi[3]

[1]ELCF, University of Geneva, [2]SILS, Waseda University and LiDiFra, Rouen University, [3]Tokyo University of Foreign Studies

isabelle.racine@unige.ch, sylvain.detey@aoni.waseda.jp, nathalie.buehler@unige.ch, sandra.schwab@unige.ch, francoise.zay@unige.ch, ykawa@tufs.ac.jp

ABSTRACT

In the past, few studies have investigated the production of French nasal vowels by non-native speakers, and none of these – as far as we know – have been corpus-based. In this study, productions of /ɑ̃/, /ɔ̃/ and /ɛ̃/ by Japanese and Spanish advanced learners of French, collected from the multitask IPFC corpus (*InterPhonology of Contemporary French*), have been assessed in a three-step process: 1) a non-expert native assessment of the vowel quality through a lexical identification task and a goodness task; 2) an expert native assessment of the postvocalic excrescences of the learners' productions; 3) an acoustic analysis of the postvocalic excrescences on a subset of productions. The results are discussed in light of the psycholinguistically distinct processes involved in the different tasks.

Keywords: L2 production, French nasal vowels, interphonology of contemporary French, multitask oral corpus, Japanese and Spanish learners of French

1. INTRODUCTION[1]

In the field of L2 phonetics and phonology, corpus-based studies have been rather scarce. However, in recent years, a number of studies have emerged: e.g. for L2 Dutch (Neri et al. 2006), Polish (Cylwik et al. 2009), German, and English in Europe (Gut 2009) and Asia (Visceglia et al. 2009). In the case of L2 French, the project *InterPhonologie du Français Contemporain* (IPFC) was launched in 2008 in order to create a large phonological corpus of oral data collected from speakers of various L1s using a single methodological protocol (Detey and Kawaguchi 2008; Racine et al. to appear; Detey et al. to appear; Detey et al. to appear). The protocol was designed after the one used in the project *Phonologie du Français Contemporain* (Phonology of Contemporary French) for native speakers (Durand et al. 2009, http://www.projet-pfc.net). The IPFC protocol consists of 5 tasks: reading aloud and repetition of a word list, text reading, formal interview with a native speaker, and semi-formal interaction between two learners. Beyond its role as a primary data provider for perceptual experiments and phonetico-phonological analyses, IPFC also aims at raising methodological issues about the articulation between psycholinguistically-oriented interphonology studies and modern corpus linguistics. The data used in the study reported here were all drawn from the IPFC corpus, more specifically from Japanese and Spanish advanced learners of French.

Among the phonological characteristics of French that non-native speakers have to learn are the nasal vowels. Even though the nasal feature [+nasal] can be found in the consonantal systems of Japanese and Spanish, and despite the existence of nasal spreading through phonetic coarticulation and assimilation processes in both languages, nasal vowels are always difficult to learn for Japanese and Spanish learners of French. So far, few studies have tackled the issue of nasal vowel learning in French as a Foreign Language: see for example Takeuchi and Arai (2009) for Japanese learners and Montagu (2002) or Garrott (2006) for American learners. This apparent lack of interest may partly be explained by the complexity of the relationship between the articulatory, acoustic and auditory properties of nasal vowels in French (Delvaux et al. 2002; Montagu 2007).

In our study, the analysis of the nasal vowels (/ɛ̃/, /ɑ̃/, /ɔ̃/) was performed according to a three-step procedure: first a non-expert perceptive assessment through both a lexical identification task and a goodness task; second, an expert perceptive assessment of the postvocalic excrescence (degree of presence of a postvocalic consonant (Johnson et al. 2007)) and third, an acoustic analysis of postvocalic excrescences. The general purpose of our study is to assess the *quality of realization*[2] of the French nasal vowels produced by non-native speakers.

2. NON-EXPERT PERCEPTIVE ASSESSMENT OF THE NASAL VOWELS

2.1. Method

Participants: The speakers were 5 Japanese learners of French (3 males and 2 females; all were students at Tokyo University of Foreign Studies and came from the Tokyo metropolitan area) and 5 Spanish learners of French (2 males and 3 females; all were students at the University of Geneva and came from Spain). They were selected from the IPFC corpus on the basis of their proficiency level in French (B2-C1 according to the Common European Framework of Reference for Languages (CEFRL)). In the perceptual experiments, 32 native listeners were used (half for the lexical identification task and half for the goodness task).

Material: Nine monosyllabic words from the word lists in the IPFC protocol were selected: 3 containing the vowel /ɛ̃/, 3 /ɑ̃/ and 3 /ɔ̃/. Each vowel appeared in 3 different contexts: VC (i.e. *Inde* “India”), CVC (i.e. *tante* “aunt”) and CV (i.e. *pont* “bridge”). All 9 words were produced twice by each learner: in a repetition task and in a reading task. The final stimulus set consisted of 180 words.

Procedure: In the lexical identification task, participants were instructed to listen to individual words and write them down. In case of hesitation (with heterographic homophones, i.e. *pense* “think” for *panse* “belly”), they were asked to write the first word that occurred to them. Each word was presented twice. If they were not able to identify a French word, they were asked to indicate it by checking an appropriate “Unknown word” field. For the goodness task, in order to avoid lexical influence and to force the participants to focus on the vowel, they were instructed to listen to syllables or parts of individual words (i.e. *–ban*, 2nd syllable of *ruban* “band”) and to judge the vowel of each stimulus for its goodness as a member of a given category (/ɛ̃/, /ɔ̃/, /ɑ̃/) using a 1-5 rating scale (1 = very good exemplar; 5 = other vowel): the better the exemplar, the lower the number.

Data analysis: For the lexical identification task, a correct nasal vowel identification rate[3] was calculated as a function of learners’ population, nasal vowel and production task. The correct vowel identification rate was calculated on the basis of the number of answers excluding those indicated as “Unknown word”. For the goodness task, a mean goodness ratings was calculated as a function of learners’ population, nasal vowel and task.

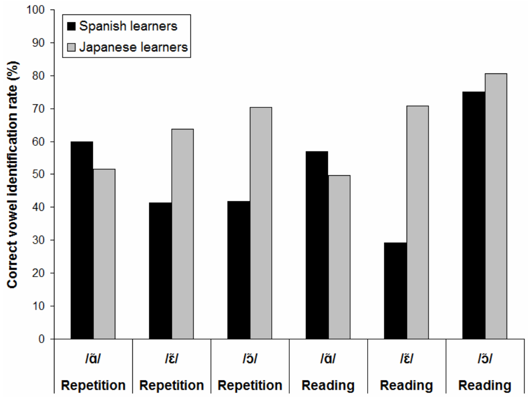
2.2. Results

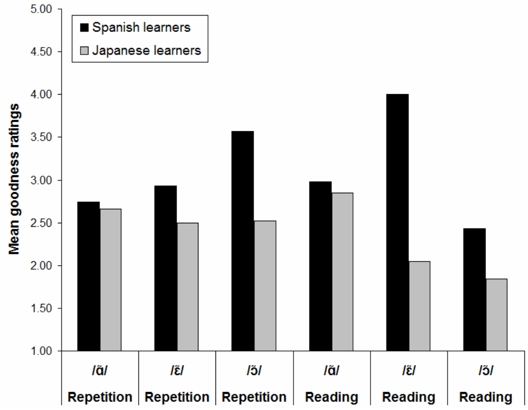
As can been seen in Figure 1 (on the left), which presents the mean correct nasal vowel identification rate (in percentage) for productions by Spanish and Japanese learners as a function of nasal vowel and production task, the correct identification rate is higher for Japanese learners’ productions (64.50%) than for Spanish ones (50.72%). The analysis of variance confirms this pattern. There is a main effect of population ($F1\ (1, 15) = 71.03$, $p<0.001$; $F2\ (1, 6) = 6.83$, $p<0.05$). There is also a main effect of task: the correct identification rate is higher for words produced in the reading task (60.42%) than for those produced in the repetition task (54.78%) (by participants only: $F1\ (1, 15) = 17.43$, $p<0.001$[4]). There is also a main effect of nasal vowel: /ɔ̃/ is better identified (67.02%) than /ɑ̃/ (54.53%) and /ɛ̃/ (51.27%) (by participants only: $F1\ (2, 30) = 7.16$, $p<0.01$).

The results of the goodness task are quite similar[5]. As can been seen in Figure 2 (on the right), which presents the mean goodness ratings for productions by the two learners’ populations as a function of nasal vowel and production task, the ratings are better for Japanese learners’ productions (2.41) than for the Spanish ones (3.11). The analyses of variance confirm this pattern. There is a main effect of population ($F1\ (1, 15) = 147.03$, $p<0.001$; $F2\ (1, 6) = 13.83$, $p<0.01$). There is also a main effect of task: the goodness ratings are better for words produced in the reading task (2.69) than for those produced in the repetition task

(2.82) (by participants only: F1 (1, 15) = 14.81, p<0.01). In this task, /ɔ̃/ obtains the best rating (2.59), followed by /ɑ̃/ (2.81) and /ɛ̃/ (2.87), although in a marginal way (by participants: F1 (2, 30) = 2.86, p = 0.07)[6].

Figures 1 and 2: Mean correct nasal vowel identification rate (Fig. 1, on the left) and mean goodness ratings on a scale of 1 (= very good exemplar) to 5 (= other vowel) (Fig. 2., on the right) for productions by Spanish learners (in black) and Japanese learners (in grey) as a function of nasal vowel (/ɑ̃/, /ɛ̃/ and /ɔ̃/) and task (repetition and reading).





3. EXPERT PERCEPTIVE EVALUATION AND ACOUSTIC ANALYSIS OF POSTVOCALIC EXCRESCENCES

3.1. Method

Participants: The speakers were 11 Japanese learners of French (3 males and 8 females) and 8 Spanish learners of French (2 males and 6 females). The experts were 4 linguists, native speakers of French.

Material: Twelve monosyllables from the word lists used in the IPFC protocol were selected for this study, each of them containing a nasal vowel in an open syllable CV or in a closed syllable VC or CVC. The final stimulus set consisted of 456 words (192 for the Spanish learners and 264 for the Japanese learners; 24 productions for each learner).

Procedure: The degree of postvocalic excrescence was assessed by the experts using a 3-point scale (1 = absence of postvocalic excrescence; 3 = clear evidence of postvocalic excrescence).

Data analysis: The experts' scores were first analyzed to determine inter-rater reliability[7]. An ICC coefficient of 0.72 (p<0.001) was obtained, which indicates high reliability. We then calculated a mean degree of postvocalic excrescence as a function of learners' population, nasal vowel and production task.

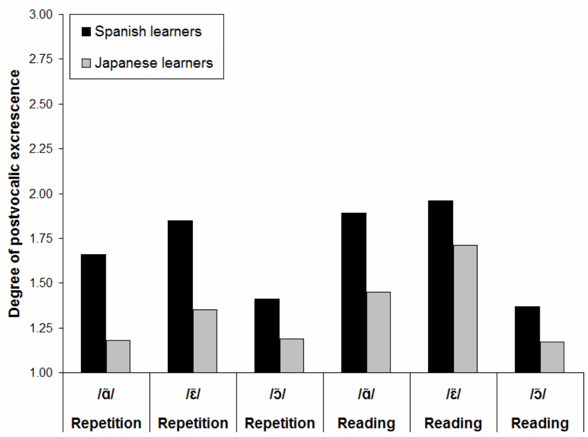
3.2. Results

As can been seen in Figure 3, which presents the mean degree of postvocalic excrescence in the productions of Spanish and Japanese learners as a function of nasal vowel and task, the degree of postvocalic excrescence is higher for Spanish learners' productions (1.69) than for Japanese ones (1.34). The analysis of variance confirms this pattern. There is a main effect of population (F1 (1, 17) = 10.10, p<0.01; F2 (1, 9) = 17.34, p<0.01). There is also a main effect of task: the degree of postvocalic excrescence is higher for the words produced in the reading task (1.69) than for those produced in the repetition task (1.44) (F1 (1, 17) = 7.75, p<0.05); F2 (1, 9) = 5.14, p<0.05). There is also a main effect of nasal vowel: the degree of postvocalic excrescence is lower for /ɔ̃/ (1.29) than for the two other vowels (/ɑ̃/: 1.55 and /ɛ̃/: 1.72) (by participants only: F1 (2, 34) = 23.52, p<0.001)[8].

In order to check the reliability of the perceptual analysis, we carried out an acoustic analysis on 2 items (38 productions of *tant* "so much" and 38 productions of *tante* "aunt"). Acoustic measures were performed by 2 phoneticians[9] using Praat (Boersma and Weenink 2009) on the basis of spectrograms. The occurrences with and without postvocalic consonant were counted after examination of the formant configuration, the formant values and the amplitude difference.

The results for acoustic analysis are convergent with the evaluation performed by experts and show two tendencies that seem to be shared by the two learners' populations. First, the degree of postvocalic excrescence is higher for the words produced in the reading task than for those produced in the repetition task. Thus, in the expert evaluation, the mean degree of postvocalic excrescence for *tante* and *tant* in the reading task is 1.88 *vs* 1.43 for the repetition task ($t(37) = 3.97$, $p<0.001$). Acoustic analyses show a similar effect: if we take into account the totality of the values ($n = 68$), a postvocalic consonant was more often present for words produced in the reading task ($n = 24$) than for words produced in the repetition task ($n = 15$) ($\chi^2 = 4.87$, $p<0.05$). Second, a postvocalic consonant was more often present for *tante* than for *tant*. In the expert evaluation, *tante* obtained a degree of postvocalic excrescence of 1.94 *vs* 1.37 for *tant* ($t(74) = 3.77$, $p<0.001$). Acoustic analysis shows the same pattern: if we take into account the totality of the values ($n = 68$), a postvocalic consonant was more often present for *tante* ($n = 26$) than for *tant* ($n = 13$) ($\chi^2 = 8.47$, $p<0.05$). A global analysis showed that the pattern observed for *tant* and *tante* can be generalized: CV words obtain the lowest degree of postvocalic excrescence. These are followed by VC or CVC words in which the last consonant is [s]. Finally, words that obtain the highest degree of postvocalic consonant are CV or CVC words in which the last consonant is a stop consonant with the same place of articulation as that of the postvocalic consonant.

Figure 3: Mean degree of postvocalic excrescence (on a scale of 1 (= absence of a postvocalic consonant) to 3 (= clear evidence of a postvocalic consonant)) in the productions of Spanish learners (in black) and Japanese learners (in grey) as a function of nasal vowel (/ɑ̃/, /ɛ̃/ and /ɔ̃/) and task (repetition on the left and reading on the right).



4. GENERAL DISCUSSION

Three global tendencies seem to emerge from our results: 1) better performance by Japanese learners as compared to Spanish learners; 2) better results in the reading task as compared to the repetition task in terms of vowel quality, but opposite results in terms of postvocalic excrescences; 3) better results for /ɔ̃/ as compared to /ɛ̃/ and /ɑ̃/.

Concerning the population effect (Japanese > Spanish), one differentiating factor that needs to be considered from a psycholinguistic viewpoint is the degree of *focus-on-form* (Ellis et al. 2002). Given the interlinguistic distance between each L1 and French, it is possible to hypothesize that the Japanese learners might have paid more attention to formal linguistic aspects of their learning than the Spanish learners. It must be borne in mind that French and Spanish differ from Japanese not only linguistically but also in their graphemic systems. The Spanish system is alphabetic with a rather shallow orthography, whereas the Japanese system is rather deep and non-alphabetic. This interlinguistic distance bears strong psycholinguistic implications for the learning process (e.g. new reading procedures and new syllabic types for the Japanese learners). Therefore, at an equal linguistic level, the attentional load may be different for the two populations: with better results for the Spanish learners on both formal and communicative dimensions in the initial stages, but better results for the Japanese learners on the formal level at a latter stage (given a constant attentional focus). Such a strong hypothesis must be tested longitudinally and the results are bound to fluctuate according to inter- and intra-learner variation.

The production task effect identified in our results must be interpreted from a psycholinguistic viewpoint, since the nature of the initial stimuli and the cognitive process at work in the reading and repetition tasks are not identical. More specifically, the repetition task involves auditory perception (and therefore possible misperception in L2, partly due to temporal constraints), whereas the reading task involves visual perception (and therefore a temporally more stable input). In the case of vowel quality, even though correct graphemic identification does not guarantee the production of a phonetically correct unit in the target language system, it seems plausible that the reading task could be more favourable to input faithfulness than the repetition task. In that case, the orthographic input would play a positive role in the identification of certain phonemic categories (Steele 2005). In terms of postvocalic excrescences on the other hand, the opposite results serve as a reminder that orthography can have an effect – a negative one here – on both suprasegmental (Detey and Nespoulous 2008) and segmental levels (Detey et al. 2005): erroneous graphemic segmentation for the first level and automatic graphophonemic activation for the second (Dijkstra et al. 1993). In the repetition task, the degree of presence of an epenthetic consonant – absent from the input – is thus unsurprisingly lower than in the reading task. If we put aside performance errors, three arguments can be put forward to explain the presence of postvocalic excrescences: 1) on a psychoacoustic or phonological level, a perceptual or interphonological reinterpretation of the nasal vowel; 2) on a psycholinguistic level, the activation of a phonological or orthographic lexical representation with a lexicalized epenthetic consonant; 3) on an articulatory level, universal or L1-transferred automatic coarticulation mechanisms.

Our results of the non-expert assessment lead to the following ranking: /ɔ̃/ > /ɑ̃/ > /ɛ̃/. If we follow the hypothesis of Paradis and Prunet (2000), according to which nasal vowels should be considered Oral vowel + Nasal consonant sequences, we must take into account the recent work of Montagu (2002, 2007), which shows that corresponding oral vowels in contemporary French are not /ɔ/, /ɑ/ and /ɛ/ respectively – as it could be assumed from the IPA symbols – but instead /o/, /ɔ/, and /a/. This points to the fact that /ɑ̃/ is the only vowel without an oral equivalent category (/ɔ/) in the L1 system (Japanese or Spanish). Yet, when we take into account the graphemic dimension, it is /ɛ̃/ that seems to be the most costly in terms of cognitive processing, as it has the highest number of graphic variants in French (as compared to /ɑ̃/ and /ɔ̃/): /ɛ̃/ was actually the only one to be represented as a trigram in the reading task (*teinte* and *teint*), and the results for /ɛ̃/ in the reading task concur with those in Garrott's work (2006). Therefore, according to our global results, /ɔ̃/ seems to be the easiest category to learn and to identify.

5. CONCLUSION

Even though a cross-task comparison including the three other tasks of the IPFC corpus (text reading and conversations) seems necessary to further our understanding of the production of the three French nasal vowels by advanced Japanese and Spanish learners, the task effect brought to light so far in our results from the words in isolation already has direct implications: on a methodological level – for research in the field of L2, and maybe even L1, phonology –, and on a pedagogical level – for oral language education. On a methodological level, multitask – and not only single- or double-task – protocols seem to be essential to build up large and multipurpose oral corpora. While this seems to be important in the case of native speakers (see for example the PFC protocol with two reading tasks and two conversation tasks), it is crucial in the study of non-native speakers, whose maturing interphonological (and interphonetic) systems are even more heavily influenced by the psycholinguistic features of each task. Ideally, both modalities (auditory and visual) should be involved, as well as both the perceptive and the productive side of the learner's interphonological system. On a pedagogical level, our results point to the necessity of providing a well-balanced learning environment in which the selected tasks allow phonetico-phonological and phonographemic skills to develop simultaneously.

6. REFERENCES

Boersma, P., Weenink, D. 2009. *Praat: doing phonetics by computer* (version 5.0). http://www.praat.org.

Cylwik, N., Wagner, A., Demenko, G. 2009. The EURONOUNCE corpus of non-native Polish for ASR-based Pronunciation Tutoring System. *Proceedings of SlaTE 2009 – 2009 ISCA Workshop on Speech and Language Technology in Education*. Birmingham, UK.

Delvaux, V., Metens, T., Soquet, A. 2002. French nasal vowels: acoustic and articulatory properties. *Proceedings of the 7th ICSLP* (1). 53-56.
Detey, S., Durand, J., Nespoulous, J.-L. 2005. Interphonologie et représentations orthographiques. Le cas des catégories /b/ et /v/ chez des apprenants japonais de Français Langue Etrangère. *Revue PArole (RPA)* 34/35/36 (supplément). 139-186.
Detey, S., Kawaguchi, Y. 2008. Interphonologie du Français Contemporain (IPFC): récolte automatisée des données et apprenants japonais. *Journées PFC: Phonologie du français contemporain: variation, interfaces, cognition*. Paris, Dec. 11-13, 2008.
Detey, S., Nespoulous, J.-L. 2008. Can orthography influence L2 syllabic segmentation? Japanese epenthetic vowels and French consonantal clusters. *Lingua. International Review of General Linguistics* 118 (1). 66-81.
Detey, S., Racine, I., Kawaguchi, Y. (to appear). Assessing non-native speakers' production of French nasal vowels: a multitask-corpus-based study. *Working Papers in Corpus based Linguistics and Language Education 4*. Tokyo: Tokyo University of Foreign Studies.
Detey, S., Racine, I., Kawaguchi, Y., Zay, F., Bühler, N., Schwab, S. (to appear). Evaluation des voyelles nasales en français L2 en production: de la nécessité d'un corpus multitâches. *Actes de CMLF 2010*, Nouvelle Orléans, 12-15 juillet 2010.
Dijkstra, A., Frauenfelder, U.H., Schreuder, R. 1993. Grapheme-phoneme interaction in a bimodal detection task. *Journal of Experimental Psychology: Human Perception and Performance* 19 (5). 1-20.
Durand, J., Laks, B., Lyche, C. 2009. Le projet PFC: une source de données primaires structurées. In: Durand, J., Laks, B., Lyche, C. (eds), *Phonologie, variation et accents du français*. Paris: Hermès. 19-61.
Ellis, R., Basturkman, H., Loewen, S. 2002. Doing focus on form. *System* 30. 419-432.
Garrott, C. L. 2006. *L1 (English) to L2 (French) transfer: The question of nasalization*. Ms. Dec. 29, 2009 <http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/1b/ee/9c.pdf>.
Gut, U. 2009. *Non-native Speech: a Corpus-based Analysis of Phonological and Phonetic Properties of L2 English and German*. Wien: Peter Lang.
Johnson, K., DiCanio, C., MacKenzie, L. 2007. The acoustic and visual basis of place of articulation in excrescent nasals. *UC Berkeley Phonology Lab Annual Report (2007)*. 529-561.
Montagu, J. 2002. L'articulation labiale des voyelles nasales postérieures du français: comparaison entre locuteurs français et anglo-américains. Actes des *XXIVèmes JEP*. 253-256.
Montagu, J. 2007. *Analyse acoustique et perceptive des voyelles nasales et nasalisées du français parisien*. Unpublished doctoral dissertation. University of Paris III.
Munro, M. J. 2008. Foreign accent and speech intelligibility. In: Hansen Edwards, J. G., Zampini, M. L. (eds), *Phonology and Second Language Acquisition*. Amsterdam: John Benjamins. 193-218.
Neri, A., Cucchiarini, C., Strik, H. 2006. Selecting segmental errors in L2 Dutch for optimal pronunciation training. *IRAL* 44. 357–404.
Paradis, C., Prunet, J.-F. 2000. Nasal vowels as two segments: evidence from borrowings. *Language* 76(2). 324-357.
Racine, I., Detey, S., Zay, F., Y. Kawaguchi (to appear). Des atouts d'un corpus multitâches pour l'étude de la phonologie en L2: l'exemple du projet « Interphonologie du français contemporain » (IPFC). In: Kamber, A., Skupiens, C. (eds). *Recherches récentes en FLE*. Berne: Peter Lang.
Steele, J. 2005. Assessing the role of orthographic versus uniquely auditory input acquiring new L2 segments. *7èmes Rencontres Internationales du Réseau Français de Phonologie*. University of Aix-en-Provence, June 2-4, 2005.
Takeuchi, K., Arai, T. 2009. Strategy for the production of French nasal vowels by Japanese students. *PTLC 2007*. London, UK.
Visceglia, T., Tseng, C.-Y., Kondo, M., Meng, H., Sagisaka, Y. 2009. Phonetic aspects of content design in AESOP (Asian English Speech cOrpus Project). *Proceedings of Oriental-COCOSDA*. Urumuqi, China.

NOTES

[1] We would like to thank Helene N. Andreassen for her help with the final version of this paper.
[2] This notion must not be confused with *intelligibility* or *accentedness*, or even *comprehensibility* or *acceptability* (see Munro 2008).
[3] This was preferred to the more traditional correct lexical identification rate because it focuses on nasal vowel recognition and does not count errors of identification triggered by other factors such as surrounding consonants.
[4] The absence of a main effect by items in the analyses of variance may partly be explained by the small number of items we had in each category (only three).
[5] A correlation of 0.72 ($p<0.001$) between correct vowel identification rate and goodness ratings was obtained, which indicates high reliability between the two tasks.
[6] Moreover both tasks show several interactions, which reveal that the identification rate and the goodness ratings for each vowel vary as a function of population and task. They thus underline the complexity of the system but further analyses are required to interpret them.
[7] The inter-rater reliability coefficient measures the consistency between the assessments of the 4 raters and varies between 0 and 1 (1 indicates a perfect consistency between the raters).
[8] Moreover there is an interaction between the vowel and the task (by participants only: $F1\ (2, 34) = 3.67$, $p<0.05$), which shows that postvocalic excrescence varies as a function of the task and underlines that the variables examined are linked together.
[9] We would like to thank Naoki Marushima for his help with the acoustic measures.

Intelligibility and foreign accentedness in a Context-and-Language-Integrated-Learning (CLIL) setting

Lucrecia Rallo Fabra and Maria Juan-Garau

University of the Balearic Islands
lucrecia.rallo@uib.es

ABSTRACT

The present paper investigates the effect of a semi-immersion setting on speech production by Catalan-Spanish learners of English as a foreign language (FL). The data presented here are part of a longitudinal research project which explores the short- and mid-term effects of CLIL, a relatively recent trend in European mainstream education in which a FL is used as the medium of instruction. Two groups of learners varying in amount and nature of formal FL instruction and a group of native English speakers were recorded performing a controlled speaking task at two data collection times. Four expert native English judges assessed five short excerpts in terms of intelligibility ("easy", "mid" or "difficult" to understand) and immediately afterwards they rated the speech samples for foreign accent on a 1-to-5 scale. As expected, the read speech samples by the CLIL learners were judged to be more intelligible and less accented than the samples by their peers in conventional formal instruction (FI), but both groups differed significantly from the native English group. No significant short-term improvement in either intelligibility or foreign accent was found after one year of CLIL instruction, indicating that gains in learners' pronunciation in a content-based learning context are more likely to appear in the long run.

Keywords: intelligibility, foreign accent, formal instruction, CLIL

1. INTRODUCTION

Empirical research exploring production of nonnative speech sounds by speakers learning the target language in a naturalistic setting has shown that extensive exposure to the target language is determinant to reach native-like attainment (Bohn and Flege 1997; Moyer 1999; Piske et al. 2002). Other accounts that compare speech learning in naturalistic and formal instruction settings (Best and Tyler 2007; Piske 2007) have suggested that the conditions of most foreign language (FL) classrooms in terms of amount and quality of FL input are not ideal. As Piske (2007: 305) argues, instructors rarely use teaching methods directed at improving the pronunciation of the target language, which, in turn, diminishes the students' chances to use it in the classroom. He finishes making a plea to FL teachers to create learning environments in which FL students are encouraged to use the target language as much as possible.

Prior work examining FL production in instructional settings has revealed that amount of formal instruction (FI) has little impact in ameliorating foreign accent in the L2. For instance, García Lecumberri and Gallardo (2003) examined intelligibility and foreign accent in a group of Basque schoolchildren who differed in starting age of exposure to English. On the basis of ratings by native English listeners, they found that students were far from acquiring nativelike pronunciation and noted the relevance of intensive aural exposure for FL speech learning. Similarly, Fullana (2006) found that overall foreign accent among Spanish-Catalan learners in a formal setting did not decrease as a function of amount of FL input quantified in hours of instruction.

Content and language integrated learning (CLIL) is a relatively recent trend in European mainstream education in which "content subjects such as history or physics are taught through the medium of a second or foreign language. This language is a target language (TL) other than the language of the learners, the teachers or the language used in the rest of the school curriculum" (Pérez-Vidal 2009: 3). CLIL appeared as an alternative to traditional teaching methods in FI, which have proved to be limited in terms of learners' speech production. So far there is evidence that CLIL programmes can facilitate FL learning in various domains (Dalton-Puffer and Nikula 2006; Lorenzo et al., forthcoming). However, to our knowledge, no previous studies have investigated the effect of CLIL instruction on intelligibility and accentedness of FL speech .

The European Sections Programme, as the CLIL scheme implemented in the Balearic Islands is known, has grown exponentially since it was first launched in the academic year 2004-2005 (see Pérez-Vidal and Juan-Garau 2010 for a more detailed account). Subject specialists with a good command of the FL are normally the ones in charge of CLIL instruction. They act in coordination with FL specialists, but tandem teaching is not the norm. In order to be admitted to the programme, learners need to have their parents' consent. Other selection criteria may additionally apply (e.g. student records). Stakeholder satisfaction with the programme tends to be high, but its real learning outcomes are still being assessed.

The aim of the present study is thus to investigate to what extent CLIL instruction can enhance Spanish-Catalan learners' oral performance in English as a FL. More specifically, we intend to find out whether there is evidence of gains in intelligibility and foreign accent after one year of CLIL instruction. Based on previous findings on L2 and FL speech production, we hypothesize that learners in the CLIL setting are likely to show a moderate improvement in overall intelligibility relative to their peers in the conventional foreign instruction setting, but they will show little or no improvement in foreign accentedness.

2. METHOD

2.1. Participants

Two groups of Catalan-Spanish EFL learners varying in target language exposure –a CLIL group and a FI group- as well as a group of native English speakers participated in the study. All participants were attending state-run secondary schools in five different locations in Majorca. These schools participated in the European Sections Programme. The participants' mean age ranged from 13 to 14 years. The two learner groups received different treatment in terms of amount and type of FL instruction. The CLIL learner group (n=64) had 3 hours a week of formal English instruction plus 3 extra hours of content instruction in English. A second group of learners (n=42) only received three hours a week of conventional FI. The native English speaker group (NE) was formed by students attending one of the schools that participated in the study (n=15) Two of the NE participants were born in England and had arrived in Majorca 5-6 years prior to the time of testing. The rest were early English-Spanish bilinguals whose father and/or mother spoke English at home.

2. 2. Speech materials

All participants were recorded in school premises using a digital recorder with a built-in microphone. The two learner groups were required to perform three oral tasks at two points in time separated by a one-year interval. The tasks included both controlled and extemporaneous speech: a role-play, an oral narrative and a reading aloud. This study reports on the speech materials from the reading aloud text only. Five excerpts were edited out from the resulting sound files using the *Praat* software and were later normalized for peak intensity for evaluation by native English listeners. The selected excerpts were: *There is very little vegetation, Life is hard and everything needs to be imported, It is the driest place in the world, We can water our plants every week* and *At night the sky is incredibly clear.*

2. 3. Procedure

The learners' and the native English speakers' utterances were grouped in five different blocs, one for each excerpt, and randomized for intelligibility and foreign accent assessment. The listening tests were run with the Multiple Force Choice application of the *Praat* software. Four native English listeners who were familiar with EFL speech participated as judges. They were all university staff and had been living in Spain 18 years on average (range 7-25). The four judges spoke non-regional varieties of standard British English. The decision of using expert judges instead of untrained or naïve listeners was based on previous studies that showed that experience with L2 speech often correlated with high levels of inter-judge reliability (see Munro 2008 for a review). Listeners were informed that they would hear an unspecified proportion of native and non-native speech samples, whose contents were provided beforehand. They were asked to perform two judgements: first, they had to assess intelligibility pressing one or three buttons labelled "easy" "mid" or "difficult" and immediately afterwards they had to rate foreign accent on a 1-to-5 scale (1= no FA, 5=strong FA). Based on prior work on L2 speech assessment, intelligibility was defined as "the degree to which a speaker's utterance is actually understood by a listener" (Munro 2008: 200). Each listener was tested individually in a quiet room and attended four testing sessions scheduled in four separate days, two for time 1 and two for time 2; two months elapsed between time 1 and time 2 assessments. Each testing session lasted about fifty minutes; there was a five-minute break between each test administration to prevent listeners' drop of attention.

3. RESULTS

3.1. Reliability

Intra-class correlations were performed to determine whether listeners agreed on the intelligibility and accentedness ratings of the learners' utterances. The intraclass correlation coefficients based on the averaged ratings of all four listeners were 0.83 for the intelligibility judgements and 0.91 for the foreign accent ratings, indicating a high level of inter-rater agreement.

3.2. Intelligibility

Table 1 shows the averaged percentage of times the five excerpts produced by two groups of Catalan-Spanish learners of English varying in amount of target language exposure were heard as "easy", "mid" or "difficult" to understand by four native English judges. Group differences were examined by means of chi-square analyses performed separately for time 1 and time 2. The effect of group was found to be significant for both time 1 ($\chi^2 = 282$ *d.f.=4 p=0.000)* and time 2 ($\chi^2 = 417$ *d.f.=4 p=0.000),* indicating that all three groups differed from one another in terms of speech intelligibility. Specifically, CLIL learners were understood with relative ease 51% of the times on average, as opposed to FI learners, who were easily understood only 31% of the times on average. No differences were found between time 1 and time 2. In fact, overall intelligibility across groups was judged to be lower at time 2. This might be an effect of listener's familiarity with the speech materials and the task, which resulted in more severe judgements even for the native English group.

Table 1: Average percentage of speech samples judged as "easy/mid/difficult to understand".

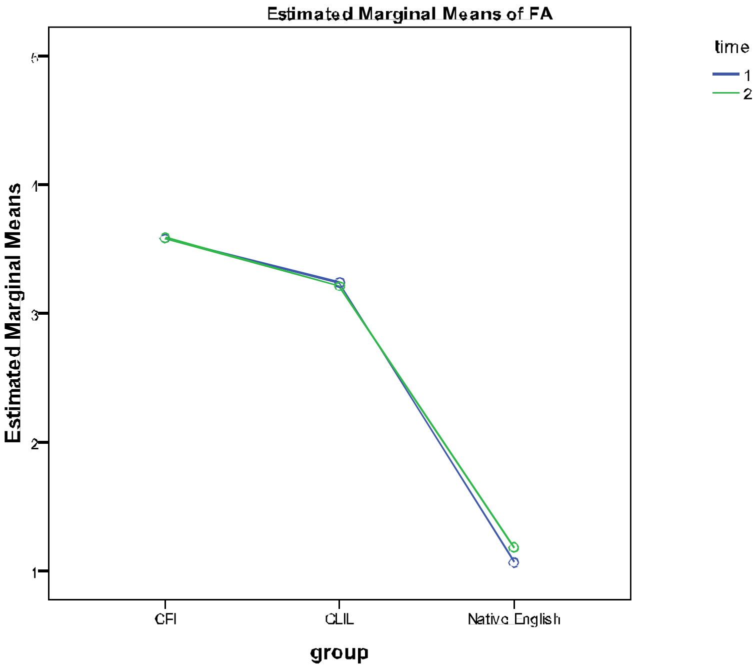
	Time 1			Time 2		
	easy	mid	difficult	easy	mid	difficult
FI	31.4	38.1	30.5	26.5	32.8	40.7
CLIL	51.5	32.7	15.8	45.6	34.7	19.8
NE	97.9	2.1	0	91.3	7.7	1.0

In order to test whether listeners' assessment of intelligibility varied depending on which excerpt they heard, an additional one-way ANOVA and Tukey's *post-hoc* tests were run. The main effect of excerpt was significant *(F (4,4040)=41.59 p=0.000)* showing that some excerpts were easier to understand than others. The pairwise comparisons indicated that, of the five excerpts presented to the listeners *We can water our plants every week* was the easiest to understand followed by *There is very little vegetation*. In turn, the excerpts *Life is hard and everything needs to be imported* and *At night the sky is incredibly clear* were judged to be more intelligible than *It is the driest place in the world*.

3.3. Foreign accent

The mean foreign accent ratings obtained by the CLIL and FI learner groups and the NE group are shown in Figure 1 below. A two-way repeated measures ANOVA with *group* as the *between-subject* factor and *time* as the within-subject factor was performed on the mean accented ratings. The main effect of group reached significance *(F (2, 4024)= 1453, p=0.000),* but neither the main effect of time *F (1, 4024)= 0.793, p=0.373)* nor the *group x time* interaction were significant *F (1, 4024)= 1.291, p=0.275).* Pairwise comparisons with Tukey's *post-hoc* analyses indicated that both groups of learners differed significantly from the native English group. The CLIL group and the FI group also differed significantly from one another.

Figure 1: Mean foreign accent ratings for CLIL, FI and NE groups.



4. DISCUSSION

This study examined the effect of content-based instruction on learners' intelligibility and accentedness in the target language. The data just reported showed that CLIL learners produced significanty more intelligible speech than their peers in the conventional FI setting. The CLIL group also differed significantly from the NE group and so did the FI group, indicating that both learner groups were far from approaching native-like standards. Foreign accent ratings were also found to be significant, revealing a substantial difference between the two groups of learners and the native English peers, and a smaller-size difference of accentedness ratings between CLIL and FI learners. No significant effects of time on speech intelligibility or foreign accent were found, suggesting that one year of CLIL instruction might be insufficient to improve learners' production in a foreign language.

The present findings are consistent with prior work on FL speech production by García Lecumberri and Gallardo (2003) and Fullana (2006). Both studies concluded that amount of formal instruction by itself does not diminish learners' degree of foreign accent, unless specific teaching methodology directed at improving learners' pronunciation skills is introduced in the foreign language school curriculum.

Some limitations of the present study should be acknowledged. Speech production in this study was assessed from read materials only. The possibility exists that lower intelligibility and higher foreign accent ratings in both the CLIL and FI groups were due to learners' lack of familiarity with some of the words that appeared in the reading passage and to orthographic interference. The significant differences on intelligibility found as a function of excerpt provide evidence for this. Previous studies examining intelligibility and foreign accent in L2 speech used both controlled and extemporaneous utterances (Munro and Derwing 1994; Munro and Derwing 1999; Piske et al. 2002). The participants in the present study were also recorded eliciting extemporaneous speech through a picture story task. Subsequent research should thus include ratings from less controlled speech to give a more accurate account of FL pronunciation in a formal setting. One methodological limitation that might arise is that CLIL and FI learners might substantially differ in terms of fluency or speaking rate, that is the number of words produced per minute, thus making a between-group comparison of extemporaneous speech problematic.

One further significant finding of the study was a tendency for judges to assess the intelligibility of time-2 utterances more severely than time-1 utterances. This trend was not only observed in the scores obtained by the two groups of learners. The native English utterances were also judged to be more intelligible at time 1 than at time 2, which is surprising because native English speakers were recorded just once and we used the same speech samples for the assessment of time-1 and time-2 speech. We speculate that the lower intelligibility ratings assigned to time-2 speech could be due to raters' familiarity with the speech materials, as it has been noted in earlier studies that assessed foreign accent in L2 speech (Flege and Fletcher 1992; Munro and Derwing 1994). Only two months had elapsed between the assessment of time-1 and time-2 speech so it is reasonable to hypothesize that, by the time time-2 speech was assessed, judges had become familiar with the contents of the read material. This might have facilitated their awareness of the most frequent pronunciation errors and, consequently, led them to assign harsher intelligibility ratings to time-2 utterances.

5. ACKNOWLEDGEMENTS

This research was funded by grant HUM2007-66053-C02-02/FILO from the Spanish Ministry of Science and Innovation and supported by the Council of Education of the Balearic Islands through the International Programmes Unit. The authors would like to express their gratitude to the students and teachers from the five secondary schools in Majorca that participated in this project: IES Alcúdia, IES Bendinat, IES Felanitx, IES Pau Casesnoves and IES Ramon Llull. We are also grateful to Celia Rosselló for assistance in the data analysis and to Pamela Butcher, Karen Jacob, Steven Pearson and Sarah Pegler for participating as judges.

6. REFERENCES

Best, C. T., Tyler, M. D. 2007. Nonnative and second-language speech perception. In Bohn, O.-S., Munro, M. J. (eds.), *Language Experience in Second Language Speech Learning. In Honor of James Emil Flege*. Amsterdam: John Benjamins. 13-34.

Bohn, O.-S., Flege, J. E. 1997. Perception and production of a new vowel category by adult second language learners. In James, A., Leather, J. (eds.), Second Language Speech: Structure and Process. Berlin: Mouton de Gruyter. 53-74.

Dalton-Puffer, C., Nikula, T. 2006. Pragmatics of content-based instruction: Teacher and student directives in Finnish and Austrian classrooms. *Applied Linguistics,* 27: 241-267.

Flege, J. E., Fletcher, K. L. 1992. Talker and listener effects on degree of perceived foreign accent. *Journal of the Acoustic Society of America,* 91, 1: 1990-1991.

Fullana, N. 2006. The development of English (FL) perception and production skills: Starting age and exposure effects. In Muñoz, C. (ed.), *Age and Rate of Foreign Language Learning*. Clevedon: Multilingual Matters. 41-60.

García Lecumberri, M. L., Gallardo, F. 2003. English FL sounds in school learners of different ages. In García-Mayo, M. P., García-Lecumberri, M.L. (eds.), *Age and the Acquisition of English as a Foreign Language*. Clevedon, England: Multilingual Matters. 115-135.

Lorenzo, F., Casal, S., Moore, P. To appear. The effects of CLIL in European Education: Key findings from the Andalusian Bilingual Sections Evaluation Project. *Applied Linguistics.* Advanced access published 20 Nov 2009, doi:10.1093/applin/amp041.

Moyer, A. 1999. Ultimate attainment in L2 phonology. The critical factors of age, motivation and instruction. *Studies in Second Language Acquisition,* 21: 81-108.
Munro, M. J. 2008. Foreign accent and speech intelligibility. In Hansen Edwards, J. G., Zampini, M. L. (eds.), *Phonology and Second Language Acquisition.* Amsterdam: John Benjamins. 193-218.
Munro, M. J., Derwing, T. M. 1994. Evaluations of foreign accent in extemporaneous and read material. *Language Testing* 11: 253-266.
Munro, M. J., Derwing, T. M. 1999. Foreign accent, comprehensibility and intelligibility in the speech of second language learners. *Language Learning,* 49, 1, 285-310.
Pérez-Vidal, C. 2009. The integration of content and language in the classroom: A European approach to education (the second time around). In Dafouz, E., Guerrini, M. (eds.), *CLIL across Educational Levels*. Madrid: Richmond. 3-16.
Pérez-Vidal, C. and Juan-Garau, M. 2009. To CLIL or not to CLIL: From bilingualism to multilingualism in Catalan/Spanish communities in Spain. In Lasagabaster, D., Ruiz de Zarobe, Y. (eds.), *CLIL in Spain: Implementation, Results and Teacher Training*. Cambridge: Cambridge Scholars Publishing. 115-138.
Piske, T. 2007. The influence of Flege's research in the foreign language classroom. In Bohn, O.-S., Munro, M. J. (eds.), *Second Language Learning and Linguistic Experience. In Honor of James Emil Flege.* Amsterdam: John Benjamins. 301-314.
Piske, T., Flege, J. E., MacKay, I. R. A. and Meador, D. 2002. The production of English vowels by fluent early and late Italian-English bilinguals. *Phonetica, 59*, 49-71

Accentual marking of information status in Dutch and French as foreign languages. Production and perception.

Laurent Rasier[1], Johanneke Caspers[2] & Vincent J. van Heuven[2]

[1]Institute for Language and Communication, Université catholique de Louvain – FNRS, [2]Leiden University Centre for Linguistics
Laurent.Rasier@uclouvain.be, J.Caspers@hum.leidenuniv.nl, V.J.J.P.Van.Heuven@hum.leidenuniv.nl

ABSTRACT

This paper explores (non-)native speakers' use and perception of accent as a cue to information status in Dutch and French as foreign languages. On the production side, French-speaking learners of Dutch tend not to vary the accent distribution according to the information status of the words in the utterance and generally do not deaccent contextually given information in a nativelike fashion, whereas Dutch-speaking learners of French, though on the whole closer to the L1 French target, tend to overuse the French "focus accent" in such a way that their accent patterns more closely reflect the information status of the words than the patterns produced by L1 French speakers. These results are explained in terms of the "markedness relationships" between Dutch and French. A pilot perception experiment then investigates (non-)native listeners' perception of prominence relationships in L1 and L2 Dutch. The data show that the accented elements produced by L2 speakers of Dutch are perceptually less prominent than accented elements produced by L1 speakers, that this reduced prominence makes it more difficult for L1 listeners to correctly identify the intended patterns, and that non-native listeners generally find it difficult to identify the intended accent patterns in their L2.

Keywords: Prosody, accentuation, L2 acquisition, Dutch, French.

1. INTRODUCTION

The last two decades have witnessed a fast-growing interest in the prosodic features of speech (e.g. stress, accent, intonation, rhythm; see Fox 2000 for an historical survey). Yet, in spite of the rapid development and increasing success of prosodic analysis, current knowledge about non-native prosody is still very limited (Chun 2002, Trouvain and Gut 2007, Gut 2009). This paper examines accentuation, which is defined as an utterance-level phenomenon related to the focus structure of the utterance (Cutler 1984, Van Heuven 2002), in non-native speech production and perception. It focuses on Dutch spoken by francophone L2 learners (L2 Dutch) and French spoken by Dutch-speaking L2 learners (L2 French). Section 2 discusses previous research on accentuation in L2 speech. Section 3 focuses on the use of (de)accentuation as a cue to information status in L2 Dutch and L2 French. Section 4 reports on a pilot perception experiment in L2 Dutch. Section 5 draws some conclusions about accentuation in non-native speech production and perception.

2. ACCENTUATION IN NON-NATIVE SPEECH

As a prosodic feature, accent has generated considerable debate (Fox 2000, Ladd 1996). In the context of L2 acquisition, by contrast, current knowledge about accentuation in a second/foreign language is still limited (Gut 2009: 221). Several studies report a general tendency for L2 speakers to produce more pitch accents than L1 speakers do (Grosser 1997, Juffs 1990). Gut (2000), in contrast, does not find any overuse in comparison with native speech, but observes differences in accent placement (see also Adams 1979, Backman 1979, Grosser 1997, Ramirez Verdugo 2002). Grosser (1997) finds for Austrian learners of English that information, which does not normally require any pitch accent, is still produced as accented in 45 to 90% of all cases. Exploring Spanish English Ramirez Verdugo (2002) also observes that contextually given information often gets accented and that the main pitch accent in the utterance tends to be realized on the utterance-final element irrespective of its information status or word class. Barlow (1998), however, does not find any accent placement differences between L1 and L2 speakers of English with various L1 backgrounds.

Rather, Barlow finds phonetic differences in accent production. L2 speakers underuse melodic and temporal cues to accent and rely on higher intensity values to mark utterance elements as accented. Phonetic differences between L1 and L2 accent production are also reported in Grosser (1997), Kelm (1987), Ramirez Verdugo (2002) and Wennerstrom (1994). Ramirez Verdugo (2002) shows that Spanish L2 speakers of English use the same type of melodic movement to signal new and given information, whereas L1 speakers use distinct pitch movements. Non-native pitch movements also appear to have a smaller excursion size than native ones. Wennerstrom (1994) examines the pitch height of new and given information in L1 and L2 English and finds that L1 speakers produce higher pitch on new information than on given information, whereas L2 speakers use the same pitch height for all elements, regardless of their information status.

The studies reviewed so far examine English as a second/foreign language. In many other non-native varieties, accentuation still constitutes uncharted ground. This is e.g. the case in L2 Dutch and L2 French.

3. ACCENTUAL MARKING OF INFORMATION STATUS IN L2 DUTCH AND L2 FRENCH

It has been claimed that local pitch movements (e.g. pitch accents) are universally used to signal contextually new or important information (e.g. Bolinger 1989). However, there exist cross-linguistic differences in the extent to which accent patterns constitute a cue to focus. Vallduvi (1991) makes a distinction between plastic and non-plastic languages. In plastic languages, prominence patterns can be manipulated in such a way that they reflect the focus structure of the utterance. In English e.g. accent patterns can be modified so that the "givenness" of the normally accented word(s) is reflected by means of deaccentuation. Non-plastic languages, in contrast, are characterized by "fixed" accentuation so that syntactic means must be used to achieve the same effect as (de)accentuation in plastic languages. Non-plastic languages are a.o. Catalan (Vallduvi 1991), Italian (Swerts et.al. 2002) and Romanian (Swerts 2007). According to Ladd (1996: 294), French is also non-plastic as "[it] seem[s] to be most resistant to moving accent out of phrase-final or sentence-final position", whereas Dutch is a plastic language (Krahmer an Swerts 2001, Swerts et.al. 2002).

3.1. Method

Rasier (2006) and Rasier and Hiligsmann (2007, 2009) have investigated accent placement in L1/L2 Dutch and French. The data were collected in a picture description task in which the participants had to describe geometrical figures (a circle, a triangle, a star, a square) appearing in different colours (red, yellow, blue, green) on a computer screen. Manipulating the presentation order of the stimuli made it possible to vary the information status of the words, resulting in four conditions on the adjective-noun pairs: New/New (NN), Given/Contrastive (GC), Contrastive/Given (CG), Contrastive/Contrastive (CC). A property is "new" when it has not yet been used in the discourse. It is called "given" when it has already been mentioned in the preceding context. A property that differs from the immediately preceding utterance is called "contrastive".

The subjects were 20 French-speaking learners of Dutch and 20 Dutch speaking-learners of French who had been learning the L2 for 8-10 years in a school setting in Belgium and in the Netherlands. All of them were second year students of economics or business communication at a Belgian or Dutch university. Their ages varied between 19-20 years. They took the test in their L1 and L2. The data were transcribed by the first author of this paper with the aid of two phonetically trained judges who were native speakers of Dutch and French, respectively. The transcriptions were then compared and discussed until agreement was reached.

3.2. Results

L1 Dutch is characterized by a high association between information status and accentuation (Cramer's V = 0.878), which indicates that the accent distribution on the elicited NPs generally reflects the information status of the words in the utterance. In the NN- and CC-conditions, both the adjective and the noun are accented in 90% and 95% of all cases, respectively. In the GC-condition, a single accent on the noun (i.e. the contrastive element) can be found in 50% of the cases, whereas both the (contextually given) adjective (in prenuclear position) and the noun are accented in 45% of the cases. In the CG-condition, a single accent on the C-element is found in 98% of the cases, which shows that L1 Dutch does not allow the accentuation of the noun in postnuclear position in the CG-condition. This is in accordance with Krahmer and Swerts (2001) and Swerts et.al. (2002) who show that in Dutch nuclear accents can sometimes be preceded by prenuclear ones, even in contrastive settings, whereas they are rarely followed by postnuclear ones.

A lower association between information status and accentuation can be observed in L1 French than in L1 Dutch (Cramer's V = 0.615), pointing to a lower degree of plasticity in L1 French. Indeed, the phrase-final element is most frequently accented in the four experimental conditions (NN: 100%, GC: 100%, CG: 70%, CC: 97%). However, it is generally preceded by another accent on the first syllable of the first lexical item (i.e. "bridge accent"; NN: 60%, GC: 60%, CG: 57%, CC: 83%) or on the indefinite article (i.e. "extended bridge accent; NN: 30%, GC: 30%, CG: 0%, CC: 7%). Interestingly, there are cases in the CG-condition where the accent is moved out of phrase-final position so that only the contrastive piece of information in non-final position is accented (i.e. "focus accent"; CG: 30%), but this is less often the case than in L1 Dutch. So, unlike Catalan or Italian, French sometimes allows given information to be deaccented.

Looking at non-native speech, it appears that L2 Dutch speakers make significantly more accent placement errors than L2 French speakers (47% vs. 78% contextually adequate accent patterns; $\chi^2 = 36.0$; df = 2; $p < 0.001$). This suggests that it is easier to shift from a plastic to a non-plastic language (i.e. Dutch → French) than the other way around (i.e. French → Dutch). L2 Dutch is also characterized by a lower association between information status and accentuation than L1 Dutch (Cramer's V = 0.600 vs. 0.878), which means that L2 Dutch speakers do not prosodically distinguish between given and new/contrastive information to the same extent as L1 Dutch speakers. Indeed, L2 Dutch speakers in most cases accentuate both the adjective and the noun notwithstanding their information status (NN: 60%, GC: 70%, CG: 50%, CC: 83%), which results in significantly lower deaccentuation rates in L2 Dutch than in L1 Dutch, i.e. 5% vs. 50% in the GC-condition ($\chi^2 = 16.2$; df = 1; $p < 0,001$) and 35% vs. 95% in the CG-condition ($\chi^2 = 28.1$; df = 1; $p < 0,001$). L1 and L2 French, by contrast, have quite close association coefficients between information status and accentuation (Cramer's V = 0.615 in L1 French vs. 0.632 in L2 French). However, the fact that the association measure is slightly higher in L2 French than in L1 French suggests that the accent patterns produced by L2 French speakers more closely reflect the information status of the words in the utterance than the patterns used by L1 French speakers. L2 French speakers use the "bridge accent" pattern in a nativelike manner in the NN- and CC-conditions. This pattern is also the most frequent one in the GC-condition, but L2 French speakers also produce a single (focus) accent on the phrase-final element more often than L1 French speakers (18% in L2 French vs. 5% in L1 French). The tendency to produce a single (focus) accent on the contrastive piece of information, leading to the deaccentuation of the contextually given element, is even clearer in the CG-condition where a single accent is produced on the contrastive piece of information in 56% of the cases (vs. 30% in L1 French), while L1 French speakers still use the "bridge accent" pattern in the majority of the cases (67% of the cases in L1 French vs. 38% in L2 French).

3.3. Summary and discussion

The L1 Dutch data support the view that Dutch is a plastic language as accentuation can be varied in such a way that it reflects the information status of the words in the utterance. However, structural constraints also apply when it comes to determining the exact position of the accent within the focussed constituent (Gussenhoven 1984, Kruyt 1985). So, accentuation in Dutch can be characterized as {pragmatic > structural}. In French, in contrast, accentuation is mainly governed structurally by the surface position in the utterance, although at times pragmatic reasons can lead to accent being moved to a constituent in non-final position. French accentuation can therefore be described as {structural > pragmatic}. This makes French different from other Romance languages such as Italian or Catalan where accent placement is governed structurally ({structural}), whereas no language has been reported where accentuation is only governed pragmatically.

Typologically, these observations imply that a language with structural accentuation rules does not necessarily have pragmatic accentuation rules as well, while the reverse is not true. In terms of markedness, a structure A is typologically more marked relative to another structure B if every language that has A also has B, but every language that has B does not necessarily have A (Gundel et al. 1986: 108). In this sense, then, pragmatic accentuation rules are more marked than structural ones. In this respect, Eckman's *Markedness Differential Hypothesis* (Eckman 1984, 2008) predicts that marked structures are more difficult to learn than unmarked ones. As pragmatic accentuation rules are more marked than structural rules and pragmatic rules have a dominant position in Dutch – but not in French – one can explain why it is easier to shift from a plastic to a non-plastic language (Dutch → French, i.e. marked → unmarked) than the other way around (French → Dutch, i.e. unmarked → marked). The diverging degrees of markedness of structural and pragmatic accentuation rules also explain why L2 French speakers produce the French "bridge accent" pattern, which is a largely structurally governed – and therefore unmarked – accent pattern, in a nativelike fashion, whereas L2 Dutch speakers generally do not succeed in applying the Dutch deaccentuation rule in a contextually adequate way in spite of the fact that this rule also exists in French (cf. "focus accent"). Deaccentuation is a marked rule in French and marked L1 rules are less likely to be transferred to the learners' L2 than unmarked ones. The French "bridge accent", by contrast, is an unmarked pattern so that L1 transfer and the surface similarity between the French "bridge accent" and the Dutch "double accent" can explain why this pattern is used in L2 Dutch regardless of the information status of the words in the utterance. As there is evidence that accenting given information can delay sentence comprehension in Dutch (Terken and Nooteboom 1987) and that sentence comprehension is sensitive to accentual structure in native listeners (Akker and Cutler 2001: 92), French-speaking learners' deaccentuation errors could be detrimental to Dutch listeners' processing of L2 speech. For Dutch-speaking learners of French, however, deaccenting given information is less marked than for L1 French speakers, and this can explain why it is more often deaccented in L2 French than in L1 French. To our knowledge, the effect of (de)accentuation errors on sentence comprehension in French has not been studied so far.

4. ACCENT PERCEPTION IN L2 DUTCH: A PILOT STUDY

On the production side, L1 Dutch and L1 French prosodically encode information status in different ways (i.c. in terms of accent distribution) and the differences between the two languages influence L2 speakers' accent placement strategies. As accent is an important cue to focus in Dutch (but less so in French), the question arises whether accentuation differences can be perceived by non-native Dutch listeners (Watanabe 1988) and to what extent they can process prosodic information for semantic structure in a nativelike fashion (Akker & Cutler 2001). This section examines pilot data on accent perception in L1 and L2 Dutch.

4.1. Method

Starting from the production data, a perception experiment was carried out in a web-based environment. The research questions were how well accentuation differences can be perceived in L1 and L2 Dutch, which accent patterns are the most/least difficult to perceive by (non-)native listeners, whether systematic "patterns of substitution" can be identified, and what these are due to. The stimuli for the perception test were selected from the production data according to the sex of the speaker (female voices), her language background (L1 and L2 speakers in equal proportions), the speech quality (no disfluencies), and the type of accent pattern (single accent on the adjective, single accent on the noun, accent on both the adjective and the noun). The three accent patterns appear in the materials in equal proportions. Each stimulus was produced with an L%- and an H%-boundary tone and was presented in the experiment twice. Four listening conditions occur in the test: L1 Dutch listeners/L1 Dutch speakers, L1 Dutch listeners/L2 Dutch speakers, L2 Dutch listeners/L1 Dutch speakers, L2 Dutch listeners/L2 Dutch speakers[1].

The listeners were 10 native speakers of Dutch and 10 francophone learners of Dutch who had been learning the L2 for 8-10 years in a school setting in Belgium. They were second year students of economics or business communication at a Belgian (L2 listeners) or Dutch university (L1 listeners). Their ages varied between 19-20 years. The test included 10 training items and 72 test items. The task consisted in evaluating the relative degree of prominence of both the adjective and the noun on a 10-point scale. The participants could listen to the same item twice before responding. Prominence is regarded as the perceptual consequence of "accentedness", which leads to the assumption that listeners' prominence judgements reflect the accent patterns they hear in speech. Therefore the accent patterns perceived by native and non-native listeners were derived from their prominence judgements. Elements with a prominence level equal to or above 5 points are considered as "accented", whereas elements with a prominence level below 5 points are regarded as "unaccented". Also there had to be a difference of at least 2 points between the adjective and the noun for one of the two elements to be regarded as more prominent than the other.

4.2. Results

Native listeners identify the expected accent pattern in 61% of the cases when they listen to L1 Dutch speakers and in 52% of the cases when they listen to L2 speakers. The difference between the two listening conditions is not significant ($\chi^2 = 3.4$; df = 1; $p = 0.065$). Non-native listeners, by contrast, identify the expected accent pattern in native speech in 40% of the cases when they listen to L1 Dutch (vs. 61% by L1 listeners; $\chi^2 = 22.7$; df = 1; $p < 0.001$), whereas they identify the expected pattern in 45% of the cases when they listen to Dutch utterances produced by other French-speaking learners. The difference between the two listening conditions does not reach significance either ($\chi^2 = 1.8$; df = 1; $p = 0.183$).

Listening to L2 speakers, L1 Dutch listeners perceive the expected accent pattern in ca. 82% of the cases in the "two accents"-condition (i.e. accent on the adjective and the noun), whereas they perceive a single accent on the noun in 26% of the cases in the "single accent on the noun"-condition and a single accent on the adjective in just 21% of the cases in the "single accent on the adjective"-condition ($\chi^2 = 71.3$; df = 2; $p < 0.001$). Instead of a single accent on the adjective or the noun, native listeners perceive an accent on *both* elements in 96% of the cases in the "single accent on the noun"-condition and in 71% of the cases in the "single accent on the adjective"-condition. In the other cases, they perceive either a single accent on the adjective (i.e. 4% of the substitutions in the "single accent on the noun"-condition) or a single accent on the noun (i.e. 29% of the substitutions in the "single accent on the adjective"-condition). Substitutions are significantly more frequent (ca.71%) in utterances ending on an L%-boundary tone than in utterances ending on an H%-boundary tone (ca. 29%; $\chi^2 = 26.6$; df = 1; $p < 0.001$).

Listening to L1 speakers, non-native listeners perceive the expected accent pattern in ca. 61% of the cases in the "two accents conditions" (i.e. accent on the adjective and the noun), whereas they perceive the expected pattern in 39% of the cases in the "single accent on the adjective"-condition (i.c. single accent on the adjective) and in only 15% of the cases in the "single accent on the noun"-condition (i.c. single accent on the noun) ($\chi^2 = 26.2$; df = 2; $p < 0.001$). The pattern with an accent on both the adjective and the noun accounts for 66% and 74% of the cases of substitution in the "single accent on the adjective"- and the "single accent on the noun"-condition. In the other cases, non-native listeners perceive a single accent on the noun in the "single accent on the adjective"-condition (34%) and a single accent on the adjective in the "single accent on the noun"-condition (26%). Most substitutions occur in utterances ending on an H%-boundary tone (54%), whereas non-native listeners perceive the expected pattern in ca. 65% of the cases when the utterance ends on a L%-boundary tone ($\chi^2 = 7.8$; df = 1; $p < 0.01$).

4.3. Summary and discussion

Native listeners identify the expected accent pattern more often when they listen to native speakers of Dutch than when they listen to non-native speakers. Though not significant, the observed difference is in accordance with the view that foreign-accented speech is generally less intelligible to native listeners than native speech (Munro & Derwing 1995, Van Wijngaarden 2001). Non-native listeners, by contrast, identify the expected accent pattern more often when they listen to non-native speakers (i.c. with the same L1) than when they listen to L1 speakers of the target language. Though not significant, the slight increase between the two listening conditions suggests nevertheless that non-native listeners benefit from listening to other L2 speakers (interlanguage speech intelligibility benefit; Bent & Bradlow 2003), while this is not the case for native listeners. Also, systematic patterns of substitution can be observed in native and non-native listeners.

In native listening (i.e. L1 Dutch listeners/L2 Dutch speakers), most substitutions consist of an accent on both the adjective and the noun where a single accent on either the adjective or the noun is expected. Such patterns mainly occur when the utterance ends on an L%-boundary tone. A possible explanation for such cases of perceptual substitution is that L2 speakers do not prosodically mark the distinction between accented and unaccented elements as clearly as L1 speakers do. Indeed, according to native listeners, there is a significant prominence difference between accented and unaccented information in L1 Dutch speech (unaccented = 3.64, StD = 1.25; accented = 6.05, StD = 1.62; $t = -15.0$; df = 131.6; $p < 0.01$), but not in L2

Dutch speech where the perceptual difference between accented and unaccented elements is less clear-cut (unaccented = 3.82, StD = 1.25; accented = 5.84, StD = 1.63; $t = -10.7$; $df = 86.4$; $p = 0.06$).

In non-native listening (i.e. L2 Dutch listeners/L1 Dutch speakers), most substitutions consist of an accent on both the adjective and the noun in cases where a single accent on either the adjective or the noun is expected. Non-native listeners also quite frequently perceive a single accent on the noun in cases where a single accent on the adjective is expected. Contrary to L1 listening, most substitutions can be observed in utterances ending on an H%-boundary tone. This suggests that the H%-boundary tone is actually interpreted as an (additional) phrase-final accent by non-native listeners. These results can be indicative of L1 influence on francophone learners' perception of accentuation in Dutch. Indeed, phrase-final elements are (nearly) always accented in French so that accent is used as a demarcative (i.e. boundary-marking) device (Lacheret-Dujour and Beaugendre 1999). Moreover, the "two accents"-pattern is also the pattern which is most often used by francophone learners in speech production (see § 3.2).

5. CONCLUSIONS

This paper has undertaken a comparative study of prominence patterns in both native and non-native speech production and perception. On the production side, it has investigated to what extent accent is used as a cue to information status in Dutch and French spoken by native and non-native speakers. On the perception side, it has examined native and non-native listeners' perception of prominence patterns in L1 and L2 Dutch.

The production data reveal that Dutch and French are quite different in the way they encode information status in accent patterns. While L1 Dutch speakers vary their accent patterns according to the information status of the words, L1 French speakers generally accentuate both the adjective and the noun ("bridge accent") regardless of their information status in the utterance, although they sometimes produce a single accent on contextually contrastive information ("focus accent"). The L2 production data show that accent placement in non-native speech is influenced by L1 characteristics. The Dutch-speaking learners of French, though on the whole closer to the L1 French target, overuse the French "focus accent" in such a way that their accent patterns more closely reflect the information status of the words than the patterns produced by L1 French speakers. The francophone learners of Dutch, on the other hand, generally produce the same accent pattern (accent on the adjective and the noun) notwithstanding the information status of the words, and therefore do not deaccent contextually given information in a nativelike fashion. These results can be explained in terms of the "markedness relationships" between Dutch and French.

The perception study shows that native listeners identify the expected accent pattern more often when they listen to L1 Dutch speakers than when they listen to L2 Dutch speakers, which suggests that non-native speech, i.e. foreign-accented prosody, is less intelligible to native listeners than native speech and prosody. Non-native listeners, however, get better identification results when they listen to non-native speakers than when they listen to native speakers of the target language, i.e. interlanguage speech intelligibility benefit. In the two listening directions (i.e. L1 listeners/L2 speakers and L2 listeners/L1 speakers), native and non-native listeners frequently perceive an accent on both the adjective and the noun instead of a single accent on either the adjective or the noun. Native listeners' perceptual substitutions, which can mainly be observed in utterances ending on an L%-boundary tone, were argued to emerge from the fact that, according to L1 listeners, the difference between accented and unaccented elements in terms of prominence is less clear-cut in L2 Dutch than in L1 Dutch. Acoustic analyses are needed to determine which properties of the non-native stimuli make them perceptually less prominent than the native stimuli. Non-native listeners' substitutions, by contrast, mainly occur in utterances ending on an H%-boundary tone, which suggests that they interpret the H%-boundary tone as an (additional) phrase-final accent. The substitution patterns that were observed in non-native listeners can be related to properties of their L1, thereby pointing to L1 influence at the perception level. Future work will examine accent perception in French by native and non-native listeners (i.e. Dutch-learners of French) in order to get a better understanding of the way prominence patterns are perceived in an L2 as well as of the role the learners' L1 plays in it.

6. REFERENCES

Adams, C. 1979. *English Speech Rhythm and the Foreign Learner*. The Hague: Mouton.

Akker, E., Cutler, A. 2001. Prosodic Cues to Semantic Structure in Native and Nonnative Listening. *Bilingualism: Language and Cognition* 6(2). 81–96.
Backman, N. 1979. Intonation Errors in Second-language Pronunciation of Eight Spanish-speaking Adults Learning English. *Interlanguage Studies Bulletin* 4. 239–265.
Barlow, J.S. 1998. *Intonation and Second Language Acquisition. A Study of the Acquisition of English Intonation by Speakers of Other Languages*. PhD dissertation, University of Hull.
Bent, T., Bradlow, A. 2003. The interlanguage speech intelligibility benefit. *Journal of the Acoustical Society of America* 114. 1600–1610.
Bolinger, D.L. 1989. *Intonation and its Uses: Melody in Grammar and Discourse*. London: Edward Arnold.
Chun, D.M. 2002. *Discourse intonation in L2. From Theory and Research to Practice*, Amsterdam/Philadelphia: John Benjamins.
Cutler, A. 1984. Stress and accent in language production and understanding. In: Gibbon, D., Richter, H. (eds.), *Intonation, Accent and Rhythm. Studies in Discourse Phonology*. Berlin: De Gruyter. 77–90.
Eckman, F. 1984. Universals, Typologies and Interlanguages. In Rutherford, W. (ed.), *Language Universals and Second Language Acquisition*. Philadelphia: John Benjamins Publishing Company. 79–105.
Eckman, F. 2008. Typological Markedness and Second Language Phonology. In: Hansen-Edwards, J.G., Zampini, M.L. (eds), *Phonology and Second Language Acquisition*. John Benjamins Publishing Company: Amsterdam/Philadelphia. 95–115.
Fox, A. 2000. *Prosodic Features and Prosodic Structure. The Phonology of Suprasegmentals*. Oxford: Oxford University Press.
Grosser, W. 1993. Aspects of intonation L2 acquisition. In: Kettemann, B., Wieden, W. (eds.), *Current Issues in European Second Language Acquisition research*. Tübingen: Gunter Narr Verlag. 81–94.
Gundel, J., Houlihan, K., Sanders, G. 1986. Markedness Distribution in Phonology and Syntax. In Eckman, F., Moravcsik, E., Wirth, J. (eds.), *Markedness*. New York: Plenum Press. 107–138.
Gussenhoven, C. 1984. *On the Grammar and Semantics of Sentence Accents*. Dordrecht: Foris Publications.
Gut, U. 2000. On the Acquisition of Rhythmic Structure. *Proceedings of New Sounds 2000, Amsterdam*.148–154
Gut, U. 2009. *Non-native Speech. A Corpus-based Analysis of Phonological and Phonetic Properties of L2 English and German*. Frankfurt am Main: Peter Lang.
Heuven, V. van. 2002. *Boven de klanken/Beyond the Segments*. Amsterdam: Koninklijke Nederlandse Academie van Wetenschappen.
Juffs, A. (1990). Tone, Syllable Structure and Interlanguage Phonology. Chinese Learners' Stress Errors. *International Review of Applied Linguistics in Language Teaching* 28, 99–117.
Kelm, O. 1987. An Acoustic Study of the Differences of Contrastive Emphasis between Native and Non-native Spanish Speakers. *Hispania* 70. 627–633.
Krahmer, E., Swerts, M. 2001. On the alleged existence of contrastive accent. *Speech communication* 34(4). 391–405.
Kruyt, J.G. 1985. *Accents from Speakers to Listeners. An Experimental Study of the Production and Perception of Accent Patterns in Dutch*. PhD dissertation, Leiden University.
Lacheret-Dujour, A., Beaugendre, F. 1999. *La prosodie du français*. Paris: CNRS Editions.
Ladd, D.R. 1996. *Intonational Phonology*. Cambridge: Cambridge University Press.
Munro, M. J., Derwing, T. M., 1995. Foreign accent, comprehensibility and intelligibility in the speech of second language learners. *Language Learning* 45, 73–97.
Ramirez Verdugo, D. 2002. Non-native Interlanguage Intonation Systems: A Study Based on a Computerized Corpus of Spanish Learners of English. *ICAME Journal* 26. 115–132.
Rasier, L. 2006. *Prosodie en vreemdetaalverwerving. Accentdistributie in het Frans en Nederlands als vreemde taal*. PhD dissertation, Université catholique de Louvain.
Rasier, L. 2008. Contrastief prosodieonderzoek Nederlands-Frans. Een contrastief-typologische kijk op de accentuering. *Verslagen en Mededelingen van de Koninklijke Academie voor Nederlandse Taal- en Letterkunde* 118(1). 49–67.
Rasier, L., Hiligsmann, Ph. 2009. Exploring the L1-L2 Relationship in the L2 Acquisition of Prosody. *Proceedings of First and Second Languages. Exploring the Relationship in Pedagogy Related Context*.
Swerts, M. 2007. Contrast and Accent in Dutch and Romanian. *Journal of Phonetics* 35. 380–397
Swerts, M., Krahmer, E., Avesani, C. 2002. Prosodic Marking of Information Status in Dutch and Italian: a Comparative Analysis, *Journal of Phonetics* 30(4). 629–654.
Terken, J., Nooteboom, S. 1987. Opposite Effects of Accentuation and Deaccentuation on Verification Latencies for 'Given' and 'New' Information. *Language and Cognitive Processes* 2. 145–163.
Trouvain, J., Gut, U. (eds.). 2007. *Non-native prosody. Phonetic Description and Teaching Practice*, Berlin: Mouton de Gruyter.
Vallduvi, E. 1991. *The Informational Component*. New York: Garland.
Watanabe, K. 1988. Sentence Stress Perception by Japanese Students. *Journal of Phonetics* 16. 181–186.
Wennerstrom, A. 1998. Intonation as Cohesion in Academic Discourse. *Studies of Second Language Acquisition* 20. 1–25.
Wijngaarden, S. J. van. 2001. Intelligibility of native and non-native Dutch speech. *Speech Communication* 35. 103–113.

NOTES

[1] This paper focusses on native listeners' perception of accent patterns in non-native speech (i.e. L1 listeners/L2 speakers) and on non-native listeners' perception of accentuation in native speech (i.e. L2 listeners/L1 speakers). The other conditions (i.e. L1 listeners/L1 speakers and L2 listeners/L2 speakers) will not be discussed in details.

Production and perception of vowel /æ/ by Polish learners of English

Arkadiusz Rojczyk

University of Silesia
arkadiusz.rojczyk@us.edu.pl

ABSTRACT

The study aims to determine the production and perception of /æ/ by Polish learners of English relative to neighbouring English /e/ and /ʌ/, and Polish /ɛ/ and /a/. The results show that, in production, /æ/ is partly subsumed by Polish /a/ and shares a significant portion of acoustic space with /ʌ/. In perception, the learners do not rely on spectral properties differentiating /æ/ and /ʌ/, which supports the claim that they are both subsumed by one native category. Rather, a significant role of duration has been observed as a contrasting cue both in production and perception. Polish learners produce longer durations for /æ/ and are sensitive to longer values in identifying /æ/ relative to /ʌ/.

Keywords: vowels, duration, production, perception

1. INTRODUCTION

The English vowel /æ/ has been found to be one of the more difficult vowels to acquire for non-native learners of English (eg., Bohn and Flege 1997, Flege et al. 1997, Strange et al. 2001). Its production reflects intensity of a foreign accent, in that instances of /æ/ produced by learners with mild accents are more intelligible than /æ/s produced by learners with stronger foreign accents (Flege 1992, Major 1987). The same correlation is also observed between production accuracy of /æ/ and English-language experience (Bohn and Flege 1992).

Polish learners of English, whose native language makes use of only six vowels (Jassem 2003), have difficulties learning a densely crowded vowel system in English. The cross-linguistic comparison reveals that a mid-low front /æ/ does not directly correspond to any vowel category in Polish. Its nearest neighbours in Polish are font mid /ɛ/ and open front /a/. Previous research based on auditory observations showed that English /æ/ is equally likely to be assimilated by Polish /ɛ/ and /a/ and that the actual choice of either alternative in the learners' pronunciation may depend on personal preferences (Sobkowiak 2003) or be conditioned by spelling convention of a given word (Gonet et al. 2010).

The Speech Learning Model (Flege 1995) assumes that the phonetic similarities and dissimilarities of L1 and L2 segments will influence the degree of ultimate success in producing and perceiving non-native sounds. Phonetic similarity and dissimilarity are defined in terms of the articulatory and acoustic characteristics of the linguistically relevant speech sounds. The actual attainment of native-like production and perception of given L2 sounds is assessed relative to the phonetic distance between L2 and L1 segments. Learners are hypothesised to be less successful in learning L2 sounds that are perceived as similar to L1 sounds because the similarity will block the formation of a new phonetic category by means of the equivalence classification. In contrast, L2 sounds perceived as new or different from L1 categories will motivate the learners to develop new L2 categories. The application of the SLM metrics to establish the similarity scale of English /æ/ for Polish learners is not unproblematic. Using the criterion of articulatory and acoustic properties, this vowel may be referred to a 'new' for Polish learners because it is located in the front low area which is unexploited in the vowel system in Polish. This prediction is weakened, however, by observations that Polish learners assimilate /æ/ in fairly equal proportions to both Polish /ɛ/ and /a/ (Gonet et al. 2010, Sobkowiak 2003). This pattern may indicate that both /ɛ/ and /a/ have a commensurate similarity index with /æ/ and that this equivalence classification will impede the formation of a native-like L2 category.

The Perceptual Assimilation Model (Best 1995, Best and Tyler 2007) predicts that an L2 contrast will be discriminated relatively poorly if it is mapped onto one L1 segment. By contrast, cases in which an L2 contrast is mapped onto two L1 sounds will yield best results, even if there is no single category corresponding to either of the L2 categories. In the course of exposure, the listeners will disentangle the accommodated L2 contrast from two L1 segments and will be likely to create native-like perceptual specification. An application of this model suggests that Polish learners will not have difficulties with correct perception and production of an English /æ/-/e/ contrast because it will be assimilated by an /a/-/ɛ/ contrast in Polish. On the other hand, contrasting /æ/-/ʌ/ will be difficult because it will be assimilated by one /a/ category in the learners' native language.

Another issue concerning the Polish learners' production and perception of /æ/ is its duration variability as a compensating contrasting mechanism. It has been observed that /æ/ is very often lengthened relative to other lax vowels (Wells 1982) and some studies even classify it as a long vowel (Peterson and Lehiste 1960, Strange et al. 2001). Previous perception research has demonstrated that in the absence of sensitivity to acoustic differences between L2 vowels, learners may resort to use durational values as a compensating mechanism, even though vowel duration itself may only be secondary or even redundant for native speakers (Bohn 1995, Escudero et al. 2009 for references). This pattern has also been found for an /æ/-/e/ contrast (Bohn and Flege 1990). Although the exact reasons for this reweighting of cues are still unclear (Escudero et al. 2009), the previous results may be taken to suggest that Polish learners will exploit vowel duration in signalling the distinction between /æ/ and two neighbouring English vowels /e/ and /ʌ/ in production and/or use duration as a robust perceptual cue in identification of /æ/. This hypothesis is strengthened by the research that demonstrated that Polish learners are sensitive to vowel duration as an overriding cue in distinguishing between English /iː/ and /ɪ/ (Bogacka 2004).

2. CURRENT STUDY

In the current, study we set out to determine the following issues: (1) the extent of overlap between spectral properties of /æ/ and both English /e/, /ʌ/ and Polish /ɛ/, /a/; (2) the relationship between duration of /æ/ and its spectral properties relative to neighbouring vowels in English in the learners' production and perception.

2.1. Experiment 1: Production

2.1.1. Participants

A total of 43 subjects participated in the study: 31 females and 12 males. They were recruited from third-year students at the Institute of English, University of Silesia. This selection guaranteed a uniform level of proficiency due to a regular administration of various tests in use of English. They considered themselves to be advanced speakers of English with no difficulties in communicating with native speakers. They had been given 120 hours of explicit phonetic training in English pronunciation in the first and second year. They ranged between 21 to 29 years of age (Mean: 21.3, Median: 21). Eighteen participants reported to have spent more than a month in an English-speaking country. They all volunteered and were not paid for their participation. None of the subject reported any speech or hearing disorders.

2.1.2. Materials

All vowels were embedded in a /bVt/ context. In two separate sessions, both Polish and English vowels were recorded. All Polish vowels were needed as corner landmarks in order to establish an acoustic space for each speaker, which is necessary for a normalization procedure. We could not use a standard /hVd/ context (Peterson and Barney 1952) because, while English uses a glottal fricative /h/, Polish has a velar /x/ (Jassem 2003). This would have made vowels from both languages incomparable due to the fact that consonantal effects may persist throughout the whole vowel portion, including its target (Fox and Jacewicz 2009).

The target /bVt/ words were embedded in carrier sentences *I say /bVt/ this time* in English and *Mówię /bVt/ tym razem* in Polish in a non-utterance final position. This position was preferred because previous research has demonstrated a significant impact of utterance final positions on spectral properties of different sounds (e.g., Turk and Shattuck-Hufnagel 2000).

2.1.3. Procedure and recording

All recordings were made in a quiet room. English and Polish words were recorded in two separate sessions separated by 4 hours. In order to avoid a language-first effect, one half of the participant recorded English first and the other half started with Polish. Each word from every speaker was recorded twice. Both recording sessions took about fifteen minutes each. In order to ensure that the speakers would be in the desired language mode, the experimenter held a five-minute conversation with each participant in a target language prior to the recording. The carrier sentences were presented graphically on separate sheets for English and Polish. Although only three English vowels are analysed in this study, all vowels were recorded from each speaker for other research projects. Special care was taken to instruct speakers to produce the sentences as if speaking to a native speaker and to avoid unnecessary pauses and hesitations.

The carrier phrases were recorded with a Media Tech MT385 USB condenser microphone positioned 10 centimetres from a speaker's mouth. The speech input was processed and recorded by an external Sound Blaster X-Fi X-MOD sound card with a 24 bit sampling rate. The recordings were sampled at 44.10 kHz (24 bit accuracy) and subsequently stored in a notebook hard drive memory as WAV files ready for inspection.

2.1.4. Acoustic measurements

Prior the measurement session, all recordings were downsampled to 11.025 Hz and a Praat 5.1.17 speech-analysis software package (Boersma 2001) was used to scroll through the audio files in order to locate an onset and offset of target vowels. Frequencies of F1, F2 and F3 were measured at vowel mid-point, where the moment of formant movement is minimal, so as to avoid transition movement from and to the neighbouring consonants. Formant frequencies were computed with a 25-ms Hanning window with a default 14-pole LPC (linear predictive coding) prediction order, using add-on vowel analysis software Akustyk 1.8 (Plichta 2009). If the automatic analysis yielded clear errors (spurious formants or missed formants), LPC spectral envelopes and FFT (fast fourier transform) power spectra were compared in order to recompute a prediction order so that it would match a particular speaker's voice quality. The total number of analysed tokens was (6 Polish vowels + 3 English vowels) x 43 speakers = 387.

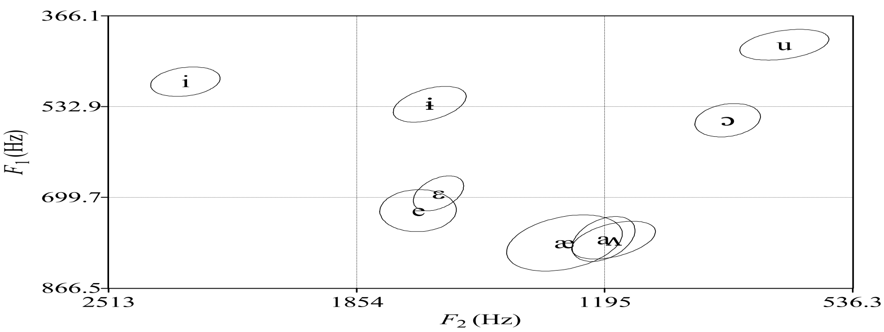
The raw measurements were subsequently normalised using the Lobanov transform (Lobanov 1971), which was found to perform exceptionally effectively in reducing anatomical and physiological variation between speakers while preserving phonemic identity in the acoustic measurements (Adank et al. 2004).

Vowel duration was measured from the onset of periodicity after the release burst showing clear formant structure to the beginning of consonant closure indicated by a rapid decrease in waveform amplitude and the cessation of energy in upper formants.

2.1.5. Analysis and results

Fig. 1 shows the interaction of analysed vowel categories computed with a Principal Component Analysis (PCA).

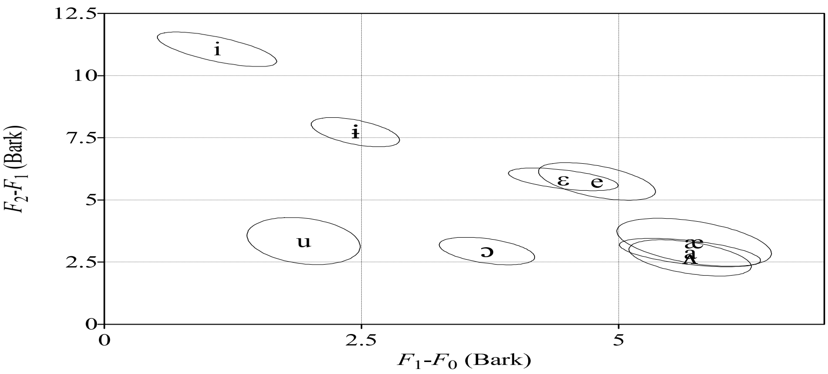
Figure 1: PCA analysis of vowel categories for all 6 Polish vowels and English /e/, /æ/ and /ʌ/.



The results show partial assimilation of both English /æ/ and /ʌ/ by a Polish /a/ category. The computed Euclidean distances indicate equal proximity of /æ/ to both /ʌ/ and /a/ and 2,5 larger distance to /e/. It is also noteworthy that /æ/ is characterised by a lack of stability along the F2 dimensions, which is evidenced by its relatively great standard deviations of F2 frequencies (155 Hz).

Subsequently, in order to quantify the category overlap between /æ/, /ʌ/ and /a/, we performed a discriminant analysis (DA) in Bark on raw measurements along the F1-F0 and F2-F1 dimensions (Syrdal and Gopal 1986). A confusion matrix revealed an equal proportion of 50% of the cases of /æ/ to be confused with /ʌ/ and /a/. The confusion with /e/ was observed for only 5% of all cases.

Figure 2: Discrimination analysis in Bark.



Vowel duration measurements in milliseconds revealed that /æ/ (Mean: 168; Std.Dev: 55) was significantly longer than /e/ (Mean: 148; Std.Dev.: 51) and /ʌ/ (Mean: 148; Std.Dev: 43) with repeated measures ANOVA, $F(2, 84)=11.020$, $p=0.000$. Post Hoc Fisher LSD did not yield a significant difference between /e/ and /ʌ/ ($p=0.88$).

2.2. Experiment 2: Perception

The objectives of the perception experiment were determined by findings from the analysis of production. More precisely, the production analysis showed that /æ/ has significant confusability with /ʌ/ in Polish learners' speech and that /æ/ is dissimilated from /ʌ/ by means of duration variability. The perception experiment was designed to determine if Polish learners are able to rely on spectral properties differentiating /æ/ and /ʌ/ or if they discriminate between the two vowels exclusively along the temporal dimension.

2.2.1. Participants

A total of 17 listeners were randomly selected from the group recruited for the production experiment.

2.2.2. Materials

We used an MBROLA diphone synthesiser (Dutoit et al. 1996) set for a male British English voice (en1 in a diphone database) to synthesise target vowels /æ/ and /ʌ/ in a /hVT/ context (/h/ 118 ms; /V/ 130 ms; /t/ 170 ms). The finals synthesis was manipulated to resemble a natural utterance in terms of the F0 contour (vowel onset 143 Hz F0; vowel offset 127 Hz F0) and normalized for intensity. The spectral properties for /æ/ were F1: 837 Hz; F2: 1592 Hz; F3: 2626 Hz, and for /ʌ/ F1: 968 Hz; F2: 1384 Hz; F3: 2802 Hz. Next, we applied local synthesis available in Akustyk 1.8 to generate 5 vowel steps between target /æ/ and /ʌ/ in equal increments (F1 22 Hz; F2 35 Hz; F3 29 Hz) and thus obtained 7 stimuli along the spectral dimension. In order to test the influence of vowel duration on vowel identification, we used PSOLA (the time-domain pitch synchronous overlap and add) to increase vowel duration from 130 ms to 200 ms in all stimuli. All the described procedures provided us with 14 stimuli for an experiment (7 short vowel steps and 7 long vowel steps).

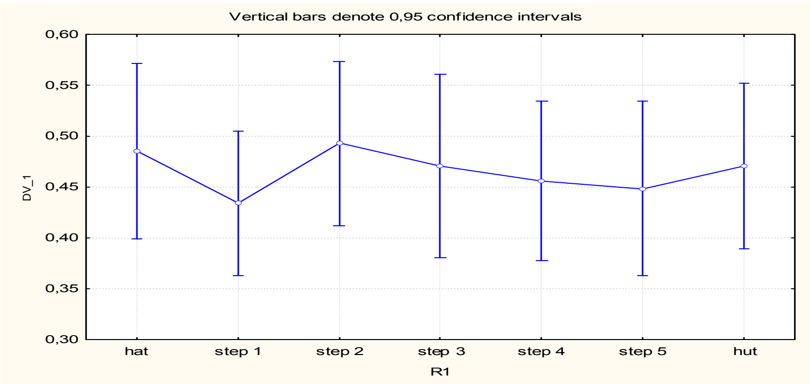
2.2.3. Procedure

The procedure was a self-paced identification task with freely available 'play again' option run by scripting Praat 5.1.17. The stimuli were blocked 7 short + 7 long x 4 repetitions and presented in a 'permute balanced no doublet' fashion. Each listener was presented with randomised 56 stimuli in a session that lasted approximately 15 minutes. The listeners were presented with orthographically rendered words 'hat' on the left and 'hut' on the right enclosed in squares on a computer screen. The 'play again' icon was in the middle position. The subjects were presented with the stimuli binaurally over headphones at a comfortable listening level at 70 dB. They were asked to click on the word they heard and, when uncertain, they were encouraged to make use of a 'play again' option. The target language mode was achieved by a short conversation and instructing in English.

2.2.4. Analysis and results

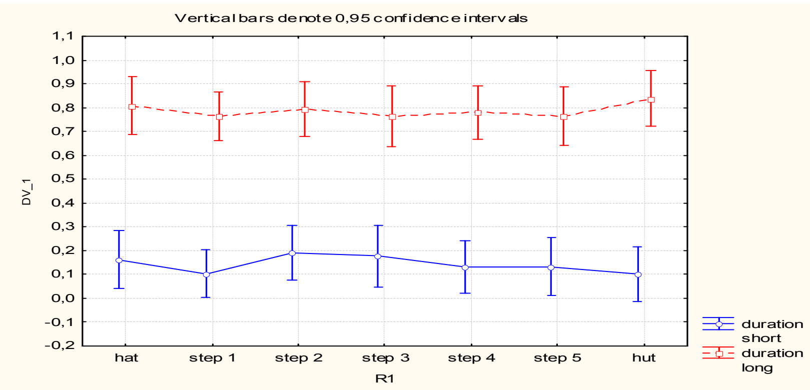
Repeated measures ANOVA was run with vowel duration as a categorical predictor and spectral steps as dependent variables. The global effect of pooled long and short stimuli along 7 steps between /æ/ and /ʌ/ was found not to be significant, $F(6, 192)=0.628$, $p=0.71$.

Figure 3: Identification rate of /æ/ for pooled long and short stimuli along 7 spectral steps.



A highly significant effect was found for vowel duration; stimuli with a longer vowel were significantly more often identified as /æ/, $F(1, 32)=98.897$, $p=0.000$.

Figure 4: Identification rate of /æ/ predicted by vowel duration for 7 spectral steps.



These results indicate that, similar to findings in the production experiment, the participants were insensitive to spectral cues between /æ/ and /ʌ/ but strongly relied on vowel duration, in that stimuli with longer vowels were consistently identified as /æ/ regardless of their formant frequencies.

3. CONCLUSIONS

The reported results for production of /æ/ in Polish learners' speech in our study lead to the following conclusions. The vowel /æ/ is attracted to and partly subsumed by Polish /a/. This scenario has also been found for Spanish speakers of English (Flege 1991, Flege et al. 1997). The vowel /ʌ/ is almost completely assimilated by /a/ in Polish and shares a significant portion of acoustic space with /æ/. The assimilation of /æ/ and /ʌ/ by Polish /a/ is predicted by PAM's single-category assimilation and SLM's assimilation of 'similar' categories. The compensating contrasting mechanism that is used to differentiate /æ/ from /ʌ/ is based on significantly increasing vowel duration for the former vowel.

A similar pattern is found in perception. Polish learners are insensitive to spectral properties differentiation /æ/ and /ʌ/. They are, however, highly reliant on vowel duration, in that longer vowels are identified as /æ/ even though they may have formant frequencies typical for /ʌ/.

The current findings contribute to results from previous research (see Escudero et al. 2009) which demonstrated that duration, which is secondary or even redundant for native speakers, may become a primary cue for non-native speakers.

4. REFERENCES

Adank, P., Smits, R., van Hout, R. 2004. A comparison of vowel normalization procedures for language variation research. *Journal of the Acoustical Society of America* 116. 3099-3107.

Best, C. T. 1995. A direct realist view of cross-language speech perception. In: Strange, W. (ed.), *Speech perception and linguistic experience: Issues in cross-language research.* Timonium: York Press. 171-204.

Best, C. T., Tyler, M. D. 2007. Nonnative and second-language speech perception: Commonalities and complementaries. In: Bohn, O-S., Munro, M. (eds.), *Language experience in second language speech learning: In Honor of James Emil Flege*. Amsterdam: John Benjamins. 13-34.

Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glot International* 10. 341-345.

Bogacka, A. 2004. On the perception of English high vowels by Polish learners of English. In: Daskalaki, E., Katsos, N., Mavrogiorgos, M., Reeve, M. (eds.), *CamLing 2004: Proceedings of the University of Cambridge second postgraduate conference in language research. Cambridge Institute of Language Research*. Cambridge: Cambridge University Press. 43-50.

Bohn, O.-S. 1995. Cross language speech production in adults: First language transfer doesn't tell it all. In: Strange, W. (ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Timonium: York Press. 279-304.

Bohn, O.-S., Flege, J. E. 1990. Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics* 11. 303-328.

Bohn, O,-S., Flege, J. E. 1992. The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition* 14. 131-158.

Bohn, O.-S., Flege, J. E. 1997. Perception and production of a new vowel category by adult second language learners. In: James, A., Leather, J. (eds.), *Second-language speech: Structure and process*. Berlin: Mouton de Gruyter. 53-73.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., van der Vrecken, O. 1996. The MBROLA project: Towards a set of high quality speech synthesizers free of use for non commercial purposes. In: Bunnell, H. T., Idsardi, W. (eds.), *Proceedings of the fourth international conference on spoken language processing*. 1393-3996.

Escudero, P., Benders, T., Lipski, S. C. 2009. Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. *Journal of Phonetics* 37. 452-465.

Flege, J. E. 1991. Orthographic evidence for the perceptual identification of vowels in Spanish and English. *Quarterly Journal of Experimental Psychology* 43. 701-731.

Flege, J. E. 1992. The intelligibility of English vowels spoken by British and Dutch talkers. In: Kent, R. (ed.), *Intelligibility in speech disorders: Theory, measurement, and management.* Amsterdam: John Benjamins. 157-232.

Flege, J. E. 1995. Second language speech learning: Theory, findings, and problems. In: Strange, W. (ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Timonium: York Press. 233-277.

Flege, J. E., Bohn, O.-S., Jang, S. 1997. Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics* 25. 437-470.

Fox, R. A., Jacewicz, E. 2009. Cross-dialectal variation in formant dynamics of American English vowels. *Journal of the Acoustical Society of America* 126. 2603-2618.

Gonet, W., Szpyra-Kozłowska, J., Święciński, R. 2010. Clashes with ashes. In: Waniek-Klimczak, E. (ed.), *Issues in accents of English 2: Variability and Norm*. Newcastle upon Tyne: Cambridge Scholars Publishing.

Jassem, W. 2003. Illustrations of the IPA: Polish. *Journal of the International Phonetic Association* 33. 103-107.

Lobanov, B. M. 1971. Classification of Russian vowels spoken by different speakers. *Journal of the Acoustical Society of America* 49. 606-608.

Major, R. 1987. Phonological similarity, markedness, and rate of L2 acquisition. *Studies in Second Language Acquisition* 9. 63-82.

Peterson, G. E., Barney, H. L. 1952. Control methods used in a study of the vowels. *Journal of the Acoustical Society of America* 24. 175-184.

Peterson, G. E., Lehiste, I. 1960. Duration of syllable nuclei in English. *Journal of the Acoustical Society of America* 32. 693-703.

Plichta, B. 2009. Akustyk for Praat (Version 1.8) [Computer program]. Retrieved August 7, 2009, from http://bartus.org/akustyk/.

Sobkowiak, W. 2003. *English phonetics for Poles*. Poznań: Wydawnictwo Poznańskie.

Strange, W., Akahane,-Yamada, R., Kubo, R., Trent, S. A., Nishi, K. 2001. Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners. *Journal of the Acoustical Society of America* 109. 1691-1704.

Syrdal, A. K., Gopal, H. S. 1986. A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America* 79. 1086-1100.

Turk, A. E., Shattuck-Hufnagel, A. 2000. Word-boundary related duration patterns in English. *Journal of Phonetics* 28. 379-440.

Wells, J. C. 1982. *Accents of English. Vol. 3.* Cambridge: Cambridge University Press.

Interaction of intrinsic vowel and consonant durational correlates with foreigner directed speech.

Joanna Sankowska[1], Maria Luisa Garcia Lecumberri[1] and Martin Cooke[1, 2]

[1] Language and Speech Laboratory (Laslab) Depto. Filologia Inglesa, Universidad del Pais Vasco, Spain; [2] Ikerbasque (Basque Science Foundation)
j.sankowska@hotmail.com, garcia.lecumberri@ehu.es, m.cooke@ikerbasque.org.uk

ABSTRACT

Foreigner-directed and Lombard speech are two examples of speech modes that have increased intelligibility compared to normal speech. Investigating the interactions between altered speech modes and phonological contrasts may throw light on the question of which details are vital in intelligibility enhancement. The present study compares the production of vowel shortening in English, a duration-based voicing correlate, realized in a listener-directed speech style (foreigner-directed speech) with native adult-directed speech and another listener-directed speech mode (Lombard speech). British speakers completed a communicative task in cooperation with an adult native speaker or adult foreigner, in quiet and in noisy conditions. Speaker productions were analyzed to examine the changes in the duration of the target vowels and following plosive consonants.

The results show that vowel shortening was present in the three speech styles. The durational voicing correlate was maintained in foreigner-directed and reduced in Lombard speech when compared with native adult-directed speech. Consonant durational differences were enhanced in foreigner-directed but reduced in Lombard speech relative to native adult-directed speech. The results suggest that foreigner-directed speech may be more intelligible in quiet conditions than Lombard speech, but less when both are presented with the same amount of noise.

Keywords: foreigner-directed speech, Lombard speech.

1. INTRODUCTION

Speech has been shown to accommodate to meet the needs of the listener. For instance, people tend to speak more loudly in the presence of noise (Lombard 1911; Dreher and O’ Neill, 1957; Summers et al., 1988; Garnier et al., 2006), and more slowly and with increased pause duration when addressing a child (DePaulo and Coleman, 1986; Biersack et al., 2005). A further listener-directed speech style is foreigner-directed speech (FDS), a speech mode that aims to enhance intelligibility. It is addressed to adult interlocutors with perceived limited linguistic capacity e.g. foreign learners. Previous research devoted to FDS has found a decrease in speech rate (Biersack et al., 2005; Scarborough et al., 2007), an expansion in vowel space (Knoll et al., 2004; Uther et al., 2007), and an increase in the duration of vowels (Scarborough et al., 2007). FDS can be compared with other speech styles that are modified in order to meet the communicative needs of the target audience. One of those listener-directed speech styles is Lombard speech (LS) that describes alterations in speaker vocal production in noisy environments. Previous research has found measurable differences in duration, pitch, intensity, and formant frequencies in the presence of noise (Summers et al., 1988; Junqua, 1993). Lombard speech is typically more intelligible than speech produced in quiet when both are presented in equivalent amounts of noise (Dreher and O’Neill, 1957; Summers et al., 1988; Lu and Cooke, 2008). In Lombard speech, reported differences in duration relate to the increase in the duration of words (Summers et al., 1988), syllables (Patel and Schell, 2008), certain consonants (Lu, 2010) and vowels (Junqua, 1993). Junqua also reports a slight decrease of the duration of consonants and Lu reports the decrease in the duration of voiceless labiodental fricatives and non-alveolar plosives.

An intriguing possibility is that changes in speech production induced by listeners’ needs might interact with phonetic correlates observed in normal speech. If foreigner-directed and Lombard speech aim to increase intelligibility then some cues may also be enhanced in order to emphasize contrasts. One well-known duration-based feature of English phonology is vowel shortening, which describes the phenomenon where a vowel followed by a voiceless consonant in the same syllable is shorter than it would be when followed by a voiced consonant (Wells, 1981; Gimson, 1989; Cruttenden, 2001) e.g. the vowel /iː/ in the word ´beat´ is shorter than in the word ´bead´. The current study seeks to describe possible interactions between duration-based contrasts and the properties of speech directed to a non-native speaker in real interactions. We investigate if vowel shortening is enhanced, maintained or reduced in FDS. Furthermore, we compare the changes occurring in FDS with those that are present in LS.

2. METHOD

Changes in vowel and consonant (plosive) duration were investigated in a communicative task in three different conditions: (i) in quiet when interacting with a native adult interlocutor, (ii) in quiet when interacting with a foreign adult interlocutor, and (iii) in the presence of stationary noise when interacting with a native adult interlocutor. Speakers produced target words in frame sentences which varied in content but were syntactically equivalent. Speaker productions were analyzed to examine differences in the duration of target vowels and following plosive consonants.

2.1. Task

Pairs of participants were involved in a communicative task designed for the purpose of this study. During the task, one of the participants (the speaker) had to compose sentences such as "Mr Gar Ven will say dog to Mr Garve again" using words on sets of cards prepared by the investigator. The sets of cards were attached to the table so that the subject could not change the structure of the sentence. Further, each set had a number in order to avoid mixing the words from two different sets. The speaker was instructed to produce the sentence for the other participant (the listener) to find it on the list of all possible sentences. The listener was instructed to find the sentence and repeat it for confirmation. After completing the task the participants were instructed to swap roles and complete the task again using different sets of cards.

2.2. Materials

Target words were chosen which contained long vowels /iː/, /ɑː/, /ɔː/ or short vowels /ɪ/, /æ/, /ɒ/, followed by either voiceless plosives /p/, /t/, /k/ or voiced plosives /b/, /d/, /g/. The target words were minimal pairs of real English words with the structure CVC and appeared sentence-medially, e.g.:

a) Mr Ghee Van will say *ford* to Mr Pork again,

b) Mr Key Fan will say *fort* to Mr Borg again,

c) Mr Gar Den will say *bag* to Mr Piece again,

d) Mr Car Ten will say *back* to Mr Bees again.

The first names that occurred sentence-initially were chosen to allow investigation of vowel shortening on the word boundary. These data and results are not presented in this paper. Vowels and consonants in coda position in the target words were used for the analysis. There were 432 tokens collected in total (144 in each of the three conditions).

2.3. Procedure

Each participant sat in a sound-attenuating booth and produced the sentences for the other person sitting in front of them but separated by a screen built for the purpose of this study. Each participant took part in 3 sessions which together lasted around 1 hour, including a prior practice and breaks. In the first session, speech was recorded in quiet conditions. During the second session both subjects were exposed to speech shaped noise at 85 dB SPL delivered through headphones. The third session involved completing the task in cooperation with an adult foreign interlocutor in quiet conditions. Recordings were done using a MOTU 8pre FireWire audio interface with Audio Desk 2 software and head-mounted microphones (Sennheiser MZA 900 P). Sessions was counterbalanced across speakers.

2.4. Speakers

Four native speakers of Standard British English with comparable southern English accents were recruited from the students and staff of the Computer Science Department at the University of Sheffield, UK. They were recorded in interactions with a British adult interlocutor and a foreign (Chinese) adult interlocutor with noticeable foreign (Mandarin) accent. None of the participants had any known history of speech or hearing impairment.

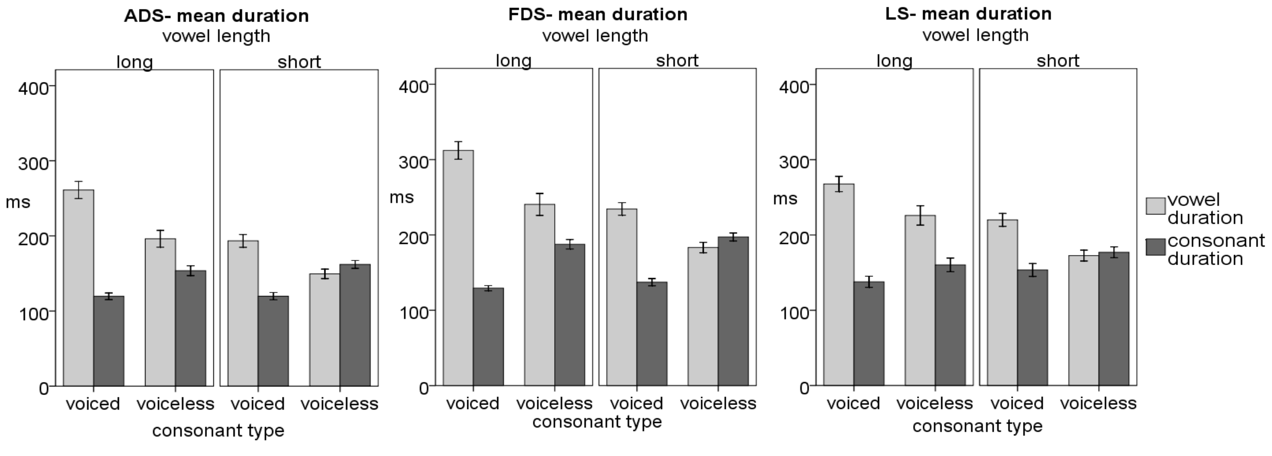
3. RESULTS

All acoustic features measurements were obtained using PRAAT (Boersma, P., Weenink, D., 2005). First, the durations of vowels and following plosives in the target words were measured. Second, the vowel durational correlate was calculated as the difference between the vowel duration in the voiced context and the vowel duration in the voiceless context, and the consonant durational contrast was calculated as the difference between the duration of the voiced consonant and the duration of the voiceless consonant.

3.1. Vowel shortening

Before analyzing the durational contrasts we checked that vowel shortening was present in all speech styles. Figure 1 shows mean vowel and consonant duration in the adult-directed speech (ADS), foreigner-directed speech (FDS), and Lombard speech (LS) conditions. Analysis of vowel and plosive duration revealed that vowel shortening due to the voicing properties of the following plosive was observed for both long and short vowels in ADS. The duration of long and short vowels was increased ($t(70)=4.083$, $p<.001$ and $t(70)=4.150$, $p<.001$ respectively) when followed by a voiced plosive. Analysis of the consonants also confirmed durational differences. Voiced plosives that followed both long and short vowels were shorter then voiceless ones ($t(70)=4.277$, $p<.001$ and $t(70)=5.882$, $p<.001$ respectively). Similar results were found for vowel and plosive duration for foreigner-directed speech. Long and short vowels were significantly longer ($t(70)=3.833$, $p<.001$ and $t(70)=4.754$, $p<.001$ respectively) when followed by a voiced plosive. Analysis of consonant duration also showed that voiced plosives that followed both long and short vowels were shorter than voiceless ones ($t(70)=8.113$, $p<.001$ and $t(70)=8.417$, $p<.001$ respectively). These results suggest that vowel shortening was still observed in this altered speech style. However, for Lombard speech, this pattern of results was observed only for vowels. The duration of long vowels was increased when the vowel was followed by a voiced plosive as opposed to a voiceless one ($t(70)=2.544$, $p<.05$) and the same trend was found for the short vowels ($t(70)=4.216$, $p<.001$). However, the analysis of the consonant duration showed differences between the plosives following long vowels and those following short vowels. For the voiced plosives following long vowels, there was a tendency for the duration to be decreased but the result fell short of significance ($t(70)=1.916$, $p=.059$). This may be due to the inter-speaker variability and the fact that some speakers may use other strategies to signal vowel shortening. However, the duration of the voiced plosives following short vowels was significantly shorter than voiceless plosives ($t(70)=2.103$, $p<.05$).

Figure 1: Mean vowel and plosive duration (error bars: +/- 1 Standard Error).

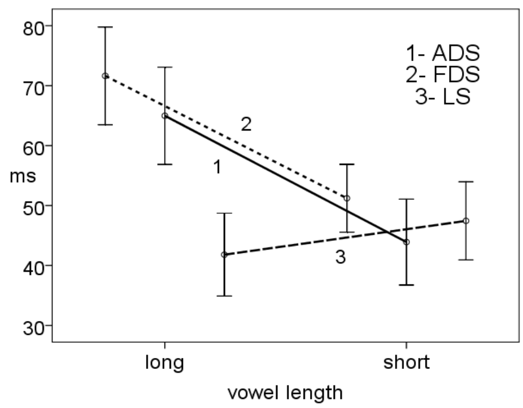


3.2. Vowel-duration-based voicing correlate

Durational differences for vowels followed by voiced vs. voiced consonants were analyzed by a repeated measures ANOVA with two factors of Length and Style, Length having two levels (intrinsically long vs. intrinsically short vowels), and Style having three levels (ADS vs. LS vs. FDS).

Figure 2 shows mean vowel duration-based voicing contrasts in the three speaking styles. We found no difference between the styles for short vowels. However, for long vowels we found that the durational contrast was decreased for LS relative to ADS and FDS. What is more, the results showed that in ADS and FDS the durational contrast was smaller for short vowels as opposed to long vowels and there was no difference in LS. The ANOVA analysis of vowel duration-based voicing contrasts confirmed these impressions and indicated a significant interaction between Style and Length ($F(2,34)=5.51$, $p<.01$) and a significant effect of Style ($F(2,34)=6.612$, $p<.01$). Different tendencies were observed for long and short vowels. For the long vowels, the durational contrast was significantly enhanced for FDS relative to LS ($p<.001$). It was significantly reduced for LS relative to ADS ($p< .01$), and no difference was found between ADS and FDS. As for the short vowels, no difference was found between the conditions. These results indicate that, at least for the long vowels, duration-based contrasts are in fact maintained in foreigner-directed speech and reduced in Lombard speech if both are compared to ADS as a base-line. Further analysis of the data revealed that, in ADS, the duration-based voicing contrast was significantly smaller for short vowels than it was for long vowels ($p< .05$). There was a tendency for the durational contrast to be smaller for short vowels than for long vowels in FDS but the difference fell short of significance ($p= .056$). No difference was found in LS between the long and short vowels. A comparison of adult-directed and Lombard speech showed that there was no significant interaction but a significant effect of Length $F(1,35)=5.4$, $p<.05$. The duration-based voicing contrast was reduced for short vowels rather than long vowels.

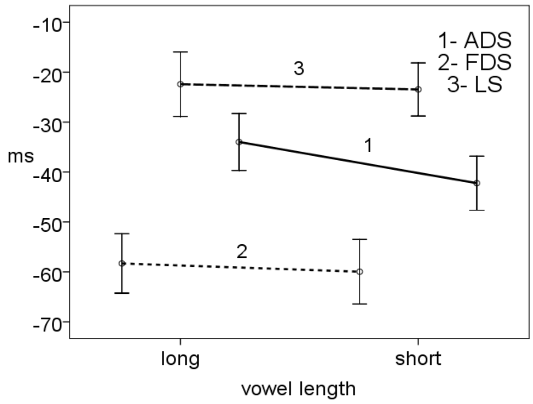
Figure 2: Mean vowel durational contrast.



3.3. Duration-based consonant voicing contrast

Figure 3 shows mean durational differences between voiceless and voiced consonants in the three speech styles. Consonant durational differences were enhanced for FDS and reduced for LS relative to ADS. A repeated measures ANOVA highlighted a significant effect of Style ($F(2,34)=28.069$, $p<.001$). The durational difference between voiced and voiceless consonants was enhanced for FDS relative to both ADS ($p< .01$) and LS ($p< .0001$). We also found that the durational difference was decreased for LS relative to ADS ($p<.01$). There was no statistically significant effect of Length. These results indicate that, although FDS and LS are both listener-directed speech styles, they exhibit different patterns of altering the speech. It seems that for plosives the duration-based voicing contrasts are reduced in LS but enhanced in FDS relative to ADS.

Figure 3: Mean consonant durational contrasts.



4. DISCUSSION

The aim of this study was to investigate a duration-based phonological contrast in speech directed to non-native speakers while conducting a communicative task, and compare it with adult-directed and Lombard speech. First, we investigated if vowel shortening before voiceless consonants was present in all speech styles. We found that speakers produced shorter vowels when followed by voiceless plosives rather than by voiced ones. Also, voiced plosives were shorter that their voiceless counterparts in all tested speaking styles. This suggests that although foreigner-directed and Lombard speech are altered modes of speech, vowel shortening is still present. Second, we investigated whether durational contrasts were maintained, reduced or enhanced in the listener-directed speech styles. Hence we compared the changes in speech in FDS with ADS as a baseline and another intelligibility enhancing speech mode i.e. Lombard speech. Analysis of the FDS data showed that vowel duration-based consonant voicing contrasts are in fact preserved in foreigner-directed speech. Different tendencies are present in Lombard speech, i.e., vowel duration-based contrasts are reduced in Lombard speech if compared to ADS as a baseline. What is more, at least for the long vowels, the durational contrast was significantly enhanced for FDS when compared with LS. Further analysis revealed that, as for normal (native adult-directed) speech, the durational contrast was smaller for short vowels than for long vowels. Also, there was a tendency for the durational contrast to be smaller for short vowels than for long vowels in FDS but not in LS. The overall increase in the duration of vowels is consistent with previous findings (Scarborough et al., 2007) who also found vowels significantly longer in FDS. Biersack (Biersack et al., 2005) found only a trend for longer vowels in FDS and Knoll et al. (2009) found no difference in vowel duration but this may be due to the fact in both studies the foreign listeners were imaginary. On the contrary, our study involved real listeners which provided real interaction and feedback to the speaker. Our results for LS are in line with the findings of Junqua (1993) who also found an increase in the duration of vowels.

To our knowledge, previous research in foreigner-directed speech has focused mainly on the durational analysis of vowels (Scarborough et al., 2007; Biersack et al., 2005). Our study extends the findings to the case of plosive duration. Analysis of the plosives showed that durational differences between voiced and voiceless consonants was substantially enhanced for FDS by nearly 50%. In contrast to the results obtained for FDS, the durational contrast in Lombard speech for plosives was reduced. As far as the mean duration of consonants is concerned, we found an increase in LS. These findings extend the results reported by Lu (2010) to the case of non-alveolar plosives (Lu found an increase of the alveolar plosives only). Our study shows some variability in acoustic changes from one speaking style to another. The overall increase in duration suggests that foreigner-directed speech is slower than normal and Lombard speech. We also assume that foreigner-directed speech may be more intelligible than both adult-directed and Lombard speech in quiet conditions. However, since it has been reported that Lombard speech is more intelligible than speech produced in quiet when both are presented in equivalent amounts of noise (Dreher and O'Neill, 1957; Summers et al., 1988; Lu and Cooke, 2008) we are unable to say how the changes seen in FDS affect intelligibility in noisy conditions. Further studies are needed to investigate potential complementary and antagonistic interactions between FDS, noise-induced speech, and phonological contrasts. Although our study involved a limited number of subjects and the task elicited read speech, the findings support the view that speakers adapt their speech to meet the needs of the target audience. Future studies on intelligibility-enhancing speech modes should involve communicative tasks that elicit more natural speech. Also, since both FDS and Lombard speech are intelligibility-enhancing modes of speech, future perception tests should show which speech styles and which phonetic details are responsible for intelligibility increases in non-native listeners in quiet and noise.

5. ACKNOWLEDGEMENTS

This work was supported by the Marie Curie RTN, Sound2Sense (S2S). The authors would like to thank Jon Barker and Guy Brown (University of Sheffield, UK) for help with recordings.

6. REFERENCES

Biersack, S., Kempe, V., Knapton, L. 2005. Fine-tuning speech registers: a comparison of the prosodic features of child-directed and foreigner-directed speech. *Proceedings of the 9th European Conference on Speech Communication and Technology, Lisbon*. 2401–2405.

Boersma, P., Weenink, D. 2005. Praat: Doing Phonetics by Computer (Version 4.3.04) [Computer Program], Available online: <http://www.praat.org>

Cruttenden, A. 2001. *Gimson´s pronunciation of English*. Oxford: A Hodder Arnold Publication.

DePaulo, B., Coleman, L. 1986. Talking to Children, Foreigners, and Retarded Adults. *Journal of Personality and Social Psychology* 51(5). 945–959.

Dreher, J. J., O' Neill, J. J. 1957. Effects of ambient noise on speaker intelligibility for words and phrases. *The Journal of the Acoustical Society of America* 29. 1320–1323.

Garnier, M., Bailly L., Dohen M., Welby P. & Lœvenbruck H. 2006. An Acoustic and Articulatory Study of Lombard Speech: Global Effects on the Utterance. *Proceedings of the Interspeech/ICSLP 2006, Pittsburgh (USA)*.

Gimson, A. C. 1989. *An introduction to the pronunciation of English*. London: Routlege.

Junqua, J. C. 1993. The Lombard reflex and its role on human listeners and automatic speech recognizers. *The Journal of the Acoustical Society of America* 93. 510–524.

Knoll, M.A., Uther, M. 2004. 'Motherese' and 'Chin-ese': evidence of acoustic changes in speech directed at infants and foreigners. *The Journal of the Acoustical Society of America* 116. 2522.

Knoll, M., Scharrer, L., Costall, A. 2009. Are actresses better simulators than female students? The effects of simulation on prosodic modifications of infant- and foreigner-directed speech. *Speech Communication* 51. 296–305.

Lombard, E. 1911. Le Signe de l'Elevation de la Voix (The sign of the rise in the voice) Ann. Maladiers Oreille, Larynx, Nez, Pharynx 37. 101–119.

Lu, Y. and Cooke, M. P. 2008. Speech production modifications produced by competing talkers, babble and stationary noise. *The Journal of the Acoustical Society of America* 124. 3261–3275.

Lu, Y. 2010. Production and perceptual analysis of speech produced in noise. Doctoral thesis, University of Sheffield.

Patel, R., Schell, K. W. 2008. The influence of linguistic content on the Lombard effect. *Journal of Speech, Language, and Hearing Research* 51. 209–220.

Sankowska, J., Cooke, M., Garcia Lecumberri, M.L. 2009. Interaction of intrinsic vowel and consonant durational correlates with Lombard speech. *4th Phonetics and Phonology in Iberia (PaPI) Conference, Las Palmas de Gran Canaria*.

Scarborough, R., Brenier, J., Zhao, Y., Hall-Lew, L., and Dmitrieva, O. 2007. An Acoustic Study of Real and Imagined Foreigner-Directed Speech. *Proceedings of the 16th ICPhS, Saarbrücken*. 2165–2168.

Summers, W., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I. & Stokes, M. A. 1988. Effects of noise on speech production: Acoustic and perceptual analysis. *The Journal of the Acoustical Society of America* 84. 486–490.

Uther, M., Knoll, M.A., Burnham, D. 2007. Do you speak E-n-g-l-i-s-h? A comparison of foreigner- and infant-directed speech. *Speech Communication* 49. 2–7.

Wells, J. C. 1981. *Accents of English. Three volumes*. London: Longman.

How sure are judges about their foreign accent judgements?

Sylwia Scheuer

University of Paris 3
sylwia_scheuer@yahoo.fr

ABSTRACT

The paper addresses the issue of subjectivity and instability of impressionistic foreign accent ratings obtained from untrained native listeners. A group of native English speakers was asked to rate the same speech samples 3 times, in an attempt to determine if the patterns of accent assessment would remain relatively constant throughout. Also, reference was made to the phonetic transcription of the samples, in order to relate the accentedness judgements to certain specific features of L2 speech.

Keywords: foreign accent, accent ratings.

1. INTRODUCTION

The composite nature of perceived foreign accent, as well as the various problems and dilemmas associated with measuring its degree, have been thoroughly discussed in the SLA literature (e.g. Piske et al. 2001, Major 2007). It has also been widely argued that individual studies into foreign accent are not always directly comparable, having involved markedly different types of speakers, listeners (if these were used at all), or experimental settings. The matter is further complicated by the fact that the listener – "the only truly linguistic measure" of foreign accent – may be more or less uncertain about his judgements, which leaves us with an "inherently subjective, and demonstrably inconstant source of information" (Markham 1997: 98). Therefore, an interesting question to explore is to what extent accent ratings – given by the same judges on different occasions – are indeed (in)constant, if the other crucial factors, like the speech samples and the overall procedure, are kept constant.

The present study compares foreign accent ratings of 7 native British English judges, received on 3 occasions: the first two sessions took place on the same day, whereas the third followed within six weeks. On each occasion the judges were asked to rate the same speech samples – elicited from a group of Polish students of English – for degree of foreign accent on a scale from 1 to 5. The study provides a follow-up to – and in some ways further validation of – an earlier foreign accent experiment carried out by the same author (Scheuer 2002), in which remarkable degree of constancy was found between accent judgements obtained from three different groups of native and non-native listeners.

2. METHOD

2.1. Speakers and speech materials

The speech samples were elicited from 15 Polish students (14 female, 1 male) at the School of English, Adam Mickiewicz University in Poznań. This was a 'free speech' task: the subjects were asked to share their impressions of student life. The recording session took place at the end of the students' first academic year, after all of them had completed an 8-month course in British English segmental phonetics, in addition to other university courses and lectures whose language of instruction was English. The students formed a fairly homogenous group in terms of age (between 19 and 22) and the amount of L2 experience, having learned English for 5-7 years, mainly from their Polish teachers at secondary school. The recordings also included speech samples produced by 3 control subjects, two female and one male, native speakers of Southern British English.

2.2. Listeners

The listeners were 7 native British English speakers, 6 male and 1 female, aged between 28 and 53. None of the judges had participated in an accent rating experiment before, and, with the exception of one listener (a teaching assistant at a French university, specialising in French-English translation), none was professionally involved in teaching English as a foreign language. 3 out of the 7 judges had been residing in France for 4 years prior to the experiment.

2.3. Procedure

The rating procedure adopted in the present study was similar to that used in Scheuer (2002), modelled on Bongaerts et al. (1995). Excerpts of 15-20 seconds in length, relatively free of syntactic deviations, were extracted from the original recordings. The listeners were asked to assess the degree of foreign accent evident in the speech samples on a scale from 1 to 5 (with possible half-point marks), where 1 stood for 'very strong foreign accent; about as far from native as it gets', and 5 for 'no foreign accent at all; definitely a native speaker, although possibly with a regional accent'. It was made clear to the listeners that the experiment was about their subjective impressions of foreign accent, and they did not need to justify their judgements in any way.

All the judges rated the same speech samples 3 times. The first two sessions (henceforth Round 1 and Round 2) were separated by an interval of approximately two hours, whereas round 3 took place between two and six weeks later. The order in which the samples were played to the listeners was identical for Rounds 1 and 2, but different for Round 3. Unlike the first two sessions, Round 3 did not include samples elicited from the control subjects.

3. RESULTS AND DISCUSSION

3.1. General remarks

3.1.1. The control group

Interestingly, one of the native speaker controls was given a score of 4 by 3 out of the 7 judges (round 1). After the experiment, the 3 judges explained that they had been misled by the subject's slow and hesitant manner of speaking. This goes to show, yet again, that foreign accent judgements are not based solely on degree of foreign accent as such, but also on other aspects of speech production, some of which may be universal, i.e. independent of the L2 in question (cf. Gut 2007). The results obtained by the control group will not be discussed any further in this paper.

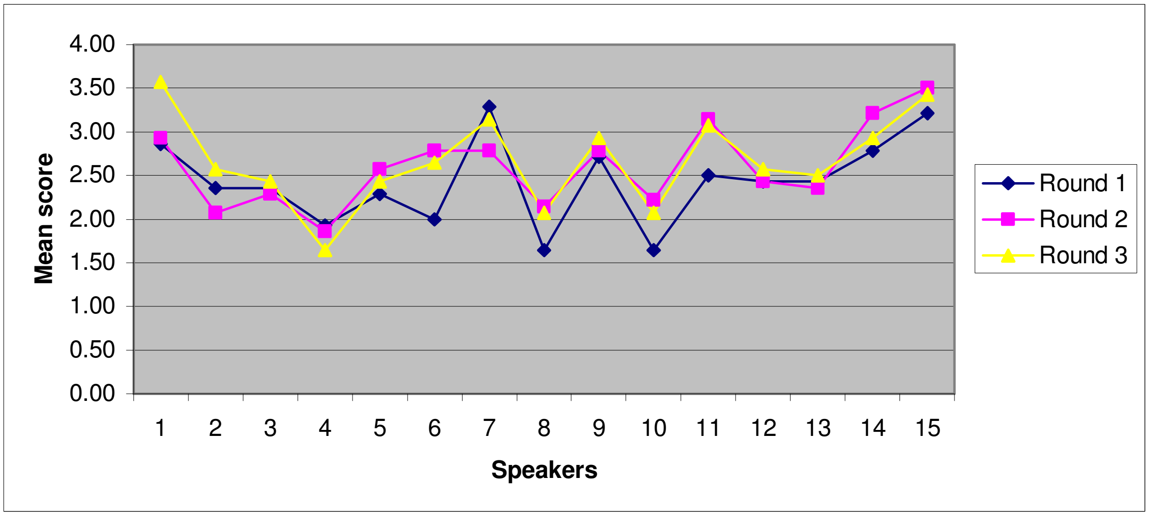
3.1.2. The experimental group

The average score given to the experimental group was 2.57. The mean score rose steadily with each judging session, from 2.43 in Round 1 through 2.60 in Round 2 to 2.67 in Round 3. It was also in Round 3 that the mean results obtained by individual speakers were most widely scattered, ranging from 1.64 (Subject 4) to 3.57 (Subject 1). Between-round comparisons are presented in the section below.

3.2. Round-to-round comparisons

The judges appeared to be more and more lenient as time went by. The mean score, averaged over the 7 listeners, significantly increased between Round 1 and Round 2, from 2.43 to 2.60 ($p<.05$), and even more so between Round 1 and Round 3 (2.43 to 2.67; $p<.005$). The difference between Rounds 2 and 3 proved to be statistically non-significant. The between-round comparison is illustrated in Figure 1.

Figure 1: Mean scores given to speakers in each round.



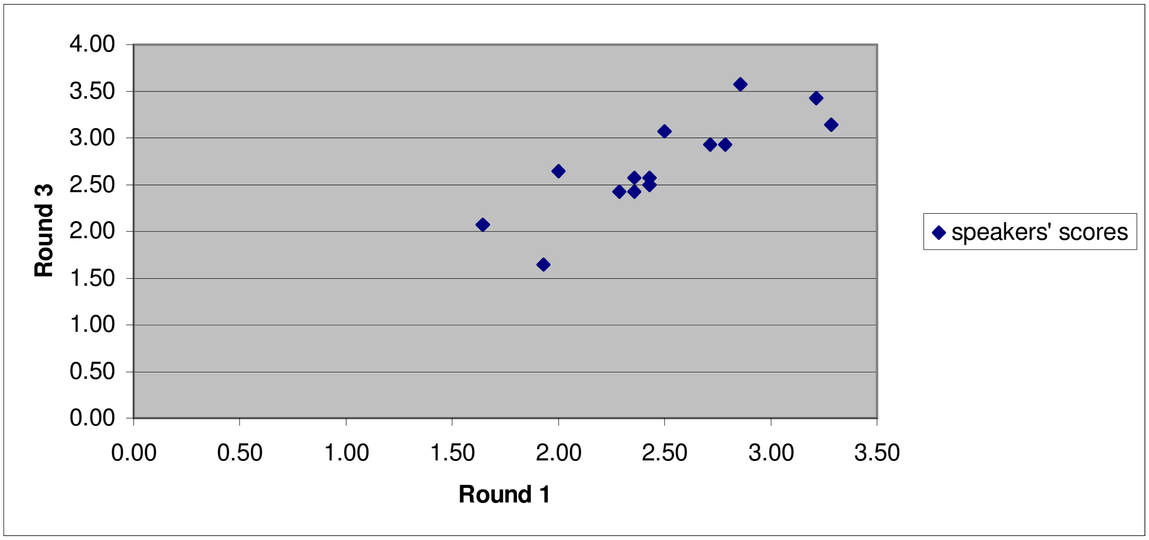
Needless to say, some sets of scores did not quite fit the general pattern. For example, speaker 4, whose accent was rated as the most foreign overall, received progressively lower scores throughout the experiment (1.93 > 1.86 > 1.64).

The steady rise in the judges' generosity deserves a word of comment. It may have resulted from their increasing familiarity with the speech samples, and therefore their increasing immunity to foreign accent evident in the recordings. Also, in Round 1, the judges were moving through unfamiliar territory; they were still gauging the range of accentedness to be encountered, so they may have been saving higher grades for better accents that might still present themselves (this is, however, a two-way argument). By the end of the experiment the listeners had gained a more complete picture of the accentedness spectrum, which would explain why the ratings for individual speakers in Round 3 were more dispersed than in either of the previous sessions.

3.3. How consistent were the judges in their judgements?

This was by far the most fundamental question that the study attempted to answer. In order to check to what extent the judges' impressions of foreign accent were of an ephemeral nature, likely to change drastically from one occasion to another, correlation coefficients were calculated for each set of paired results. These turned out to be impressively high. As far as the mean scores (the 7 judges lumped together) are concerned, the between-rounds correlations were highly significant: 0.72 for Rounds 1 and 2 ($p<.005$), 0.86 for Rounds 2 and 3 ($p<.0005$), and 0.85 for Rounds 1 and 3 ($p<.0005$). The latter correlation is presented in Figure 2.

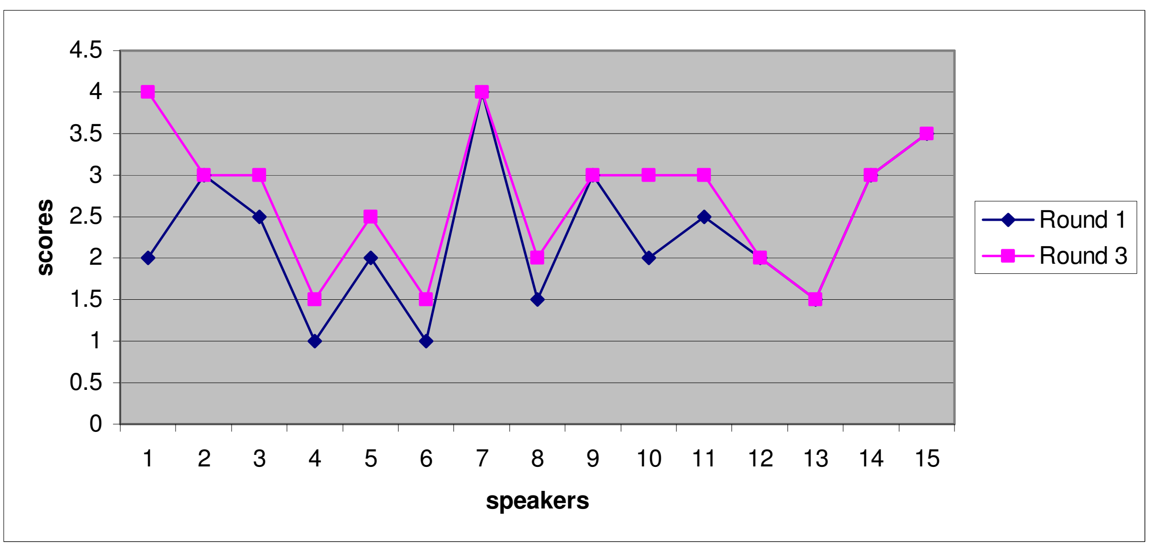
Figure 2: Mean scores given to speakers in Rounds 1 and 3.



When considered individually, the judges displayed varying degrees of consistency. Remarkably, once again a higher level of correlation was observed between Round 1 and Round 3 than between Rounds 1 and 2, which, after all, took place on the same day. In the former case, the correlation coefficients were highly significant for 6 out of the 7 judges, ranging from 0.48 to 0.8 ($p<.0005$). The one judge whose ratings failed to be statistically correlated ($r=0.33$) only used scores of either '2' or '3', which meant there was more room for chance results here. A significant level of correlation between Rounds 1 and 2 was achieved by 5 out of the 7 judges, with individual scores ranging from 0.5 to 0.84.

Figure 3 illustrates the rating pattern shown by one of the judges, for Rounds 1 and 3. Even though his ratings were significantly higher in Round 3 than Round 1 ($p<.01$), both sets of results were highly correlated ($r=0.8$; $p<.0005$), which means he consistently awarded relatively high/low scores to the same speakers.

Figure 3: The scores given to speakers by one of the judges in Rounds 1 and 3.



3.4. Phonetic signals of foreign accent

It is a commonplace observation that perceived foreign accent is a composite phenomenon, determined by a complex array of segmental and suprasegmental features of L2 speech. In my 2002 paper I attempted to pinpoint certain types of vocalic and consonantal errors that appeared to be statistically correlated with accentedness judgements passed by the native and the non-native listeners. While my conclusions were naturally very tentative, I found no significant correlation between the frequency of erroneous renditions of dental fricatives (pronunciations like *[dɪs] or *[wɪs] for ‘this’ and ‘with’) and foreign accent scores given by the two groups of native judges used in the experiment. As for the present study, at this stage I will limit myself to this particular phonetic variable. Rather unexpectedly, the relative frequency of ‘th’ errors was significantly correlated with the mean scores in each of the 3 judging sessions. The correlation coefficients ranged from -0.48 in Round 2 ($p<.05$) to -0.57 in Round 3 ($p<.025$). However, these findings must be treated with a great amount of caution, bearing in mind that the extracts played to the judges were very short, which meant that no high numbers of potential contexts were likely to arise: en error rate of 67% – seemingly very high – could mean just two erroneous and one correct rendition of ‘th’.

4. CONCLUSIONS

Although accentedness judgements are ultimately subjective and impressionistic, it stands to reason that they cannot be totally random or accidental. The most important finding of the present experiment was that the overall patterns of foreign accent ratings were impressively consistent throughout the 3 judging sessions, thus demonstrating that the native listener may not be as inconstant a source of information as is sometimes suggested. The study also lends support to the observation that degree of perceived accentedness cannot be dissociated from other – linguistic and paralinguistic – aspects of speech production like rate of delivery.

Traditionally, the author wishes to conclude by stating that more research is needed into the complex and elusive phenomenon of foreign accent.

5. REFERENCES

Bongaerts, T., Planken, B., Schils, E. 1995. Can late starters attain a native accent in a foreign language? A test of the Critical Period Hypothesis. In: Singleton, D., Lengyel, Z. (eds.), *The age factor in second language acquisition*. Clevedon: Multilingual Matters Ltd. 30-50.

Gut, U. 2007. Foreign accent. In: Müller, C. (ed.), *Speaker classification I*. Berlin/Heidelberg: Springer Verlag. 75-87.

Major, R.C. 2007. Identifying a foreign accent in an unfamiliar language. Studies in Second Language *Acquisition* 29, 539-556.

Markham, D. 1997. Phonetic imitation, accent and the learner. Lund: Lund University Press.

Piske, T., MacKay, I.R.A., Flege, J.E. 2001. Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics* 29. 191-215.

Scheuer, S. 2002. What makes foreign accent sound foreign? In: James, A., Leather, J. (eds.), *New Sounds 2000*. University of Klagenfurt. 306-314.

Voiced obstruents in L2 French: the case of Swiss German learners

Stephan Schmid

Phonetisches Laboratorium der Universität Zürich
schmidst@pholab.uzh.ch

ABSTRACT

The phonemic inventory of French coherently exploits the contrast between voiced and unvoiced obstruents. This opposition is distinctive in almost all phonotactic contexts, with two exceptions: i) /z/ does not occur word-initially; ii) the feature [±voice] may be assimilated in accordance with a following obstruent. Quite a different system is found in Swiss German dialects, where pairs of obstruents sharing the same place and manner of articulation are not differentiated through the presence/absence of laryngeal activity, but rather in terms of longer or shorter duration; such an opposition is maintained even word-finally, though some kind of postlexical fortition does occur if two obstruents follow each other.

Therefore, one may predict that Swiss German learners will find major difficulties in realizing French voiced obstruents before other obstruents, whereas they might be more prone to achieve voicing in intervocalic contexts. All in all, these hypotheses are confirmed by the acoustic analysis of a corpus of read speech: the overall degree of voicing in the whole data only amounts to 46%; intervocalically, 68% of the obstruents were voiced, whereas in the prepausal position the subjects only obtained a degree of voicing of 6%.

Keywords: Obstruents, voicing, French, Swiss German.

1. INTRODUCTION

Previous research on French as a second language has pointed at ‘Voice Onset Time’ (VOT) as a major source of pronunciation difficulties. In particular, the examination of Dutch and English learners has focused on the positive VOT in unvoiced stops, i.e. on the degree of aspiration of these sounds (Bongaerts 1999, Birdsong 2007). Quite differently, the present study takes into account another setting: for native speakers of Swiss German, it is rather the pronunciation of voiced obstruents that appears to contribute to their ‘foreign accent’ in French.

This contribution reports on an empirical study that investigates how Swiss German learners cope with the voicing contrast in L2 French. In section 2, a sketch of contrastive analysis describes the obstruents of French and Swiss German, illustrating the relevant subsets of the respective phoneme inventories as well as the phonotactic constraints and the allophonic rules that govern their distribution in the speech chain. Section 3 documents the procedures of data collection and data analysis, whereas sections 4 and 5 provide the presentation and a discussion of the main results of the study.

2. THE FEATURES [±VOICE] AND [±TENSE]: CONTRASTING FRENCH AND SWISS GERMAN

2.1. Voiced obstruents in French

As regards the subsystem of obstruents, the phoneme inventory of French is rather simple and coherent. As appears in Table 1, based on the illustration of the International Phonetic Alphabet provided by Fougeron and Smith (1999: 79), there are only twelve obstruent phonemes, i.e. three pairs of stops – at the labial, coronal and dorsal places of articulation – as well as three pairs of fricatives, which are labiodental, dental and postalveolar. In the core lexicon, French lacks affricates. For our purpose, it is important to note that, phonetically, “French voiced stops are typically voiced throughout” (Fougeron and Smith 1999: 80); the same – one could add – holds for fricatives as well.

Table 1: Obstruent phonemes in French

	Bilabial	Labiodental	Dental	Palato-alveolar	Velar
Plosive	p b		t d		k G
Fricative		f v	s z	ʃ ʒ	

The functional load of the voicing contrast is rather high, given that – in principle – it proves to be fully distinctive in word-initial, medial and also in word-final position, as one may observe in minimal pairs like /ʃu/ "cabbage" ~ /ʒu/ "to play (3rd pers. sg.)", /buʃe/ "butcher" ~ /buʒe/ "to move", /buʃ/ "mouth" ~ /buʒ/ "to move (3rd pers. sg.)". Thus, in the lexical phonology of French there is no such thing as the final obstruent devoicing known from several Germanic languages. However, a phonotactic constraint bans /z/ from the word-initial position and, moreover, an assimilation process may neutralize the feature [±voice] according to the specification of a following obstruent (cf. Léon 2007: 100). This phenomenon happens both within words (e.g. [ɔpsɛʀve] "to observe", [af̬gɑ̃] "Afghan") and across word boundaries (e.g., /ʒə kʀwa/ [ʒ̊kʁ̥wa] "I believe", /dɑ̃ sə baʀ/ [dɑ̃ s̬ baʀ] "in this bar"); the latter examples show that such postlexical devoicing or voicing is likely to occur after schwa deletion. From experimental studies on this topic we know that these allophonic rules are – to some extent – variable, rather than categorical (Rigault 1970; Snoeren and Segui 2003; Darcy and Kügler 2007; D'Apolito and Gili Fivela 2009).

2.2. The 'fortis' vs. 'lenis' contrast in Swiss German

Contrarily to French, Swiss German dialects completely lack voiced obstruents, instead showing binary opposition between so-called 'fortis' and 'lenis' consonants; it appears that the phonetic correlate of this distinction basically rests on the amount of closure duration (Willi 1996; Nocchi and Schmid 2006). Phonologically, the 'fortis' vs. 'lenis' contrast can be expressed by means of the feature [±tense] (Jakobson and Halle 1964: 100), but an alternative account has been proposed for the Thurgovian dialect, opposing singleton to geminate consonants (Kraehenmann 2003). In the following, I will refer to the Zurich dialect as described by Fleischer and Schmid (2006), adhering to the traditional distinction between 'fortis' and 'lenis' obstruents, the latter being transcribed with the IPA diacritic for voicelessness.

As emerges from Table 2, the feature [±tense] is regularly exploited for plosive and fricative phonemes, whereas there is only a single series of affricates. However, with respect to the feature [±tense], plosives differ from fricatives with respect to their phonotactic distribution: fricatives occur word-initially only as 'lenes', whereas tenseness is phonemically exploited in both word-internal and word-final contexts. For plosives, however, the contrast is relevant in word-initial, word-internal and word-final position. There is thus no equivalent to the final devoicing process observed in Standard German; both 'fortis' and 'lenis' obstruents can occur word-finally.

Table 2: Obstruent phonemes in Swiss German

	Bilabial	Labiodental	Alveolar	Palato-alveolar	Velar	Glottal
Plosive	p b̥		t d̥		k g̊	
Fricative		f v̥	s z̥	ʃ ʒ̊	x ɣ̊	h
Affricate		pf	ʦ	ʧ	kx	

Nevertheless, the fortis-lenis contrast is neutralized in another context, e.g. when a lenis plosive stands before a fortis plosive: in this case, fortition of the lenis plosive occurs, e.g., /d̥ə z̥æb̥ tɒːg̊/ → [d̥ə z̥æp tɒːg̊] "that day". Also, if a lenis plosive stands before another lenis plosive, the result is again a fortis cluster, e.g. /heb̥ d̥i/ → [hep ti] "hold tight!" (Fleischer and Schmid 2006: 248). Thus, unlike in French, postlexical neutralization in Swiss German only yields the unmarked (fortis) realization.

On the basis of this short sketch of contrastive analysis, one may hypothesize that a Swiss German learner of French will tend to pronounce L2 voiced obstruents as (voiceless) 'lenes'; devoicing is supposed to be particularly frequent in a consonantal context.

3. DATA AND METHODS

In order to test the hypotheses formulated above, a corpus of read speech has been collected at the Phonetics Laboratory of the University of Zurich. The corpus consists of twenty sentences containing the six voiced obstruents /b d g v z ʒ/ in six different phonotactic contexts: i) ##_V, ii) V#_V, iii) V_V, iv) V_C[+voice], v) V_C[-voice], vi) V_##. In the first context, the examined segment occurred utterance-initially, whereas in the second case the word-initial consonant was preceded by a word-final vowel. Within the word, the segment also occurred intervocalically and before a voiced or an unvoiced consonant; the last context contained the consonant in a word-final and utterance-final (prepausal) position. All in all, the sentences yielded 34 different segment types: six consonants multiplied per six phonotactic contexts, minus two contexts given that /z/ is not allowed in word-initially (see Schmid 2009: 260, 267-268 for a detailed description of the read words/sentences).

The twenty sentences were read aloud by ten students of a public high school in the town of Zurich; at the time of the recording, they were aged sixteen and seventeen and had experienced seven years of formal instruction in French. During the recording session, the students had to read the sentences two times; between the two readings, they were engaged in a short informal conversation regarding topics such as spare time and holidays. The recordings took place in the library of the high school building by means of a digital recorder Edirol R-1 and a Sennheiser ME66 supercardioid microphone; a sample rate of 44.1 kHz and a quantization of 16 bit were employed.

As regards the acoustic analysis of the 340 tokens (the 34 segment types mentioned above read by 10 subjects), waveforms and spectrograms were inspected manually using the software *Praat* (Boersma and Weenink 2010). For each token the duration of the periodic signal was measured by means of two indices, i.e. the presence of glottal pulses and of a fundamental frequency contour. Considering that voicing is not a matter of everything or nothing, the percentage of voicing time was calculated for the duration of each segment.

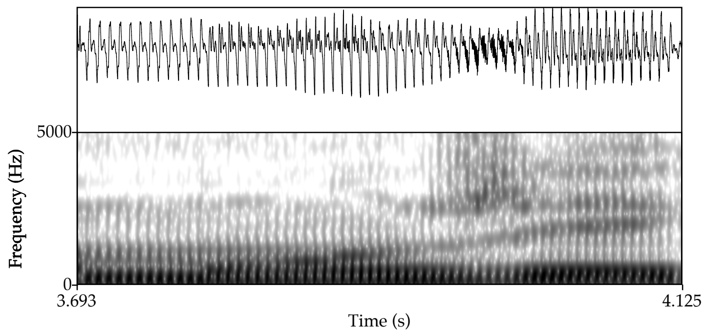
4. RESULTS

In the following sections, the recorded obstruents are analyzed from different point of views. First, a spectrographic illustration is provided of three types of fricatives – i.e. fully voiced, partially voiced, and fully devoiced. Second, the degrees of voicing are illustrated according to the phonotactic contexts in which the segments where uttered. Finally, differences in pronunciation accuracy among the 10 speakers are shown.

4.1. Voiced, partially voiced and fully devoiced fricatives

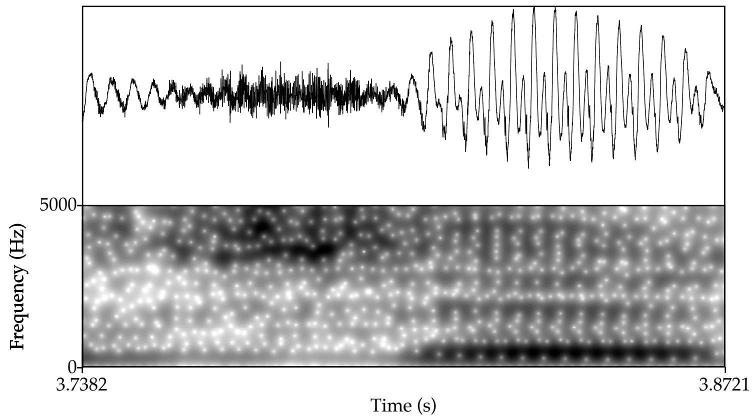
Fig. 1 shows the spectrogram of a fully voiced fricative in intervocalic position.

Figure 1: Wave form and spectrogram of the word *manger* "to eat" as pronounced by the speaker Dav.



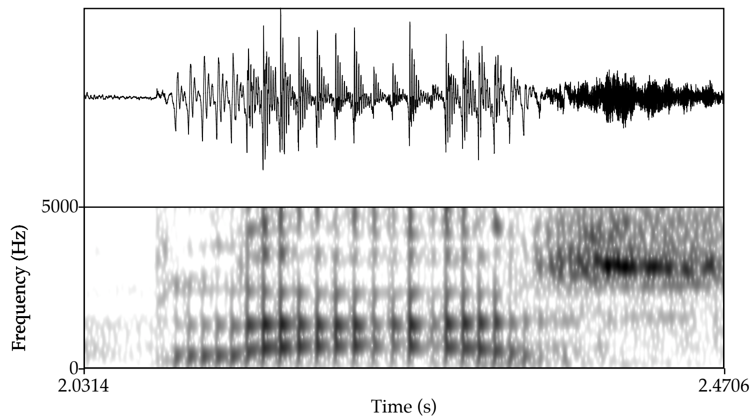
As is evident from the periodic oscillation in the wave form as well as from the continuous voice bar in the lower part in the spectrogram, the vocal folds are vibrating throughout the articulation of the fricative [ʒ]. However, in the word *je* “I” the same speaker pronounces a fricative [ʒ] which looses the initial periodicity during its articulation, as is shown in fig. 2.

Figure 2: Wave form and spectrogram of the word *je* “I” as pronounced by the speaker Lis.



Note that, word-finally, the first subject (Dav) also produces a fricative [ʒ] which is fully devoiced, as becomes clear from fig. 3

Figure 3: Wave form and spectrogram of the word *plage* “beach” as pronounced by the speaker Dav.



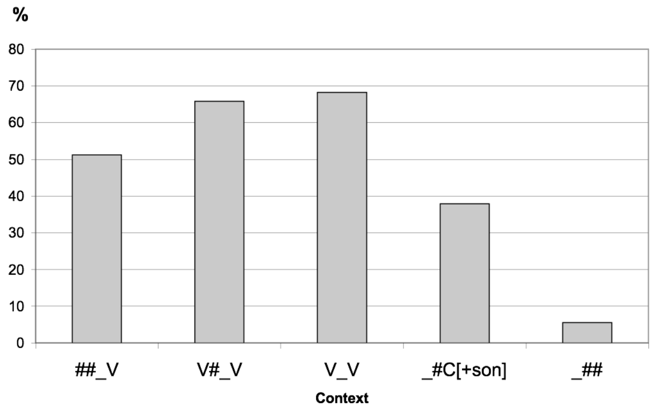
At this point, the question arises to which extent degrees of voicing are determined by phonotactic contexts and/or by individual differences among speakers.

4.2. Degree of voicing according to the phonotactic context

Fig. 4 illustrates the degree of voicing in five phonotactic contexts (the sixth context, V_C[-voice], is not considered here, as in standard French devoicing does apply as well).

The histogram shows that the amount of voicing of the investigated obstruents is indeed affected by their position in the sound chain. In particular, it appears that in word-internal intervocalic position speakers attain an accuracy of 68%. Note that the presence of a word boundary has no impact on the pronunciation of the obstruents, as the degree of voicing still amounts to 66%; this finding is far from surprising, as word boundaries do not play any substantial role in the phonology of neither French nor Swiss German. Utterance-initially, we find a degree of voicing of 51%, whereas the percentage decreases further to 38% before a voiced consonant (_#C[+son]). Most striking is the almost complete devoicing in utterance-final, prepausal position: in the whole corpus, the voiced French obstruent phonemes are pronounced with a degree of accuracy of only 6%.

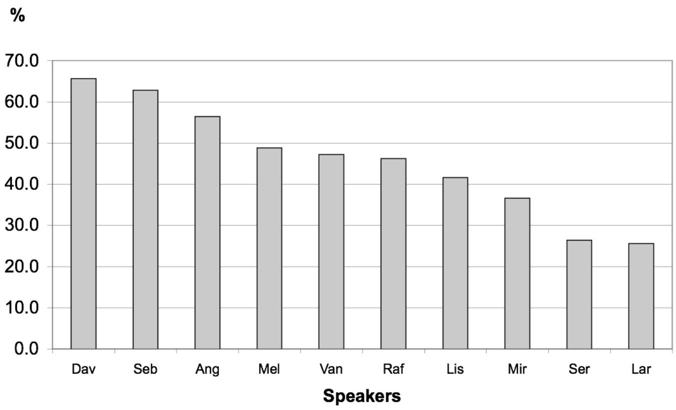
Figure 4: Degree of voicing of French L2 obstruents according to phonotactic contexts.



4.3. Degree of voicing according to speakers

Fig. 5 illustrates the individual differences observed among the subjects of our study, as far as degree of voicing is concerned.

Figure 5: Degree of voicing of French L2 obstruents according to speakers.



On the average, less than half of the whole amount of voicing time is realised by our speakers (46%). But obviously, some do better and some do worse. In fact, we find a considerable variation among the subjects, ranging from 24% and 25% (in the case of Ser and Lar) to 63% and 65% in the case of Seb and Dav; it is interesting to note that the subjects who performed most accurately are bilinguals, their mother tongue being Italian – a language that is characterized by highly voiced obstruents (cf. De Rosa and Schmid 2002; Schmid 2005).

5. DISCUSSION

There is no doubt that devoicing of voiced obstruents contributes to the 'foreign accent' in the French of Swiss German learners: given the pervasiveness of the feature [±tense] in their native dialects, they will tend to perceive and to produce voiced obstruents as 'lenes'. Not surprisingly, such a prediction is explicitely formulated in a grammar for Swiss teachers of French (Hilty and Wüest 1985: 28), and a previous experimental study has reported the devoicing of obstruents in the French of four subjects from Zurich

(Horner 1989: 48-49). It is important to note, however, that the substitution of L2 voiced obstruents with L1 'lenes' is not a purely paradigmatic process in the interlanguage phonology; distinguishing tokens by their phonotactic contexts (cf. fig. 4), one may detect two additional factors affecting the pronunciation of our subjects. The first syntagmatic process is still L1-based and derives from the postlexical fortition of two adjacent obstruents in Swiss German; here, our subjects behave quite differently from the Italian learners of French analyzed in the study of D'Apolito and Gili Fivela (2009), who were indeed able to pronounce sequences of two voiced obstruents. On the other hand, a second syntagmatic process that emerges from our data does not necessarily follow from the phonology of Swiss German, namely prepausal devoicing; note that this natural phonetic process has an even stronger impact on the performance of our subjects (cf. fig. 4).

Now, if interference from the L1 is not the only force that shapes the French interlanguage of Swiss German speakers, we must also stress the fact that not all voiced obstruents undergo devoicing; at least, this is the picture that emerges from our data. Even if the average accuracy of 46% is far from a native-like pronunciation, both intersubject variability (fig. 5) as well as variation due to the phonotactic contexts (fig. 4) demonstrate that our subjects have acquired the feature [±voice] to a certain extent. It may be unusual for a native speaker of Swiss German to pronounce a voiced obstruent in a second language, but it is not impossible.

6. REFERENCES

Birdsong, D. 2007. Nativelike pronunciation among late learners of French as a second language. In: Bohn, O., Munray, M. (eds.), *Language experience in second language speech learning: In honor of James Emil Flege*. Amsterdam: Benjamins. 100-116.

Boersma, P., Weenink, D. 2010. *Praat: doing phonetics by computer.* <www.praat.org>. Last access 15 Mar 2010.

Bongaerts, T. 1999. Ultimate attainment in L2 pronunciation: the case of very advanced late L2 learners. In: Birdsong, D. (ed.), *Second language acquisition and the Critical Period Hypothesis*. Mahwah NJ: Erlbaum. 133-159.

D'Apolito, S. and Gili Fivela, B. 2009. Il francese come lingue straniera: l'assimilazione del tratto di sonorità. In: Romito, L., Galatà, V., Lio, R. (eds.). *La fonetica sperimentale. Metodo ed applicazioni.* Torriana: EDK. 216-241.

Darcy, I., Kügler, F. 2007. Phonological context effects for voicing and devoicing in French. *Proceedings of the 16th ICPhS, Saarbrücken*. 1257–1260. <www.icphs2007.de>. Last access 15 Mar 2010.

De Rosa, R., Schmid, S. 2002. On the relationship between spelling and pronunciation in Italian-Swiss German bilinguals. In: James, A., Leather, J. (eds.), *New Sounds 2000*. University of Klagenfurt. 97-106.

Fleischer, J., Schmid, S. 2006. Zurich German. *Journal of the International Phonetic Association* 26. 243-253.

Fougeron, C., Smith, S. 1999. French. *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press. 78-81.

Hilty, G., Wüest, J. 1985. *Langue française: Phonétique, morphologie, syntaxe, différences de structure entre le français et l'allemand.* Zürich: Romanisches Seminar.

Horner, A. 1989. *Etude de problèmes phonétiques des Zurichois apprenant le français*. Mémoire de licence, Université de Strasbourg.

Jakobson, R., Halle, M. 1964. Tenseness and laxness. In: Abercrombie, D. et al. (eds.), *To honour Daniel Jones*. London: Longmans. 96-101.

Kraehenmann, A. 2003. *Quantity and prosodic asymmetries in Alemannic. Synchronic and diachronic perspectives*. Berlin: de Gruyter.

Léon, P. 2007. *Phonétisme et prononciation du français*. (5th edition). Paris: Armand Colin.

Nocchi, N., Schmid, S. 2006. Labiodentale Konsonanten im Schweizerdeutschen. In: Klausmann, H. (ed.), *Raumstrukturen im Alemannischen.* Bregenz: Neugebauer. 25-35.

Rigault, A. 1970. L'assimilation consonantique de sonorité en français: étude acoustique et perceptuelle. *Proceedings of the 6th International Congress of Phonetic Sciences*, *Prague*. 763-766.

Schmid, S. 2005. Spelling and pronunciation in migrant children: the case of Italian-Swiss German bilinguals. In: Cook, V., Bassetti, B. (eds.), *Second Language Writing Systems*. Clevedon: Multilingual Matters. 184-211.

Schmid, S. 2009. La prononciation du français par des élèves d'un lycée zurichois. In: Nüesch, H. (ed.), *Galloromanica et Romanica. Mélanges de linguistique offerts à Jakob Wüest*. Basel: Francke. 253-268.

Snoeren, N., Segui, J. 2003. A voice for the voiceless: voice assimilation in French. *Proceedings of the 15th International Congress of Phonetic Sciences*, *Barcelona*. 2325-2328.

Willi, U. 1996. *Die segmentale Dauer als phonetischer Parameter von 'fortis' und 'lenis' bei Plosiven im Zürichdeutschen: eine akustische und perzeptorische Untersuchung.* Stuttgart: Steiner.

The perception of Spanish lexical stress by French speakers: stress identification and time cost

Sandra Schwab[1,2] *and Joaquim Llisterri*[2]

[1] Université de Genève – [2] Universitat Autònoma de Barcelona
Sandra.Schwab@unige.ch – Joaquim.Llisterri@uab.cat

ABSTRACT

The perception of lexical stress in Spanish by French speakers with and without knowledge of the language has been studied with a technique allowing the evaluation of the individual or combined effects of the acoustic parameters related to the perception of stress. Results suggest in first place that the exposure to L2 makes the French speakers more sensitive to stress. Secondly, although F_0 seems to constitute the crucial cue in the identification of stress position, results point out that, when stress is accurately perceived, the time necessary to detect it is affected by manipulations involving amplitude.

Keywords: L2 perception, lexical stress, stress 'deafness', time cost.

1. INTRODUCTION

One of the most salient features in the production of French speakers learning Spanish as an L2 is the tendency to place the lexical stress in the last syllable of a word or phrase. Since French is a fixed-stress language in which stress generally appears in final position, an accentual transfer seems to take place when Francophone speakers attempt to pronounce proparoxytone or paroxytone words in a free-stress language such as Spanish. It has been hypothesized that the origin of this difficulty can be explained with the metaphor of the 'phonological filter' (Troubetzkoy 1939) that would be responsible for an insensitivity to perceive, and therefore, an inability to produce, contrastive stress differences. Along the same lines, the notion of 'stress deafness' has been put forward by Dupoux and his coworkers (Dupoux, Pallier, Sebastián and Mehler 1997; Peperkamp, Dupoux and Sebastián, 1999; Dupoux, Peperkamp and Sebastián 2001; Dupoux, Sebastián, Navarrete and Peperkamp 2008; Dupoux, Peperkamp and Sebastián 2010).

The results of a series of studies dealing with the perception of lexical stress in Spanish by Francophone speakers carried out by Dupoux, Peperkamp, Sebastián and other collaborators using different experimental procedures seem to suggest that subjects' sensitivity to stress placement depends on the cognitive charge required by the task and on the phonetic variability and the lexical status of the stimuli. Moreover, the performance in lexical decision or in repetition tasks does not appear to be strongly influenced by the degree of knowledge of Spanish. Taken together, these experiments lead to the conclusion that French speakers are unable to encode contrastive stress in their phonological representations although they might be capable, in certain tasks, to make use of the acoustic cues which are present in the speech signal. The sensitivity to acoustic cues such as fundamental frequency (F_0) has also been shown by Mora, Courtois and Cavé (1997) in an experiment on stress placement in Spanish utterances by French speakers. Their results, together with those of Muñoz, Panissal, Billières and Baqué (2009), mitigate the idea of a complete stress deafness in Francophone subjects.

The present study intends to shed some more light on the perceptual role of the three parameters involved in the phonetic realization of stress in Spanish (F_0, amplitude and duration) by considering the effect of their manipulation in the identification of lexical stress in Spanish isolated words by French speakers. The level of competence in L2, the lexical status of the items presented (words vs. pseudowords) and the accentual pattern of the word (proparoxytone, paroxytone or oxytone) are also taken into account, since they appear to be factors influencing the results of similar experiments (Alfano, Llisterri and Savy 2007; Alfano, Savy and Llisterri 2008, 2009; Alfano, Schwab, Savy and Llisterri 2010).

2. METHOD

The experimental procedure adopted has been initially designed for a study with native Spanish speakers (Llisterri, Machuca, de la Mota, Riera and Ríos 2005) and already followed in other works with non-native subjects (Alfano et al. 2007, 2008, 2009, 2010).

2.1. Participants

Two groups of French speaking participants took part in this experiment: a group with advanced knowledge of Spanish and another one with no knowledge of the language. The advanced group was composed of 10 subjects. They were between 21 and 36 years old and were all raised in a French speaking environment with only one language, French. They had been studying Spanish at University of Neuchâtel (Switzerland) during 6-11 years. The group with no knowledge of Spanish was formed by 10 students of the University of Neuchâtel. They were between 19 and 24 years old and were all raised in monolingual environment. None of them reported good knowledge of Italian, which excludes the eventual bias of knowing a free-stress Romance language.

2.2. Material

The corpus used, taken from Llisterri et al. (2005), was composed of 4 triplets of trisyllabic words (CV.CV.CV) and 4 triplets of trisyllabic analog pseudowords. All words and pseudowords could be proparoxytones (e.g. *número* 'a number'), paroxytones (e.g. *numero* 'I number') and oxytones (e.g. *numeró* 'he/she numbered'). The corpus was read 10 times by a native Spanish speaker. For each of the three vowels of the target words, the following measure were taken: F_0 at the beginning, at the centre and at the end of the segment; amplitude (Ampl) in five equidistant points along the vowel; and, finally, vowel duration (Dur).

The test stimuli were created in the following way: first of all, the original values of F_0, amplitude and duration were replaced in each vowel of each stimulus by the values averaged over the 10 repetitions (hereafter, *Base stimuli*); in a second stage, in proparoxytone words, F_0, amplitude and duration values for each vowel were replaced by the corresponding F_0, amplitude and duration values found in the equivalent paroxytone words (PP>P *Manipulated stimuli*); likewise, in paroxytone words, F_0, amplitude and duration values for each vowel were replaced by the corresponding F_0, amplitude and duration values found in the equivalent oxytone words (P>O *Manipulated stimuli*). In fact, manipulated stimuli resulted in a shift –to the right– of the accentual information.

The values were modified not only individually, but also simultaneously, obtaining the seven possible combinations of parameters: F_0, Ampl, Dur, F_0+Dur, F_0+Ampl, Dur+Ampl, F_0+Dur+Ampl. This strategy has allowed the study of the effects of each acoustic cue both in isolation and in combination with the others. All the manipulations were performed by resynthesis, using the PSOLA algorithm implemented in Praat (Boersma and Weenink 2010).

2.3. Procedure

A total of 136 stimuli (24 base items without manipulation plus 16 x 7 items with manipulations) were presented in the experiment, divided into 4 blocs of 34 items each containing the same number of base and manipulated stimuli and of words and pseudowords.

Subjects performed a stress identification task and were run individually. The stimuli were presented online from a laptop using the DMDX software (Forster 2010), which also recorded the subjects' responses and their reaction times. Subjects were instructed to listen to each stimulus (e.g. *número* 'a number'), to make a selection among the three possible choices (e.g. *número* 'a number', *numero* 'I number', *numeró* 'he/she numbered') that appeared in a row on the computer screen, and to press the corresponding button in a response box. The left-to right order of the three choices was always the same across trials: Position 1 corresponded to stimuli with stress on the first syllable, position 2 to stimuli with stress on the second syllable, and position 3 to the stimuli with stress on the third syllable. Each subject received a different randomization of the stimuli.

2.4. Data analysis

Reaction times (RT) were measured from the beginning of the stimuli. RTs inferior to 200 ms were removed (0.18% of the data set) and missing values were not replaced (3.3% of the data set). In order to avoid a possible bias, stimuli duration was subtracted from reaction times. Only RTs on correct responses (n=334) are examined in this paper. It should be noted that 'correct' in this context means that the subject has identified the intended position of the stress (i.e. on the second syllable in PP>P manipulations and on the last syllable in P>O manipulations). Reaction times have been analyzed using mixed-effect models (Baayen, Davidson and Bates 2008), since they do not only account for the fixed-effects factors but also for the random-effect factors, such as stimuli and/or participants.

3. RESULTS

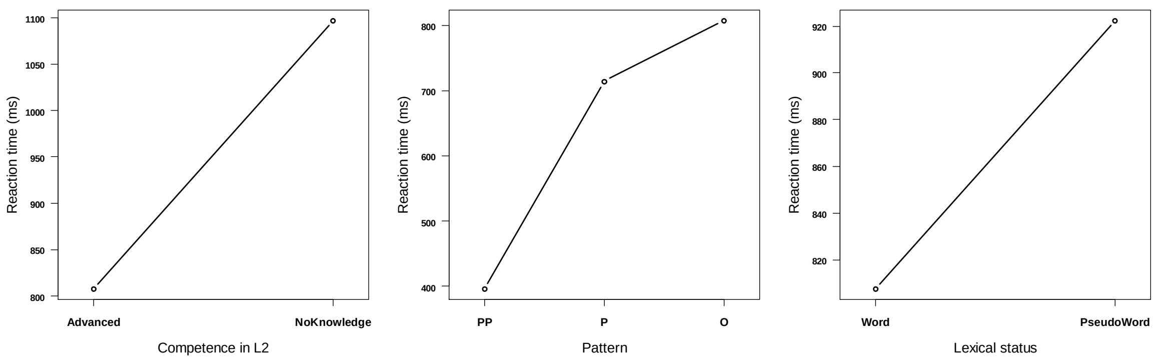
The first part of this section is dedicated to base stimuli, and the second part to manipulated stimuli. For both types of stimuli, we summarize the results of correct identification rate reported in Schwab and Llisterri (to appear) and in Alfano et al. (2010) on the same participants described in § 2.1, and we present the results of reaction times for correct responses, i.e. when participants accurately perceived the intended stress position.

3.1. Base stimuli

As far as percent correct identification of the base stimuli is concerned, Schwab and Llisterri (to appear) and Alfano et al. (2010) reported in the first place that French speakers correctly perceive stress in 71.5% of the cases, suggesting that they might not be so deaf to stress as it has been assumed. Secondly, the advanced learners of Spanish perceive stress more accurately than those with no knowledge of the language, indicating that the exposure to L2 makes French speakers more sensitive to stress. Thirdly, whatever the competence in L2 might be, stress on the first syllable is better perceived than stress on the second syllable, that is in turn better identified than stress on the third syllable. Finally, as for lexical status, no effect and no interactions with other variables were found.

Reaction times were analyzed by means of a mixed-effects model with participant and stimulus as random effects and reaction time as the dependent variable was run on correct responses. The predictors were group (advanced/no knowledge), lexical status (word/pseudoword) and pattern (PP, P, O). Following Baayen (2008), residuals larger than 2.5 times the standard deviation (5 data points out of 334, forming 1.5% of the data) were considered outliers and removed. The model refitted without these data points showed an effect of group, lexical status and pattern. Regarding group (see Fig. 1), advanced participants present shorter reaction times than participants with no knowledge of Spanish (β = -289.37, t = 2.568, $p < 0.05$)[1]. As far as pattern is concerned (see Fig. 2), PP pattern shows shorter reaction times in comparison with P pattern (β = -318.42, t = -7.425, $p < 0.001$) and O pattern (β = -411.89, t = -8.967, $p < 0.001$), and reaction times are marginally shorter for P pattern than for O pattern (β = -93.47, t = -1.949, p = 0.052). Finally, as for lexical status (see Fig. 3), reaction times are shorter for words than for pseudowords (β = -114.82, t = 3.161, $p < 0.01$). No interaction modulates these effects.

Figures 1, 2 and 3: Reaction times (ms) for the base stimuli as a function of group (in Figure 1, on the left), as a function of pattern (in Figure 2, at the center), and as a function of lexical status (in Figure 3, on the right).



3.2. Manipulated stimuli

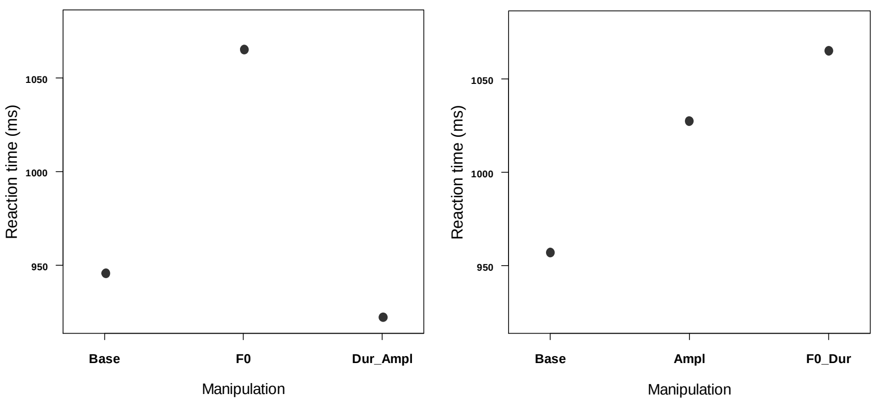
The most relevant results reported by Schwab and Llisterri (to appear) and Alfano et al. (2010) concerning the percent correct identification of the manipulated stimuli are the following: firstly, a combined manipulation of F_0, duration and amplitude leads to a better perception of the accentual shift than the separate manipulation of each acoustic parameter. This suggests that stress is perceptually not defined by a single parameter, but by the combination of parameters. Secondly, and more interestingly, both groups of French speakers (advanced and with no knowledge) don't behave in the same way according to the different acoustic manipulations. On the one hand, the advanced group perceives better the accentual shift when the three parameters (F_0, amplitude and duration) are jointly manipulated. On the other hand, while both groups are equally sensitive to the manipulation of F_0 (in isolation or in combination with amplitude of duration), the group with no knowledge of Spanish is more sensitive to the manipulations of duration or amplitude (in isolation or combined). It appears thus that French speakers with no knowledge of the L2 process stress in a more acoustic way. Thirdly, F_0 (alone or combined with duration or amplitude) seems the most important cue for a syllable to be perceived as stressed by French speakers. Indeed, researches in French (Rigault 1962; Dahan and Bernard 1996) have shown that F_0 is the decisive parameter in the perception of prominences in French L1.

As far as reaction times are concerned, a mixed-effects model with participant and stimulus as random effects and reaction time as the dependent variable was run on correct responses (i.e. when the participants perceived the intended stress position on the second syllable in PP>P manipulations and on the last syllable in P>O manipulations). The predictors were group, lexical status, pattern and manipulation. Since manipulation was the only significant fixed effect and no interaction modulated this effect, we ran different mixed-effects models in order to examine in detail the time cost induced by the manipulations in comparison with the base stimuli. In other words, instead of considering the whole set of manipulations, we performed separate analysis on subsets of manipulations.

First of all, we considered the subset of base stimuli and the stimuli in which the three parameters (F_0, duration and amplitude) were manipulated. A mixed-effects model with participant and stimulus as random effects, reaction time as the dependent variable and manipulation as predictor (base vs. F0_Dur_Ampl) shows no effect of the manipulation, indicating that the modification of the three parameters does not slow down the identification of stress, when stress is accurately identified. This suggests that the result of the acoustic manipulation was natural enough not to be noted by the listeners. Then, we examined different subsets of manipulated stimuli in comparison with base stimuli. Each subset was composed of a pair of complementary manipulations, in the sense that one manipulation concerns only one parameter in isolation (e.g. amplitude), while the other takes simultaneously into account the other two parameters (e.g. F_0 and duration). In this way, we can also observe the role of the parameters which were not manipulated (e.g. amplitude in the case of a combined manipulation of duration and F_0). The base stimuli were the third component of each subset.

The first subset of complementary manipulations we looked at was composed of stimuli with an isolated manipulation of F_0, stimuli with a combined manipulation of duration and amplitude, and base stimuli. A mixed-effects model[2] shows an effect of the manipulation, as can be seen in Fig. 4: reaction times are longer in F_0 manipulated stimuli than in base stimuli ($\beta = 119.78$, $t = 2.310$, $p < 0.05$) and than in stimuli with a combined manipulation of duration and amplitude ($\beta = 143.2$, $t = 2.396$, $p < 0.05$), whereas there is no significant difference between the base stimuli and the stimuli simultaneously manipulated in duration and amplitude ($\beta = 23.46$, $t = 0.447$, $p = 0.66$). These results suggest the presence of an inhibitory effect of F_0 in the perception of stress, while the combined manipulation of duration and amplitude does not slow down stress perception.

Figures 4 and 5: Reaction times (ms) as a function of manipulation (in Figure 4, on the left: base, isolated manipulation of F_0 and combined manipulation of duration and amplitude; in Figure 5, on the right: base, isolated manipulation of amplitude and combined manipulation of F_0 and duration).



The second subset of complementary manipulations consisted of stimuli with an isolated manipulation of amplitude, stimuli with a combined manipulation of F_0 and duration, and base stimuli. A mixed-effects model again shows an effect of the manipulation: despite the difference we can observe in Fig. 5 between base stimuli and stimuli manipulated in amplitude, reaction times in stimuli manipulated in amplitude don't differ from base stimuli ($\beta = 70.25$, $t = 1.115$, $p = 0.27$) nor do they from stimuli manipulated simultaneously in F_0 and duration ($\beta = 37.96$, $t = 0.597$, $p = 0.55$). On the other hand, reaction times in stimuli manipulated simultaneously in F_0 and duration are longer than in base stimuli ($\beta = 108.20$, $t = 2.567$, $p < 0.05$). These results seem to indicate the presence of an inhibitory effect of the combined manipulation of F_0 and duration, whereas the isolated manipulation of amplitude does not appear to slow down stress perception.

Finally, the third subset of complementary manipulations was formed of stimuli with the isolated manipulation of duration, stimuli with the combined manipulation of F_0 and amplitude, and base stimuli. A mixed-effects model shows no effect of manipulation, meaning that a manipulation of duration as well as a combined manipulation of F_0 and amplitude does not inhibit stress perception.

4. GENERAL DISCUSSION

We summarize and discuss here the results of stress identification rate (Schwab and Llisterri to appear; Alfano et al. 2010) and the results of reaction times, for base stimuli as well as for manipulated stimuli. As far as base stimuli –which didn't undergo any acoustic changes– are concerned, advanced learners of Spanish do not only perceive stress more accurately than participants with no knowledge of the language, but they are also faster in correctly identifying stress position. This strongly confirms that the exposure to Spanish as an L2 makes the French speakers more sensitive to stress. Secondly, whatever the competence in L2 might be, stress in the first syllable is detected more accurately and quicker than stress in the second syllable, which is, in turn, better and more rapidly identified than stress on the final syllable. Faster reaction times for PP pattern in comparison with P and O patterns are easily accounted for, given the fact that accentual information appears sooner in the PP stimuli than in the P and O stimuli. Thirdly, independently of the competence in L2 and the pattern, stress is perceived as accurately in words as in pseudowords, but when correctly perceived, stress in words is faster identified than in pseudowords. This discrepancy between the similarity in the identification of stress position in words and pseudowords and the difference in the time needed to process stress in words and pseudowords deserves further research.

Regarding manipulated stimuli, Schwab and Llisterri (to appear) and Alfano et al. (2010) reported that listeners are sensitive to different acoustic cues according to their competence in L2. This conclusion is not supported by results in reaction times, as there is no interaction between competence in L2 and the type of manipulation. In other words, when listeners are able to correctly perceive the stress shift induced by the different acoustic manipulations, advanced learners of Spanish and listeners with no knowledge of the language take the same amount of time to detect stress shift.

Schwab and Llisterri (to appear) and Alfano et al. (2010) also found that a manipulation of F_0 (alone or in combination with other parameters) leads to a more accurate perception of stress shift. When reaction times are considered, the picture is somehow different. We examined pairs of complementary manipulations compared to base stimuli in order to study, in terms of time cost, the role of the accentual information which is shifted to the next syllable (e.g. duration and F_0), and the role of the accentual information which is maintained as in the base stimuli (e.g. amplitude in stimuli manipulated in duration and F_0). We found first that a manipulation of F_0 has an inhibitory effect in the perception of stress, meaning that, when listeners perceive the stress shift induced only by F_0, reaction times are longer than for base stimuli. One possible interpretation is that the unmodified parameters (duration and amplitude) on the originally stressed syllable are strong enough to generate a conflict with the F_0 information on the next syllable. Consequently, the conflict between the original accentual information (duration and amplitude in the originally stressed syllable) and the shifted accentual information (F_0 on the next syllable) slows down the perception of stress in comparison with base stimuli. As far as the combined manipulation of duration and amplitude is concerned, results show no inhibitory effect. In that case, the original accentual information (F_0 in the originally stressed syllable) does not strongly conflict with the shifted accentual information (duration and amplitude on the next syllable), and does not slow down stress perception. This seems to suggest that the combination of duration and amplitude has more weight than F_0 in the perception of stress, given that, when they are maintained in the original syllable, there is a time cost, while there is none when shifted.

Secondly, results showed that a combined manipulation of F_0 and duration has an inhibitory effect in stress perception: when listeners perceive the stress shift induced by F_0 and duration, reaction times are longer than for base stimuli. On the other hand, the isolated manipulation of amplitude has no inhibitory effect. It seems thus that, in the first case, amplitude on the originally stressed syllable is strong enough to conflict with the F_0 and duration information on the next syllable, whereas, in the second case, the original accentual information (F_0 and duration in the originally stressed syllable) does not generate any conflict with the shifted accentual information (amplitude on the next syllable). It appears thus that amplitude has a stronger weight than the combined parameters of F_0 and duration, since, when it is preserved in the original syllable, there is a time cost, whereas there is none when shifted.

Finally, results indicated that neither the isolated manipulation of duration nor the joined manipulation of F_0 and amplitude has an inhibitory effect in stress perception: when listeners detect the stress shift, none of the two manipulations slows them down in comparison with base stimuli.

Schwab and Llisterri (to appear) and Alfano et al. (2010) concluded that F_0 is the most relevant parameter for the identification of stress position. Considering only cases in which listeners perceive stress shift, we found that the time needed to correctly detect stress is related to amplitude (alone or combined with duration): an unmodified amplitude implies a longer detection time of the shifted intended stress, while a shift in amplitude values does not. Taken together, these results do not minimize the role of F_0 in Spanish stress identification by French speakers (i.e. stress position is better identified when a change in F_0 is involved), but they draw attention to the effect of amplitude on the time necessary to perceive stress, when its position is correctly identified. Therefore, this research highlights the need to combine identification rate and time cost to get a more coherent picture of the processes involved in the perception of stress in an L2.

5. REFERENCES

Alfano, I., Llisterri, J., Savy, R. 2007. The perception of Italian and Spanish lexical stress: A first cross-linguistic study. *ICPhS 2007. Proceedings of the 16th International Congress of Phonetic Sciences*. 1793-6.

Alfano, I., Savy, R., Llisterri, J. 2008. Las características acústicas y perceptivas del acento léxico en español y en italiano: Los patrones acentuales paroxítonos. *Language Design. Journal of Theoretical and Experimental Linguistics. Special Issue 2: Experimental Prosody* 2. 23-30.

Alfano, I., Savy, R., Llisterri, J. 2009. Sulla realtà acustica dell'acento lessicale in italiano ed in spagnolo: La durata vocalica in produzione e percezione. In: Romito, L., Galatà, V., Lio, R. (eds.), *AISV 2007. La fonetica sperimentale. Metodo e applicazioni*. Torriana: EDK Editore. 22-39

Alfano, I., Schwab, S., Savy, R., Llisterri, J. 2010. Cross-Language speech perception: Lexical stress in Spanish with Italian and Francophone subjects. In: Schmid, S., Schwarzenbach, M., Studer, D. (eds.), *AISV 2009. La dimensione temporale del parlato*. Torriana: EDK Editore. 455-74.

Baayen, H. 2008. Analysing linguistic data. A practical introduction to statistics using R. Cambridge: Cambridge University Press.

Baayen, R. H., Davidson, D. J., Bates, D. M. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4). 390-412.

Boersma, P., Weenink, D. 2010. *Praat: Doing phonetics by computer* (version 5.1.25). <http://www.praat.org/>. Last access 26 Jan 2010.

Dahan, D., Bernard, J. M. 1996. Interspeaker variability in emphatic accent production in French. *Language and Speech*, 39(4). 341-374.

Dupoux, E., Pallier, C., Sebastián, N., Mehler, J. 1997. A destressing "deafness" in French?. *Journal of Memory and Language* 36(3). 406-421.

Dupoux, E., Peperkamp, S., Sebastián, N. 2001. A robust method to study stress 'deafness'. *The Journal of the Acoustical Society of America* 110(3-1). 1606-1618.

Dupoux, E., Peperkamp, S., Sebastián, N. 2010. Limits on bilingualism revisited: Stress 'deafness' in simultaneous French–Spanish bilinguals. *Cognition* 114(2). 266-275.

Dupoux, E., Sebastián, N., Navarrete, E., Peperkamp, S. 2008. Persistent stress 'deafness': The case of French learners of Spanish. Cognition 106(2). 682-706.

Forster, J. C. 2010. *DMDX updates page*. <http://www.u.arizona.edu/%7Ejforster/dmdx.htm>. Last access 07 March 2010

Llisterri, J., Machuca, M. J., de la Mota, C., Riera, M., Ríos, A. 2005. La percepción del acento léxico en español. *Filología y lingüística. Estudios ofrecidos a Antonio Quilis*. Madrid: Consejo Superior de Investigaciones Científicas - Universidad Nacional de Educación a Distancia - Universidad de Valladolid. 271-97.

Mora, E., Courtois, F., Cavé, C. 1997. Étude comparative de la perception par des sujets francophones et hispanophones de l'accent lexical en espagnol. *Revue PArole* 1. 75-86.

Muñoz, M., Panissal, N., Billières, M., Baqué, L. 2009. ¿La metáfora de la criba fonológica se puede aplicar a la percepción del acento léxico español? Estudio experimental con estudiantes francófonos. In: Bretones Callejas, M. C., et al. (eds.), *La lingüística aplicada actual: Comprendiendo el lenguaje y la mente*. Almería: Universidad de Almería – AESLA. 489-99.

Peperkamp, S., Dupoux, E., Sebastián, N. 1999. Perception of stress by French, Spanish and bilingual subjects. *Eurospeech 1999. Proceedings of the 6th European Conference on Speech Communication and Technology*. 2683-6.

Rigault, A. 1962. Rôle de la fréquence, de l'intensité et de la durée vocalique dans la perception de l'accent en français. *ICPhS 1962. Proceedings of the 4th International Congress of Phonetic Sciences*. 735-48.

Schwab, S., Llisterri, J. La perception de l'accent lexical espagnol par des apprenants francophones. To appear.

Troubetzkoy, N. S. (1939). Grundzüge der phonologie. *Travaux du Cercle Linguistique de Prague 7*.

NOTES

[1] β corresponds to the estimated coefficients in a regression model (i.e. the increase per unit change in x).

[2] All mixed-effects models carried out on subsets of manipulations were run with participant and stimulus as random effects, reaction time as the dependent variable and manipulation as predictor. Moreover, residuals larger than 2.5 times the standard deviation were considered outliers and removed (Baayen 2008).

Avoiding stop insertion after English /ŋ/ - a representational solution

Geoff Schwartz

UAM - Poznań
geoff@ifa.amu.edu.pl

ABSTRACT

A preliminary acoustic study of Polish learners' production of English angma is presented. The study provides support for a representational proposal by which angma is specified for the robust formant pattern on neighboring vowels. In this pattern F2 is raised to resemble palatal vocoids such as /i/ or /j/. A fronted articulation of angma is assumed to alleviate the articulatory difficulty in avoiding the insertion of a velar stop after /ŋ/. Results indicated that those learners who have acquired angma without stop insertion produce vowels with a consistently higher F2 before /ŋ/ than before /g/, while those having trouble with stop insertion do not. Textbook representations of angma that incorporate the segment's “palatal” quality on preceding vowels may assist learners trying to avoid stop insertion.

Keywords: Angma, Stop insertion, Phonological representation, Velars.

1. INTRODUCTION

The nasal /ŋ/, or *angma*, presents a significant obstacle for many foreign learners of English. The most serious problems appear in words such as *sings* or *singer*, in which the nasal is not followed by a velar stop – learners have a tendency in such cases to insert one. From the traditional description of angma as a velar nasal, we may assume that stop insertion results from difficulty controlling the release of the dorsal constriction on the soft palate. The pliant nature of the velar tissue serves to delay the release of the constriction, resulting in a buildup of pressure and a ballistic stop-like release. Because they may be said to insert a sound that isn't there, such errors are perceptually quite noticeable, and have become one of the trademarks of a stereotypical Slavic accent in English.

If stop insertion stems from an articulatory difficulty that is presumably universal, we must immediately raise the question of how native speakers of English avoid producing a stop after /ŋ/. Two possibilities present themselves. English speakers may perhaps achieve greater motor precision than speakers of other languages to control the release of angma. Unfortunately, evaluating this hypothesis would require advenced physiological study which I am not equipped to carry out. Alternatively, we might hypothesize that the traditional description of angma as a “velar” nasal might be misguided, and a more accurate representation might reveal that in the native language the articulatory challenge of producing /ŋ/ without stop insertion is not so daunting after all. This claim is a somewhat more accessible; acoustic data may be collected that reflects place of articulation.

This paper will provide a preliminary acoustic study of Polish learners' productions of angma in English, comparing tokens of students who have trouble with stop insertion to those who do not. In particular we shall test the hypothesis that the representation of /ŋ/ should be specified with a place of articulation somewhat further forward than the soft palate. Section 2 will present issues with the class of velars that arise out of transcription conventions. Section 3 will discuss the acoustic and perceptual properties of angma. Section 4 will translate these properties into a concrete representational proposal. Section 5 will present the results of the acoustic study. Section 6 will offer further discussion.

2. MOVING BEYOND THE SYMBOL /ŋ/

Traditional characterizations of angma illustrate the power that alphabetic transcription wields in the formation of “established” phonetic and phonological knowledge. The symbol /ŋ/ bears an obvious resemblance to the symbol /g/, so the notion that both sounds are “velar” is widely, if not univserally, accepted. The nasal has therefore “gone along for the ride” in descriptions of velars, and not received significant descriptive attention on its own. For instance, in Cruttenden's description of the velar stops (Cruttenden 1994: 153), we may note the observation that /k/ and /g/ may be articulated as palatals [c, ɟ] before high front vowels; two separate sagittal sections are offered, one for the front vowel context and one for the back vowel context. Conversely, in the section on /ŋ/ (Cruttenden 1994: 180), although it is mentioned that there is more advanced articulation for *sing* than for *song*, there is only a single sagittal diagram in which the constriction is so retracted to appear almost uvular. Aside from Cruttenden, most other textbook descriptions I have encountered treat velar as a homogeneous category, equal in status to any other place of articulation. Detailed articulatory descriptions of angma, at least as far as tongue position is concerned, are meant to be inferred from those of the velar stops. Authors who devote attention to the nasal generally concentrate discussion on the questionable phonemic status of /ŋ/ in English (see e.g. Roach 1991).

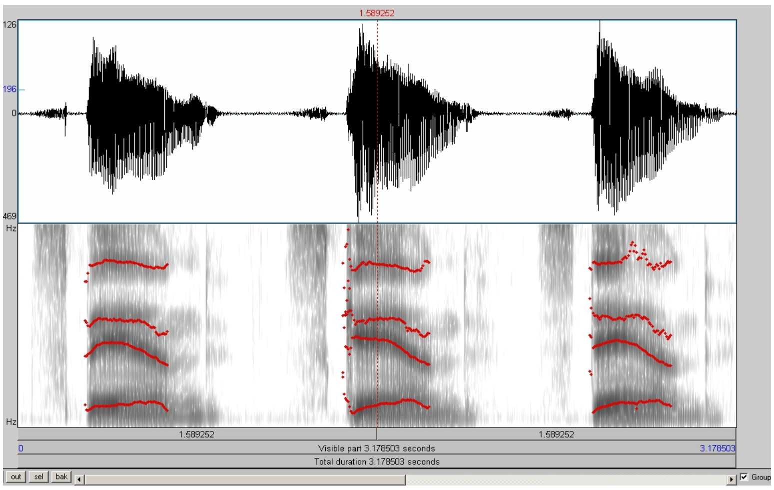
When it comes to “velars”, it has long been known that their place of articulaton is subject to significant variation. Notably, the articulation of velar stops is fronted in the environment of front vowels. However, the variation goes beyond simple fronting. In a detailed articulatory and acoustic study, Keating and Lahiri (1993) show that a fronted velar, as found in English in a word such as *key*, is distinct from a palatalized velar that that occurs in Russian. Moreover, both of these sounds are distinct from the palatal stop [c]. Unfortunately nasals were not researced in this study. However, Keating and Lahiri's findings suggest the possibility that variation within the class of velars need not be limited to coarticulatory effects. In other words, the place of articulation of angma may be distinct from /g/ and /k/, yet still fall into the larger category traditionally referred to as “velar”.

The essence of the hypothesis that we will test is that angma is what Keating and Lahiri refer to as a fronted velar, made with a dorsal constriction on the rear section of the hard palate, and subject to less contextual variation in place of articulation than /k/ and /g/. Articulation on the non-pliant hard palate does not present the same the difficulty in avoiding stop insertion, since there is no need to compensate for the softness of the velar tissue when planning the release of the constriction. Our experiment will seek acoustic evidence to support the proposal. We must therefore turn to the acoustic and perceptual properties of angma as well as the class of velars as a whole.

3. THE VELAR WEDGE AND THE PALATAL CONNECTION

Velar consonants are traditionally associated with two basic acoustic properties. The first is obstruent noise with a compact spectrum, the frequency of which is largely dependent on context. The second acoustic property associated with velars is housed on the formants of neighboring vowels. Velar formant transitions are known to produce a wedge (e.g. Lass 1996) or pinch (e.g. Baker et al 2008), in which the second and third formants converge. This can be seen in Figure 1, which shows a waveform and spectrogram display of the words *spanned*, *scanned*, and *stand* produced by a native speaker of American English. The formant tracings clearly show the velar wedge in the second token.

Figure 1: English *spanned*, *scanned*, and *stand* produced by a native speaker of American English.



Looking closely at the early portion of the vowel in these tokens, one notices that the velar formant transitions reflect an /i/-like quality – the velar wedge serves to raise the second formant, a feature typically associated with palatal vocoids. The F2 at vowel onset for these tokens was 1790 Hz after the /p/, 1895 Hz after the /t/, but 2210 Hz after the /k/. With regard to angma, Baker et al (2008) observe a longer, more salient velar pinch for /ŋ/ than for /g/, which they hypothesize is responsible for diphthongization processes in western US dialects of English that turn /æ/ into [eɪ] before angma and /g/. They also note that this change is more widespread before angma than before /g/. The implication of a /æ/>[eɪ] shift is that the palatal-like acoustic pattern associated with the velar pinch is reinterpreted on the preceding vowel. As Baker et al note, this quality is more salient in the case of angma than the oral stops, reflected in an implicational relation by which the diphthongization must happen before angma in dialects where it happends before /g/.

The title of Baker et al.'s study, "more velar than /g/", stems from a claim that the since the velar pinch is a defining property of velars, the angma, which features a more robust pinching pattern, is the most prototypical of the velar class. Alternatively, since angma produces more salient /i/-like formant transitions with a high F2, we could claim that it is "more palatal" than /g/. This palatal quality will the key to the representational proposal presented in the next section.

4. REPRESENTATIONAL CONSIDERATIONS

From a representational point of view, velars are rarely grouped with palatals such as /j/. Granted, in standard feature theory both velars and /j/ share a specification of [+high] or [dorsal]. However, the differing values the feature [back] – velars are [+back] while /j/ is [-back] – have seemingly taken precedence in most representational accounts of consonant behavior. This is probably due to the fact that velar palatalization (perhaps better labelled coronalization, as in Flemming 2002), which shifts the specification for the feature [back], is a thoroughly described phenomenon occurring in well-studied languages. Sound patterns in which palatals and velars behave similarly on the basis of their [+high] or [Dorsal] specification are less familiar, though certainly not unheard of. Examples include the hardening of /j/ to /k/ in Cypriot Greek (Kaisse 1992).

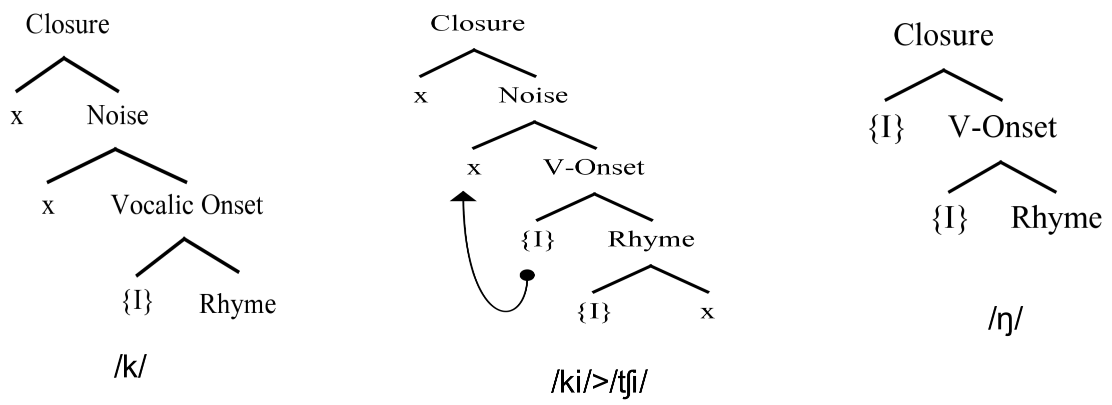
At the same time, works within the framework of Government Phonology (Kaye et al 1990), citing cases of prothetic velars on vowel-initial syllables (see e.g. Jensen 1994), suggest that the velar class must lack melodic specification. Phonetically, this lack of specification might be reflected in the fact that the noise spectrum associated with velar obstruents is largely context dependent, implying deficient melody that is susceptible to assimilation. In light of this discussion we must conclude that representations of velars must have the flexibility to capture both the palatal qualities of their formant patterns, and the context-dependence of their aperiodic noise.

In an Onset-based theory of representation (Schwartz, submitted), the structure of consonants is represented with binary branching trees (see also Pöchtrager 2006). Each layer in these trees is constructed from phonetic properties associated with manner of articulation. Place of articulation is represented by means of privative annotations to the terminal nodes. An important property of these trees is that melody may placed on different layers of the tree. In the representation of velars, which produce /i/-like formant transitions on neighboring vowels, we posit a palatal prime (either an element {I}, a [-back] specification, or simply a HighF2 annotation) on the Vocalic Onset layer of structure. This representational strategy offers an interesting perspective on the tendency for sequences like /ki/ to under go sound change, most frequently to /tʃi/. Since /k/ is specified for its /i/-like formant transitions, such sequences entail two consecutive {I} specifications and a violation of the Obligatory Contour Principle (OCP), which should result in perceptual ambiguity. The change to /tʃ/ moves the first of these specifications up the tree to the noise layer, removing the velar pinch and alleviating the OCP violation. At the same time, languages in which velars appear melodically deficient may be assumed to lack specification for the velar formant transition. In such cases "velarness" would then be based on the phonetic properties of the aperiodic noise.

To represent the idea that angma is "more palatal" than /k/, it is enough to posit that /ŋ/ contains an additional {I} specification higher up on the tree, on the Closure node. Angma thus contains two palatal specifictions to only one for the oral stops. In coda position, where the Vocalic Onset layer is removed from the representation, angma still contains a single palatal specification, while /k/ and /g/ have none. (Note also that we may distinguish angma from the palatal nasal /ɲ/. This segment is [coronal] and would not be specified for the velar pinch on Vocalic Onset since its F3 would be higher). The representational proposals are illustrated in Figure 2. On the left we see /k/ with the velar pinch specified on the vocalic onset layer. In the center /k/ loses its palatal specification to avoid an OCP violation, resulting in palatalization. On the right is the proposed representation for angma, which with two velar specifications is more "palatal" than the oral velar stops.

In sum, the discussion so far suggests that angma should be specified for its palatal properties, and that incorporating such a specification into textbook descriptions may have benefits for learners who have problems with stop insertion. The experiment described in the next section is designed to test this hypothesis. Learners who have acquired angma without stop insertion are predicted to have acquired its palatal qualities.

Figure 2: Proposed representations for /k/ (left), /ki/, and angma



5. EXPERIMENTAL METHOD

5.1 Subjects

12 first-year students of English at Adam Mickiewicz University in Poznań, Poland, divided into two categories based on evaluations by their English pronunciation teachers. The first group is described by their teacher as having acquired angma successfully with only sporadic stop insertion. This group also included one native speaker of American English. The second group is described by their teacher as having systematic and serious problems with stop insertion.

5.2 Data

A word list containing tokens of angma in two basic positions and two separate vocalic conexts: following either /ɪ/ or /æ/, and preceding a stop or not preceding a stop. The list also contained the oral stop /g/ when not preceded by angma, producing triplets. The /ɪ/-words included triplets of the type *sink/sing/dig*, *sings/sinks/digs* or *sinker/singer/digger*, while the /æ/ words were exemplified by pairs such as *bank/bang/bag, banks/bangs/bags,* and *hanger/anger/dagger*. In total the list contained five of the /ɪ/ triplets and three of the /æ/ triplets, as well as numerous filler words.

5.3 Analysis

The second formant was measured by hand at the onset of significant nasality, visible in the spectrogram as a weakening of formant structure, especially in the higher frequencies. Before the oral stops F2 was measured at the onset of closure. A set of t-tests were carried out on both group and individual data. Additionally, the number of stop insertion errors for each subject was noted for the purposes of testing any correlation, irrespective of the category designation, between the formant data (the ratio of F2 in tokens containing angma to F2 of those that did not) and the likelihood of stop insertion.

5.4 Results

The results are summarized in Tables 1 and 2. Table 1 shows the group data: mean F2 values for vowels before angma and before /g/. The non-stop inserters produced a significantly higher F2 before angma in both vowel contexts ($p=0.03$ for /ɪ/; $p<0.01$ for /æ/). The stop-inserters produced did not produce a significantly higher F2 in either context ($p=0.38$ for /ɪ/; $p= 0.11$ for /æ/), although in the case of /æ/ the distinction approached significance. The results for the /æ/ may have been confounded by the fact that this is a notoriously difficult vowel for Polish learners. Two substitutions are common, /e/ and /a/, and there was some inconsistency in the subjects' productions of this vowel. We may conclude that the results for the /ɪ/ vowel are probably more reliable.

Table 1: Group results

Mean F2	/ɪŋ/	/ɪg/	/æŋ/	/æg/
Group 1 – non-stop inserters	2325	2186	2110	1856
Group 2 – stop inserters	2275	2229	2047	1827

The results for the individual subjects, irrespective of the teachers' grouping, are presented in Table 2. The number of stop-insertion errors made by each subject is noted, as well as the ratio of mean F2 before angma to F2 before /g/ in both vowel contexts. This calculation was intended to quantify the degree to which these students acquired a higher F2 for vowels before angma. Pearson correlation coefficients were calculated between the number of errors and the F2 ratios. A solid inverse correlation (-0.78) was found for /ɪ/, implying that the Higher the F2 for a subject's vowel before angma, the lesser the likelihood of stop insertion. There seemed to be little or no correlation for /æ/ (-0.2), though the inconsistencies in subjects' pronunciation of 'ash' may have affected the data for this vowel. Significance levels are given for the /ɪ/ context only in cases which were not significant ($p<0.05$ for all subjects where significance is not noted). Significance levels were not calculated for the /æ/ as a result of inconsistencies in students' pronunciation, as well as the low correlation score with the number of errors.

Table 2: Individual results

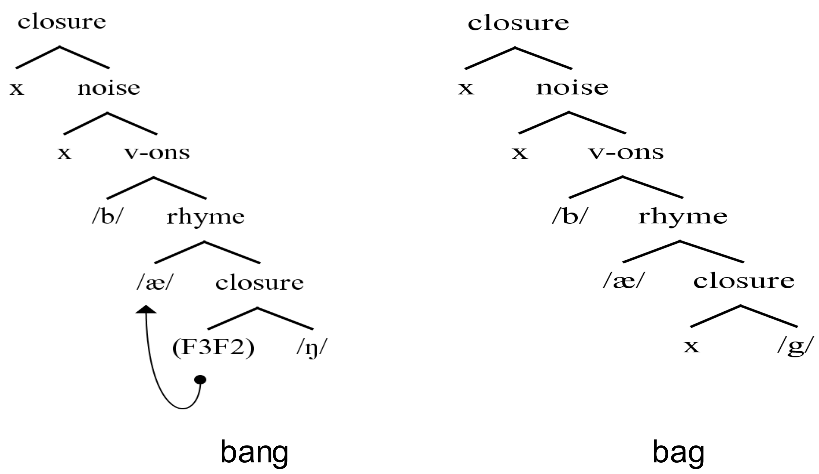
Subject	# of errors	F2(ɪŋ)/F2 (/ig/)	F2(æŋ)/F2 (/æg/)
1	2	1.11	1.07
2	7	1.09	1.08
3	6	1.07	1.24
4	2	1.04 (p=0.09)	1.01
5	0	1.12	1.27
6	8	1.01 (p=0.85)	1.15
7	9	1.02 (p=0.34)	1.38
8	6	1.04	1.22
9	11	0.99 (p=0.84)	1.07
10	7	1.04 (p=0.38)	1.03
11	11	1.01 (p=0.49)	1.01
12	13	1.02 (p=0.3)	1.03

6. DISCUSSION

The results of the present study offer support for the hypothesis that acquiring angma without stop insertion implies the acquisition of a higher F2 on preceding vowels, suggesting a fronted velar articulation (cf. Keating and Lahiri 1993). In particular the results for the vowel /ɪ/ were quite suggestive. Further confirmation of the fronted quality of angma may provide a real benefit for textbook descriptions of this sound. The notion that the nasal is distinct from /g/ might open the eyes and ears of both teachers and learners struggling with stop insertion words containing /ŋ/. Since angma has always been represented with a symbol that “looks velar”, it's velarness may have been taken for granted. This study suggests that for successful acquisition of second language sounds we sometimes need to go beyond the conventions of segmental transcription and orthography, and look more closely at the dynamic properties of speech.

From the perspective of the phonetics-phonology relationship, it is worth mentioning again that the F2 measurements for the present experiment were made not on the nasal itself, but on the offset of preceding vowels. If the observed difference between angma and /g/ is indeed systematic in English, phonological representations must be capable of modeling it. Thus, representations of consonants should be expressible in terms of formant transitions on neighboring vowels. Such a strategy is adopted in the representational proposals in Section 4. The angma-/g/ distinction in coda position is illustrated from this perspective in Figure 3, which offers representations of the English words *bang* (left) and *bag*. The angma contains the velar pinch (denoted with the annotation F3F2) on the Closure node, which may spread to the preceding vowel and cause diphthongization. The oral stop /g/ is annotated for the velar wedge lower down in the structure, on the Vocalic Onset node. In coda position this node is assumed to be pruned from the structure, so spreading of the velar formant pattern on the preceding vowel is less likely in the case of *bag*.

Figure 3: representations of *bang* and *bag*



7. REFERENCES

Baker, A., J. Mielke, and D. Archangeli. 2008. More velar than /g/: consonant coarticulation as a cause of diphthongization. In *Proceedings of the 26th West Coast Conference on Formal Linguistics,* ed. Charles B. Chang and Hannah J. Haynie. Somerville, MA: Cascadilla Proceedings Project. 60-68.

Cruttenden, A.. 1994. *Gimson's pronunciation of English: Fifth Edition*. London: Arnold.

Flemming, E. 2002. *Auditory representations in phonology*. New York: Routledge.

Jensen, S. 1994. Is ʔ an element? Towards a Non-segmental Phonology. *SOAS Working Papers in Linguistics and Phonetics* 4. 71-78.

Kaisse, E., 1992. Can [Consonantal] spread? *Language* 68(2). 313–332.

Kaye, J., J. Lowenstamm, and J-R. Vergnaud. 1990. Constituent structure and government in phonology. *Phonology Yearbook* 7. 193-231.

Keating. P. and A. Lahiri. 1993. Fronted velars, palatalized velars, and palatals. *Phonetica* 50. 73-101.

Lass, N. J. 1996. *Principles of expermental phonetics*. St. Louis: Mosby.

Pöchtrager, M. 2006. The structure of length. Ph.D. dissertation. University of Vienna.

Roach, P. 1991. *English Phonetics and Phonology – A practical course.* Cambridge: CUP.

Schwartz, G. Submitted. An onset-based theory of segmental representation. Under review for publication in *Lingua*.

Prosody in Hong Kong English: Aspects of speech rhythm and intonation

Jane Setter

Department of Applied Linguistics, The University of Reading
j.e.setter@reading.ac.uk

ABSTRACT

Setter (2006) adopted a pedagogically oriented, hierarchical methodology to examine Hong Kong English (HKE) speech rhythm, in which the duration and proportion of weak, unstressed, stressed and nuclear syllables in a speech task were compared with an existing corpus of British English (BrE) data. It was found that rhythmic patterns in HKE differ significantly in comparison with BrE, and that there were more unstressed syllables than weak syllables in the HKE data. In this paper, the material from Setter (2006) is re-examined using the Pairwise Variability Index (PVI) (Low, Grabe and Nolan 2000), in this case applied at the level of the syllable (sPVI); a similar conclusion is reached using the sPVI. In addition, the PVI is applied at the level of the syllable peak to new HKE data (nPVI), and findings compared with Low *et al.*'s existing study of British and Singapore English. The HKE data are shown to be more similar to the Singapore English data than the BrE data, thus reinforcing the findings of the 2006 study and the re-examination of that data. However, great individual speaker variation is found.

The paper also examines some features of tonicity and tone in HKE speech data, collected using communicative tasks.

Keywords: Rhythm, intonation, prosody, PVI.

1. INTRODUCTION

Considering prosodic features of English from a traditional English language teaching (ELT) perspective, Taylor (1981) describes English speech rhythm as one of the most difficult aspects for foreign learners of English to acquire, and goes as far as to say that intonation is “not teachable, and possibly not learnable either” (Taylor 1993: 2). Roach, in his discussion of intonation in English, suggests that classroom teaching may not be effective in learning some aspects of intonational meaning when he concludes that “the attitudinal use of intonation is something that is best acquired through talking with and listening to English speakers” (2009: 151).

Though often viewed as problematic from a teaching point of view, in recent years there has been an increasing body of research on the subject of prosodic features of various types of English, both from learners and from speakers of varieties of World Englishes. This paper starts by looking at speech rhythm in Hong Kong English before going on to examine some features of intonation in that variety. The results are considered in terms of the possible implications for the use of English in international contexts, and their contribution to classroom practice.

2. SPEECH RHYTHM IN HONG KONG ENGLISH

Traditionally, varieties of English belonging to Kachru’s (1982) Inner Circle – e.g. British, North American and Australian English – are said to have a stress-timed rhythm. However, varieties which belong to the Outer Circle, such as Singapore English (SE), often display a different pattern, and could be described as having a syllable-timed rhythm. Indeed, studies such as that of Low, Grabe and Nolan (2000) have demonstrated that the rhythmical properties of SE tend more towards syllable-timing than stress-timing.

Setter (2006) applied a pedagogically motivated hierarchical approach to the study of speech rhythm in HKE and BrE. Essentially, and using categories derived from Roach (2009), the duration of tonic, stressed, unstressed and weak syllables in HKE and BrE corpora was measured and compared statistically. It was found that, although speakers of HKE retained a statistically significant difference in duration between the four syllable types, when compared to BrE, this difference was not as great. In addition, the largest amount of syllable types in the HKE was unstressed, whereas in the BrE data, weak syllables were more numerous. This lack of deprominencing (see Low *et al.* 2000), it was hypothesised, together with the lesser difference in syllable type duration, could lead to intelligibility difficulties between HKE and BrE speakers. Figure 1 shows the relative durations of HKE and BrE, and Figure 2 shows the proportion of syllable types in the study.

Figure 1: Line plot of syllable duration according to stress level for HKE and BrE (Setter 2006).

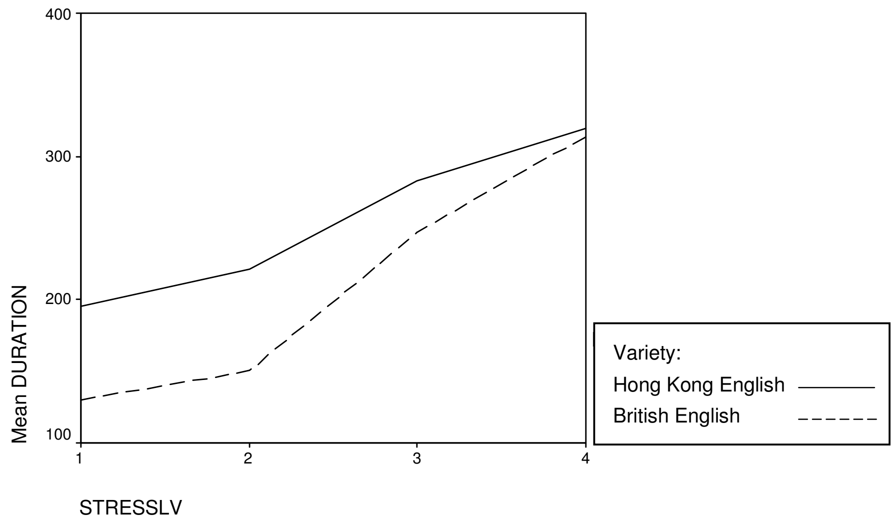
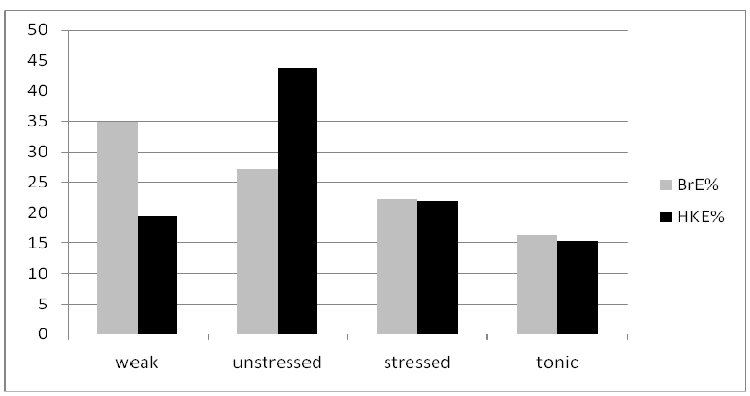


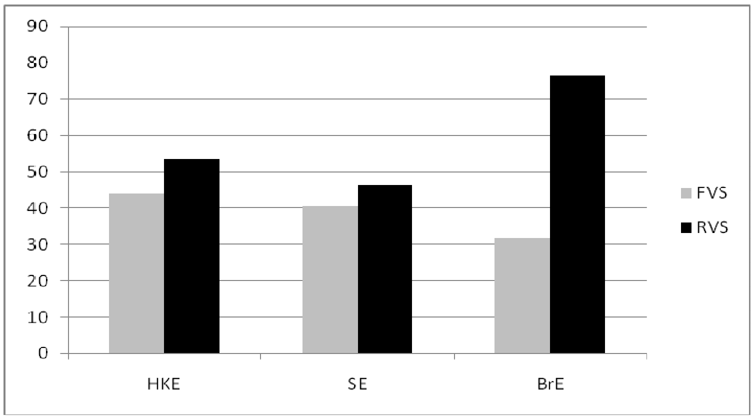
Figure 2: Proportion of weak, unstressed, stressed and tonic syllables in HKE and BrE (Setter 2006).



This hierarchical methodology is pedagogically motivated in that teachers of English language often know and understand the idea of different levels of stress in syllables, and the effect this can have on e.g. syllable duration and pitch. Low *et al.* (2000) used a metric known as the Pairwise Variability Index (PVI or nPVI), which compares the successive duration of syllable nuclei using a mathematical formula. In order to do this, two sets of sentences were prepared, one containing weak and strong syllables (e.g., *Grace was tired of Matthew Freeman*), referred to as the reduced vowel set or RVS, and the other containing all strong syllables (e.g., *Grace works through huge mounds each Friday*), referred to as the full vowel set or FVS. The data collected from SE and BrE speakers producing these sentences were then compared. SE scored 46.37 and 40.37 for the RVS and FVS respectively, and BrE scored 76.57 and 31.79 for the RVS and FVS respectively (Low *et al.* 2000). The high score of 76.57 for the RVS in BrE indicates greater variability in syllable nuclei duration in sentences of this type, whereas the low score of 31.79 shows sentences containing all stressed syllables are produced with much less difference. This is in contrast with the SE group, whose speech rhythm differed little in their production of the RVS and FVS sentences, and were more towards the FVS end of the nPVI for the BrE speakers in both sets. Where does HKE sit on this scale?

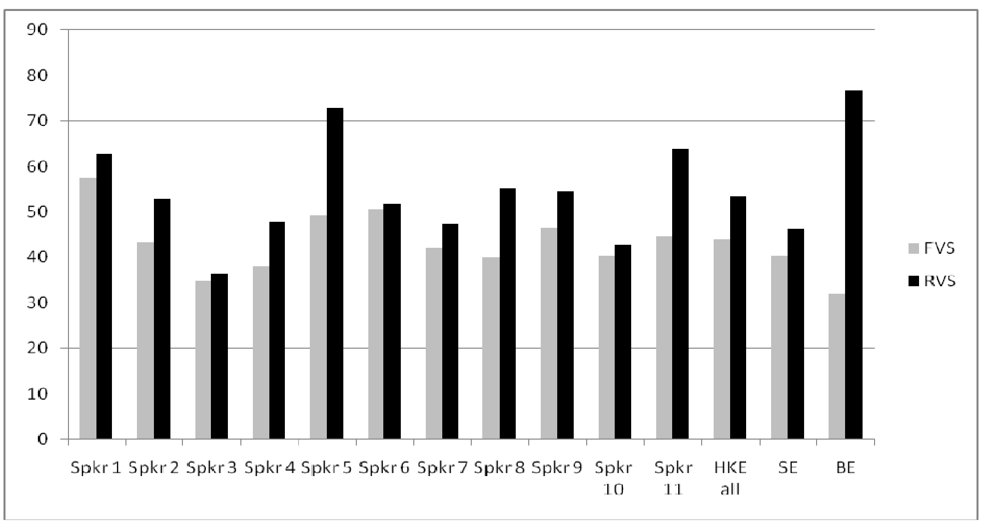
Data were collected from 11 speakers of HKE reading the RVS and FVS sentences from Low *et al.*'s study, and the nPVI was applied to the syllable nuclei in these sets. The results show that HKE has an nPVI of 53.3 for the RVS and 43.3 for the FVS. This indicates that there is greater variation in HKE than there is in SE, but still nothing of the magnitude of difference seen in the BrE data. Figure 3 shows the results in visual form.

Figure 3: nPVI in HKE, SE and BrE.



However, there was a great deal of speaker variation in the HKE data. Figure 4 shows a breakdown for each of the speakers, with the averages for the HKE, SE and BrE given at the end.

Figure 4: Individual scores for nPVI in HKE, compared with averages for HKE, SE and BrE.



From this graph, one can see that some of the speakers showed very little difference in the rhythmical properties of the RVS and FVS sentences (e.g., Speakers 3, 6, 7 and 10), whereas others showed a lot of difference (e.g., Speakers 2, 4, 5, 8 and 11). This points to the instability of speech rhythm in the variety, and could also indicate that HKE cannot as yet be regarded as a variety in a similar way to SE, if there is a lack of homogeneity amongst speakers. However, as Low *et al.* (2000) did not provide a breakdown for individual speakers in their study, it is difficult to draw any further comparisons.

I was also curious to apply the PVI to the HKE data from my original study (Setter 2006), comprising 4404 syllables of connected speech. As the data were analysed at the level of the syllable (hence sPVI), however, and were in the form of semi-scripted student presentations, they are very different from those prepared for the replication of Low *et al.*'s study above. I used an online PVI calculator (see http://vesicle.nsi.edu/users/patel/npvi_calculator.html), which derived an sPVI of 49.78 for the HKE data, and 67.93 for the BrE data used as a comparison in that study (1847 syllables). This is not dissimilar from the average results for the variety shown above, and supports the findings of the hierarchical method first used.

In conclusion, it appears that HKE is closer to SE rhythmically than it is to BrE, and can be regarded as nearer the syllable-timed end of any stress-timed/syllable-timed continuum.

3. INTONATION IN HONG KONG ENGLISH

Until recently, there has been very little work indeed on pitch and nuclear tones in HKE. Bolton and Kwok (1990), for example, only have a paragraph on suprasegmental features in the variety. They assert that speakers tend to use a rising intonation on all question types, including wh-questions, and show in their displays of the pitch contour that statements often have a fall on them (Bolton and Kwok 1990: 154-160). Most other works on HKE do not treat this area at all. The exception is Cheng, Greaves and Warren (2008), which is a comprehensive non-acoustic study of discourse intonation in a corpus of HKE. Here, I make look at some of the patterns which arise in data collected for Setter, Wong and Chan (forthcoming), and caution that this is not intended to be a full treatment of pitch patterns in the variety.

Our data were collected from 11 speakers of HKE using two communicative tasks, a Map Task (for details see http://www.hcrc.ed.ac.uk/maptask/) and discussion of a happy event. Collection took place at the same time as the first PVI study reported here. The initial finding was that the tonic syllable in the HKE tone units tends to be (but is not exclusively) on the last word. On the whole, the speakers use a range of nuclear tones in their speech, including all those covered in Cruttenden (2008): high fall and low fall; high rise and low rise; fall-rise; rise-fall; and level. In our data, the highest incidence tone, possibly owing to the amount of hesitation, was the level tone (43.32%) – as observed also in Cheng *et al*.'s corpus (2008: 126) – followed by the rise (24.39%), the fall (23.02%), the fall-rise (8.68%) and finally the rise-fall (0.59%).

The rise-fall, a strongly dominant tone which in BrE can indicate indignation, sarcasm or, conversely, being surprised or very impressed by something (Cruttenden 2008: 284), is very low incidence in the data, as it was in Cheng *et al.*'s corpus. In the small number of instances in which it is used in our recordings, it has none of the meanings associated with it by Cruttenden (2008). In SE, the rise-fall is often used to indicate extra emphasis (Deterding 2007: 37); Speaker 5 could be indicating extra emphasis in (1), but this explanation seems unlikely in Speaker 10's extract (2):

(1) and then ... er after that you can see the finish point ... on your right and you can see me ↗↘there {Spkr 5}

(2) and ↗↘then ... um ... on your left hand side {Spkr 10}

There is an interesting use of intonation by several speakers on the phrase *and then*, in which is it produced with a rising pitch on *then* when used as a lead into another phrase or sentence. This seems to be a phrasal pattern specific to this sequence of words. Speakers 9 and 10 use this pattern quite regularly in the Map Task recordings.

(3) and ↗then you turn to ... er then you go east {Spkr 9}

(4) yeah and ↗then you will you will be ah at the foot of the mountain and ↗then you keep going north and just um when you see ... ah you you pass the mountain and ↗then you turn east again {Spkr 9}

(5) and ↗then you will possibly see a field station {Spkr 9}

(6) and ↗then keep going {Spkr 10}

(7) and ↗then um ... in the in in the direction of north of you {Spkr 10}

This pattern, however, may be a general feature of South East Asian Englishes; data collected recently on Malay English show a similar pattern on the phrase *and then*.

In some accents of English, including Australian, New Zealand and among younger speakers in the UK, there is a tendency towards a phenomenon known as 'upspeak' or 'uptalk' (see Bradford 1997; Cruttenden 1997: 129-130; Wells 2006: 37-38). This is where the intonation rises at the end of declaratives (shown by '↗') rather than falling (shown by '↘'). Cruttenden (2008) suggests that this might be to ensure the listener is paying attention, because a rising tone often requires some kind of response from the listener. There are examples of up-speak in our data from all five speakers. Here are examples of declaratives with a falling tone and a rising tone for Speakers 1, 5 and 8. In the first example (8), we can also see Speaker 1 using rising tones to indicate he has not finished his turn yet.

(8) Spkr 1: when I was a ↗kid erm around the age of erm ↗ten
Int: ten years
Spkr 1: eleven around and mmm I actually I studied in er in the same primary school with my ↘brother

(9) I I was just being you know so fascinated about you know ... to to be a ↗kid again I'll say you know in a way like compared to ↗now {Spkr 1}

(10) because I used to study in a very prestigious school in Hong ↘Kong {Spkr 5}

(11) well I got a quite interesting ↗childhood {Spkr 5}

(12) let me talk about my trip to ↘Canada {Spkr 8}

(13) and we ... spent about four days there because I've never been to Vancouver be↗fore {Spkr 8}

Another interesting phenomenon is the lack of compound stress in the variety; the main stress in a compound falls on the last word, which is similar to the pattern found in premodifier + noun combinations in varieties such as BrE. There are several noun compounds in the data collected, amongst them *diamond mine* and *banana tree*, which are stressed as follows in RP: *DIAmond mine*; *baNAna tree*. Here are some examples from the data, with the main stress in capitals.

(15) erm can you see a diamond MINE next to you {Spkr 5}

(16) mmm first you have to go north and on your left hand side you will see a diamond MINE {Spkr 8}

(17) in front of you can you see the diamond MINE {Spkr 10}

(18) can you can you see um ... a banana TREE {Spkr 5}

(19) follow that road and you will see the banana TREE {Spkr 8}

(20) you can see a banana TREE ... can you see it {Spkr 10}

In addition, there is a lack of de-accenting of repeated items in the data. Here is an example from Speaker 5, in which she fails to de-accent *point*, even though she has referred to the *highest viewpoint* before:

(21) Spkr 5: can you see a highest viewpoint
Int: yes there is one yes there's one on ... er well towards the north east direction of where I am […]
Sprk 5: that's right so um ... um the springboks actually is before the highest viewPOINT on your left […]
Int: so where shall I go now
Sprk 5: and then um you should um go past the highest viewPOINT fr- from the back of it

In this section, I have illustrated a few patterns which are found in HKE intonation, but clearly there is much more work to do on the data collected.

4. DISCUSSION

The patterns described here are of interest for various reasons, not least because they could be indicative of an emerging variety of English. However, it is important to guard against issues which may lead to unintelligibility in communicative settings. Jenkins (2009), among others, points out that English as a lingua franca is not used solely for speakers to communicate with monolingual English speakers (MES) from places such as Britain and America, even though those two varieties still dominate as models in ELT materials; the communicative and pedagogical goal, therefore, must be to ensure speakers of HKE can be understood in a variety of settings, and that this is not adversely affected by strongly variety-specific patterns of speech rhythm and intonation. As I have illustrated, HKE is more similar to SE in terms of speech rhythm, and it may be assumed that other varieties in the region also have similar rhythmic patterns, and this may well facilitate communication. In communicating with MES, however – and there is still quite a lot of call for such communication in Hong Kong – it may be necessary for both sets of interlocutors to move a little towards the other.

5. REFERENCES

Bolton, K., Kwok, H. 1990. The dynamics of the Hong Kong accent: social identity and sociolinguistic description. *Journal of Asian Pacific Communication* 1(1). 147-172.

Bradford, B. 1997. Upspeak in British English. *English Today* 51(13.3). 29-36.

Cheng, W., Greaves, C., Warren, M. 2008. *A corpus-driven study of discourse intonation*. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Cruttenden, A. 1997. *Intonation*. (2nd ed.). Cambridge: Cambridge University Press.

Cruttenden, A. (ed.). 2008. *Gimson's pronunciation of English*. (7th edn.). London: Edward Arnold.

Deterding, D. 2007. *Singapore English*. Edinburgh: Edinburgh University Press.

Jenkins, J. 2009. *World Englishes*. (2nd ed.). Abingdon: Routledge.

Kachru, B. B. 1982. *The other tongue: English across cultures*. Oxford: Pergamon Press.

Low, E. L., Grabe, E., Nolan, F. 2000. Quantitative characterisations of speech rhythm: syllable-timing in Singapore English. *Language and Speech* 43(4). 377-401.

Roach, P. 2009. *English phonetics and phonology: A practical course*. (4th ed.). Cambridge: Cambridge University Press.

Setter, J. 2006. Speech rhythm in World Englishes: The case of Hong Kong. *TESOL Quarterly* 40(4). 763-782.

Setter, J., Wong, C. S. P., Chan, B. H. S. Forthcoming. *Hong Kong English*. Edinburgh: Edinburgh University Press.

Taylor, D. S. 1981. Non-native speakers and the rhythm of English. *International Review of Applied Linguistics (IRAL)* 19(3). 219-226.

Taylor, D. S. 1993. Intonation and accent in English: what teachers need to know. *International Review of Applied Linguistics in Language Teaching* 31(1). 1-22.

Wells, J. C. 2006. *English intonation: An introduction*. Cambridge: Cambridge University Press.

Discovering the Relationship Between Context and Allophones in a Second Language: Evidence for Distribution-Based Learning

Christine E. Shea

Department of Linguistics, University of Calgary
ceshea@ucalgary.ca

ABSTRACT

This study examined the identification of stressed syllables in CVCV nonce words by beginner and intermediate L1 English/L2 Spanish learners to see if it is influenced by an allophonic alternation driven by word position and stress. The allophones utilized were the Spanish voiced stop-approximant alternation (bdg~βðɣ), where stops occur in word onsets and stressed syllable onsets. In Experiment 1, allophone onset and vowel stress were crossed while in Experiment 2, only the allophone onset alternated. More experienced groups were predicted to perceive stress on stop-initial syllables with greater likelihood than approximant-initial syllables, following the Spanish distributional information. In other words, these listeners would perceive a 'stress illusion', induced by the onset allophone. Results confirmed this, showing that the Spanish proficient group was more likely to perceive stress on syllables with stop allophones than the lower proficiency group. This suggests that learning the interplay between allophonic distributions and their conditioning factors is possible with experience. This knowledge is distributional in nature, suggesting that second language learners use a statistical mechanism in acquisition. In order to account for this, we adapt the PRIMIR-bilingual framework (**P**rocessing **R**ich **I**nformation from **M**ultidimensional **I**nteractive **R**epresentations; Curtin, Byers-Heinlein & Werker, in press) to adult second language acquisition.

Keywords: distribution-based learning, allophones, L2 phonology, Spanish, PRIMIR

1. INTRODUCTION

Studies of L2 speech perception have primarily explored how target language sounds fit into the sound system of the speaker's native language, in particular, whether non-native sounds represent new categories, are classified into existing native-language phoneme categories, or if they are similar to existing allophones. In the experiments discussed here, the acquisition task is approached from a slightly different perspective. Instead of focusing on the acquisition of new sound categories and how L2 allophones assimilate into the native language sound inventory, I examine whether learners are sensitive to the contextual factors found in the phonological environment that condition the allophones' distribution. In other words, do learners recognize and store information about the specific context in which each variant occurs? I examine the acquisition of the stop-approximant alternation in Spanish by L1 English speakers. The occurrence of either the stop or approximant allophone is contingent upon the phonological environment – where in the word it occurs and whether the syllable is stressed. Given this, it was posited that more experienced learners use stress and word position as probabilistic cues to the stop-approximant alternation. Learners with greater language experience are predicted to be aware of which factors condition the allophonic alternation and this awareness will shift their perception of the target language. Thus, the goal of the experiments presented here is to assess how L1 English/L2 Spanish learners perceive and make use of the conditioning factors driving the stop-approximant alternation in their target language.

One way learners might do so is by means of a distribution-based learning mechanism. Researchers have shown that both adult and infant listeners are able to form categories based on the distributions of speech sounds and shift their perception of allophones by using this type of distribution-based mechanism (e.g. Maye & Gerken, 2001; Maye, Werker & Gerken, 2002). By their very nature, distribution-based models assume learners create and store highly detailed, rich, exemplar-type representations and grammar emerges as the result of generalizations across all the stored items in the lexicon (see, e.g., Goldinger, 1997, Johnson, 1997; Pierrehumbert, 2001, 2003). In the present case, the learning task for L1 English learners of Spanish involves creating a new distribution for the approximant category, separate from the Spanish stop category that is very similar to their L1 category.

If adult second language learners track distributional information in the input, they will expect to hear more stops in word initial, stressed position than word medial unstressed position. Thus, learners will be more likely to perceive stress when the stressed syllable is accompanied by a stop consonant and less likely to perceive a syllable as stressed when it begins with an approximant. In Experiment 1, listeners heard CVCV nonce words, crossed for allophone onset and stressed/unstressed vowels to determine if the perception of stress shifts according to the allophone onset. Following the predictions stated above, learners with knowledge of the relationship between phonological environment and allophones will be more likely to select a syllable as stressed if it begins with a stop consonant and has a stressed vowel, than a syllable with an approximant onset and stressed vowel. In Experiment 2, the vowel was equated for stress and only the syllable onsets alternated between stops and approximants. That is, stress is not explicitly present in the signal, but rather inferable from the presence of a stop onset, providing that the listener is sensitive to the distributional information connecting stress and stops in the Spanish input. Listeners with more Spanish experience should select stop-onset syllables as stressed with greater likelihood than groups with less Spanish experience, given their increased knowledge of Spanish distributional information. This indirect behavioural method circumvents problems with phonetic vs. phonological representations and also arrives at the key question motivating these experiments: are learners aware of the contextual factors that drive allophonic alternations in their target language?

2. EXPERIMENT 1: CONSONANT AND VOWEL STRESS SHIFT

2.1. Method

2.1.1. Participants

Participants were 15 Low Intermediate and 15 High Intermediate L1 English/L2 Spanish learners, recruited from second and third-year university-level Spanish classes. No participant from either group had spent more than six weeks in a Spanish-speaking country and none spoke Spanish outside of the classroom context. Fifteen Native Spanish speakers (NSS) were recruited from the Center for the Teaching of Foreign Languages at the National Autonomous University of Mexico, (CELE-UNAM) in Mexico City, Mexico. None of the participants had ever lived abroad, none had attended a bilingual school nor did they have more than three hours per week of contact with English. Finally, 15 Monolingual English speakers (ME) were recruited from a university psychology subject pool and were also age-matched with the two learner groups. None of these participants spoke any language other than English. All participants were paid the equivalent of $15.00 for their time or, in the case of the Monolingual English speakers, received course credit.

2.1.2. Stimuli

The stimuli were created from CVCV non-words, with first syllable stress, taken from naturalistic speech samples recorded by a native female speaker of Spanish from Mexico City. The consonants were [b], [d] or [g] and the vowel was [a]. Using PRAAT 5.1 (Boersma & Weenink, 2008), the consonants were spliced from the vowels to create four separate sounds: stop (word onset), approximant (second syllable onset), stressed vowel and unstressed vowel. For example, the nonce word ‘baba’ [báβa] provided four separate segments; [b], stressed [a], [β] and unstressed [a]. These four sounds were combined to create four different tokens: [báβa], [βába], [baβá] and [βabá]. This procedure was repeated for both [d] and [g], creating a total of 12 tokens. Stimuli ranged in length from 67ms to 78ms. All stimuli were presented to two native English speakers and two native Spanish speakers and judged for naturalness on a scale of 1 (natural-sounding) -5 (artificial sounding). Stimuli that did not originally receive a rating of 4.5 or higher was re-spliced and presented to the judges again.

2.1.3. Procedure

All participants first completed a stress detection task, which involved listening to a series of 20 nonce words which which followed the phonotactic requirements of Spanish. Participants indicated by means of pressing keys on the computer keyboard whether they thought stress fell on syllable ‘1’ ‘2’ or ‘3’. Only participants who obtained at least 75% accuracy on this task had their results included for analysis. Participants were instructed to select the syllable they perceived as stressed by means of pressing a key on the computer keyboard. The keys were marked with a sticker indicating either ‘1’ or ‘2’. Subsequent tokens were played after the participant made their selection, with an ISI of 1000ms, and timed out after 1500ms.

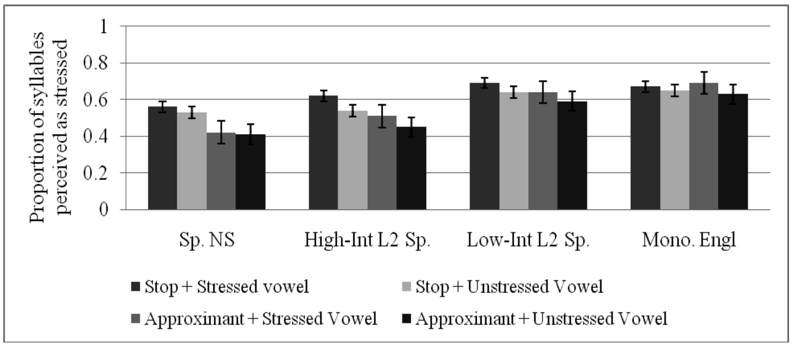
2.2. Results

Recall the prediction that stress perception for low-level learners would not be affected by the allophone in the onset position of the syllable and instead would only be affected by the vowel. Low-level learners are predicted to perceive stress in accordance with the prominence of the vowel. To test this, a one-way ANOVA was carried out to determine whether there were overall differences among the groups in terms of connecting stressed vowels to one or the other allophone. Group was the independent variable and for the dependent variable, a ratio value was calculated as follows: stops + stressed vowel perceived as stresse***d/*** approximants + stressed vowel perceived as stressed**.** The Native Spanish and High Intermediate groups are predicted to have ratios greater than 1, indicating more syllables with initial stops and stressed vowels were perceived as stressed than syllables with initial approximants and vowel stress. For the Low Intermediate and Monolingual English groups, the ratios are predicted to be around 1, indicating a lack of preference for either allophone.

A one-way ANOVA was conducted on the ratio values. The results were significant [F(3,56)=30.85, p <0.001]. The groups with less Spanish experience had ratios that were close to 1 (Low Intermediate: M=1.3, SD=.25; Monolingual English: M=.95,SD =.15) while the ratios for the two groups with more experience were significantly greater (Native Spanish: M=2.78, SD=.122; High Intermediate: M =2.1, SD=.7). Tukey's HSD post hoc tests showed significant differences among the Native Spanish speaker group and the two lower proficiency groups, and between the High Intermediate group and the two lower proficiency groups (all ps<0.05), but not between the High Intermediate and Native Spanish speaker groups. These results show that the more experienced groups perceived stress significantly more often on stop-initial syllables than approximant-initial syllables, suggesting listeners with more Spanish experience associate stress with the stop allophone.

I next conducted a goodness-of-fit chi-square test on the proportion of syllables perceived as stressed for each of the four possible onset-vowel combinations, reported for each group. If in fact the groups with more experience prefer stops as onsets to stressed syllables, there should be a difference amongst the four combinations, with the different allophone-types clustering together for the more proficient learners and the stressed-unstressed vowel factor clustering together for the less proficient learners. These results are presented in Figure 2:

Figure 1: Proportion of syllables perceived as stressed



For the Native Spanish speaker group, preference for syllable stress was not equally distributed, χ^2 (3, N=176) = 4.21, p<0.05. For the High Intermediate group, the same results held, χ^2 (3, N=174) = 7.22, p<0.05. For the Low Intermediate and Monolingual English speaker groups there were no significant differences across the four contexts (all ps>0.05). The finding that language experience led to significant differences in the perception of stress across the four contexts shows the pivotal role played by this variable in terms of how the allophone drives stress perception in Spanish.

It is possible, however, that the selection of stressed syllables is also influenced by knowledge of the predominant stress pattern in English and Spanish, which is trochaic (Alameda & Cuetos, 1995; Cutler & Carter, 1987). I calculated the proportion of first syllables perceived as stressed, independent of vowel prominence, followed by the proportion of trials perceived as initial stress with stop allophones. While it is possible that all groups potentially demonstrate a trochaic bias for CVCV forms, I predict that only the groups with greater Spanish experience will show a preference for syllables with stops over approximants in trochaic contexts. The less experienced groups are predicted to be around chance (0.5). To permit adequate comparisons among the groups, I calculated ratio values (proportion stop allophoneσ1+stressed vowel /proportion approximant allophoneσ1+stressed vowel) and carried out a one-way ANOVA with groups as the independent variable and the ratio values as the dependent variable. There was no significant effect for stress detection on the first syllable ($F[3, 55]= 1.97$, $p>.05$). A second one-way ANOVA was conducted with stop-onset syllables in initial position as the dependent variable (also a ratio). The results were significant ($F[3,55]=22.03$, $p<0.001$). Tukey HSD post hoc tests revealed significant differences between the Native Spanish speaker group and the other three groups ($p<0.001$), but no significant differences emerged among the High Intermediate, Low Intermediate and Monolingual English groups (all ps >0.05). These results show that all four groups of listeners show a bias towards hearing trochaic stress patterns, but the Native Spanish speakers demonstrate an additional bias towards perceiving stress on syllables that have stop onsets, consistent with the Spanish distributional information.

3. EXPERIMENT 2: ALLOPHONE ALTERNATION, VOWEL STEADY

3.1. Method

3.1.1. Participants

The same participants from Experiment 1 took part in Experiment 2.

3.1.2. Stimuli

The stimuli for this experiment consisted of CVCV nonwords, created from the same naturalistic speech samples as Experiment 1. However, vowel [a] was held steady and only the consonant onsets were alternated. A stressed vowel token was taken from the CVCV stimuli used in Experiment 1 and the intensity was adjusted to 75dB. The F1 value was 806Hz and F2 was 1628Hz and the duration was 74ms, for example [γ]A[g]A. Stimuli ranged in length from 171ms to 201ms. Finally, the consonants were spliced onto the vowel and counterbalanced for allophone variant. The place of articulation was held constant within each CVCV sequence. As with the first experiment, only stimuli rated 4.5 or higher were used for the experiment.

3.1.3. Procedure

The same procedure as in Experiment 1 was used.

3.2. Results

A one-way ANOVA was run to determine whether participants perceived stress in higher proportions on stop syllables or on approximant syllables. Group was the independent variable and the following ratio measure was the dependent variable: stop-initial syllables perceived as stressed/approximant-initial syllables perceived as stressed. There was a significant difference among the means ($F[3, 55]=20.18$, $p<0.001$). Post hoc Tukey HSD tests revealed significant differences between the Native Spanish speaker group and the other three ($p<0.01$) and the Monolingual English group and the other three groups ($p<0.01$). There were no significant differences between the High Intermediate group and the Low Intermediate group.

As with Experiment 1, responses were examined to see if there was for a bias for perceiving stress on first syllable of the word. To permit adequate comparisons among the groups, I calculated ratio values (proportion stop allophone$_{\sigma 1}$/ proportion approximant allophone$_{\sigma 1}$) and carried out a one-way ANOVA. The results did not reach significance ($F[1,55]=1.8$, $p>0.05$), possibly because the two higher-proficiency groups clustered together, as did the two lower-proficiency groups. The mean ratio for the Native Spanish speakers was 1.5 (SD = 0.6) and for the High Intermediate group it was 1.5 (SD= 0.7) demonstrating that these participants prefer to associate stress with stop syllable onsets. For the Low Intermediate group the mean was 1.1 (SD= 0.37) and for the Monolingual English speakers, the mean was .99 (SD=0.4) suggesting that participants in these groups did not distinguish between stop and approximant onsets as they related to stress.

The results from Experiment 2 indicate that with more Spanish experience, the onset allophone – whether stop or approximant- can lead to an illusory stress perception effect. The Native Spanish speaker group heard stress significantly more often on stop-initial syllables than on approximant-initial syllables as compared to the other three groups and the Monolingual English group perceived stress significantly less often than the other three groups when the syllable had a stop in onset position. These results further suggest that with increased Spanish experience, L1 speakers of English perceive an illusory stress effect, induced by the onset allophone in bisyllabic nonwords. Learners associate the stop allophone with stress and the approximant allophone with absence of stress, but only after considerable experience with Spanish.

4. GENERAL DISCUSSION AND CONCLUSIONS

In these experiments, I investigated whether L2 learners connect each allophone to its expected phonological environment and if so, whether language experience plays a role. The results suggest that learners are able to track the distribution of the allophones and over time, they begin to learn the relationship between the allophones and the contexts in which they surface. One possible way to explain how L2 learners acquire allophones is through distributional learning. As expected, based on the predominant pattern for main stress in both Spanish and English, all four groups showed a preference for perceiving stress on the first syllable. However, upon closer examination, the bias in favour of stop initial, stressed syllables only occurred with the Spanish-proficient groups. This suggests that participants in the Native Spanish and High Intermediate groups have acquired knowledge about the distribution of these allophones. In particular, these listeners have connected the phonological environment of stress to a stop onset and lack of stress to an approximant onset. This shows that experience with Spanish can actually shift the perception of stress in non-native speakers in the direction of the distributional information found in their target language.

Distribution-based learning mechanisms play a fundamental role in exemplar-based models of phonological acquisition. Under an exemplar-based model, such effects arise when listeners rely upon information they have stored and probabilistically draw upon when exposed to input. The experienced Spanish listeners have representations that probabilistically associate stress with stop onsets. Their perception is biased towards perceiving stops and stress, yielding phonotactic sequences that are highly probable in Spanish. They are biased against hearing stress on approximant-initial syllables for the same reason. The groups with less Spanish experience have not built up sufficiently dense representations and are thus not biased in one direction or the other. This could be part of the explanation for the results from Experiment 2, where more Spanish-proficient listeners demonstrated an allophone-induced ‘stress illusion’ (see Dupoux, Kakehi, Hirose, Pallier & Mehler, 1999 for similar effects in Japanese L1 listeners). As language experience increases, listeners are more affected by the contextual cues, or conditioning factors that drive the allophonic alternation. Specifically, knowledge of probabilistic, distribution-based information allowed more advanced learners to recognize the factors that condition the allophonic alternation. In the present experiment, context effects – i.e., the onset allophone – actually shifted the perception of stress in the learners with greater Spanish proficiency. Lower-proficiency learners did not demonstrate any such effects. This suggests that adult L2 speech perception shifts over time and becomes sensitive to the phonotactics of their target language.

To summarize, the results of this study suggest that adult second language learners use contextual information in their acquisition of target language allophones: the perception of stress was conditioned by the onset allophone and the position in the word, as a function of language experience. In a broader sense, these results point to the availability of a distribution-based mechanism for adult second language acquisition and further suggest that language experience plays a strong role in how exactly this mechanism is used over the time-course of second language acquisition.

In order to account for these results, we adapt the PRIMIR-bilingual framework (Processing Rich Information from Multidimensional Interactive Representations; Curtin, Byers-Heinlein & Werker, in press) to adult second language acquisition. The bilingual extension to PRIMIR posits that learners use a statistical learning mechanism to form the sound categories of their native language and in the case of bilingual or adult second language acquisition, the statistical mechanism is complemented by an additional mechanism that allows learners to compare and contrast the incoming input and track statistics independently across the learner’s languages. Thus, learners develop two different sets of statistics for each language being acquired. In the case of L1 English/L2 Spanish learners, acquisition of the stop-approximant alternation requires tracking where each allophone occurs with respect to stress and word position. The learner hears the incoming allophone and compares it to previously stored exemplars of that category and when the contrast mechanism reveals that the token cannot be classified with previously experienced tokens, the learner creates a new category. This new category contains nested information regarding position and prosodic characteristics of the context.

This process requires extensive experience with the target language in order to build sufficiently robust representations. The fact that our advanced level learners were able to 'hear' stress more often on syllables that started with stops than those with approximants supports the argument that such a statistical mechanism of comparison and contrast is operational in adult second language acquisition. These results provide evidence for a phonological system capable of tracking distributional information in the speech stream. Furthermore, this distributional knowledge is gradually accumulated, as shown by the different effects for the contextual factors across distinct levels of experience with Spanish: listeners with greater Spanish experience demonstrated an illusory effect for stress, induced by the presence of a stop allophone in syllable onset position. These results speak directly to how contextual factors drive listener expectations regarding the allophone alternation and suggests that learner representations encode statistical information such as co-occurrence likelihoods.

5. REFERENCES

Alameda, J.R., & Cuetos, F. (1995). *Diccionario de frecuencias de las unidades lingüísticas del castellano*. Oviedo, Spain: University of Oviedo Press.

Boersma, P., & Weenink, D. (2009). Praat: doing phonetics by computer (Version 5.1.) [Computer program]. Retrieved February 2, 2009, from http://www.praat.org/.

Curtin, S., Byers-Heinlein, K., & Werker, J. (under review). Bilingual beginnings as a lens for theory development: PRIMIR in focus.

Cutler, A., & Carter, D.M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language, 2*, 133–142.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C. & Mehler, J. (1999). Epenthetic vowels in Japanese: a perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance, 25*, 1568-1578.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology:Learning, Memory, and Cognition, 22*(5), 1166–1183.

Johnson, K. (1997). Speech perception without speaker normalization. In K.Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 145–165). San Diego: Academic.

Maye, J., & Gerken, L. (2001). Learning phonemes: how far can the input take us? In A.H.-J. Do, L. Dominguez & A. Johansen (Eds.), *Proc. 25th Annual Boston University Conference on Language Development* (pp. 480–490). Somerville: Cascadilla Press.

Maye, J.,Werker, J. & Gerken, L. (2002) Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*, B101-B111.

Pierrehumbert, J.B. (2001). Stochastic phonology. *Glot International*, 5(6), 195–207.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech, 46*, 115-154.

Nativelike Attainment in L2 Listening: The Segmentation of Spoken French

Ellenor Shoemaker

University of Texas at Austin
Université Paris 13
ellishoe@gmail.com

ABSTRACT

French is a language that poses particular difficulties for the L2 learner in the processing of continuous speech. The phonological processes of *liaison* and *enchaînement* (resyllabification) can render syllable and word boundaries ambiguous (e.g., *un air* 'a melody' and *un nerf* 'a nerve', both [œ̃.nɛʁ]). Some research has suggested that speakers of French give listeners acoustic cues to word boundaries by varying the duration of liaison (e.g., /n/ in *un air)* and initial consonants (e.g., /n/ in *un nerf*) and that access to mental representations in the lexicon is facilitated by these cues (e.g., Spinelli et al., 2003); however no study to date has directly demonstrated that durational differences are exploited in the online segmentation of speech.

One way to directly test the exploitation of duration as a parsing cue by both native and non-native speakers is to manipulate and exaggerate this single acoustic factor while holding all other factors constant. To this end, the current study employed ambiguous French phrases in which the pivotal consonants (i.e. /n/ in *un air/nerf*) were instrumentally shortened and lengthened while the rest of the phrase remained unaltered. Eighteen native speakers of French and 18 advanced late learners of L2 French were tested on a forced-choice identification task incorporating these manipulated stimuli. Results suggest that duration alone can indeed modulate the lexical interpretation of sequences rendered ambiguous by liaison in spoken French. In addition, six out of 18 non-native participants scored at or above the native mean, suggesting nativelike sensitivity to non-contrastive phonological variation in a L2.

Keywords: speech segmentation, spoken word recognition, allophonic variation, L2A

1. INTRODUCTION

A substantial body of research has established that acoustic and phonological cues to speech segmentation are not exploited to the same extent and in the same manner cross-linguistically (e.g., Cutler et al., 1989; Cutler & Norris, 1988; Sebastian-Galles et al., 1992). Segmentation cues differ from language to language and are thus assumed to pose problems for the segmentation of a second language (L2). Thus while native segmentation strategies render speech perception in one's first language (L1) automatic and effortless, the ease of L1 speech processing stands in sharp contrast to the conscious effort that can be required in the aural comprehension of a L2. Research dealing with specific cues to speech segmentation such as phonotactics (Weber, 2001) and prosody (Cutler et al., 1989; Dupoux et al., 1997) has suggested that L2 learners are constrained by L1 segmentation routines.

Research on the notion of a critical period for language learning has attributed the discrepancy between native and non-native language processing to a post-pubescent pruning of phonological sensitivity that leads to perceptual deficiencies for those who undertake the study of a L2 later in life. Many researchers hold that this decline in sensitivity leads to a *perceptual foreign accent* (Strange, 1995) that leaves late learners with possibly insurmountable deficits in the perception of L2 phonological systems (for a review of research on *non-native listening* see Cutler 2001). However, more recent research has suggested that learners can not only suppress the use of L1 segmentation strategies in the processing of an L2 (Cutler et al., 1997), but can acquire and implement novel L2 segmentation routines as well (Golato, 2002), challenging strong claims of limitations on the plasticity of phonological learning and perceptual processing in adult learners.

2. SPEECH SEGMENTATION IN FRENCH

The current study examines the exploitation of fine-grained acoustic detail as a segmentation cue by adult learners in the processing of L2 French. French is a language that poses particular challenges for the L2 learner in the comprehension of running speech. Lexical ambiguities often arise in spoken French as syllable and word boundaries can mismatch due to the phonological processes *enchaînement* (resyllabification) and *liaison*. These processes serve both to avoid hiatus at the boundary between two words (henceforth W_1 and W_2) and to preserve an open syllable structure. *Enchaînement* occurs when W_1 is consonant-final and W_2 is vowel-initial. The coda of W_1 is resyllabified across the word boundary to become the onset of W_2. The phrase *une amie* 'a friend' (feminine) is thus produced as [y.na.mi] where syllable and word boundaries are mismatched, instead of [yn.a.mi] where boundaries would be aligned. *Liaison* on the other hand concerns consonants in final position that are represented graphically, but are not realized phonetically when the word is pronounced in isolation or followed by a consonant-initial W_2. The latent consonant is realized before a vowel-initial W_2 and then resyllabified through *enchaînement* explained above. For example, the determiner *un* (singular, masculine indefinite article) is pronounced [œ̃] in isolation or before a consonant (e.g., *un stylo* [œ̃.sti.lo] 'a pen'). When preceding a vowel onset in W_2, however, as in *un ami* 'a friend' (masculine), the latent /n/ surfaces and is syllabified as the onset of *ami*. Accordingly, the phrase is syllabified [œ̃.na.mi] instead of [œ̃n.a.mi] where word boundaries would be respected.

The effects of resyllabification and the misalignment of syllable and word boundaries on the perception of spoken French by native speakers have generated extensive research (for a review see Nguyen et al. 2007), mainly due to a body of work suggesting that the syllable serves as the basic perceptual unit for speech processing in French (Cutler et al., 1989; Mehler et al., 1981). Given the prominent role of the syllable and syllable boundaries in the processing of spoken French, the prevalence of resyllabification would presumably incur severe processing costs and impede speech segmentation processes. However, Spinelli et al. (2003) found that in the case of liaison, perceptual efficacy and processing in native speakers are not hindered by resyllabification. They probed lexical access processes and revealed significant priming effects for both consonant-initial and vowel-initial words in globally ambiguous sentence pairs such as *c'est le dernier rognon*, 'it's the last kidney', and *c'est le dernier oignon,* 'it's the last onion', both [se.lə.dɛʁ.nje.ʁɔ̃.ɲɔ̃], even though liaison and resyllabification render the two phrases putatively homophonous.

The majority of the classical literature on the acoustic-phonetics of French has maintained that consonants are identical at the acoustic level whether they appear as liaison consonants or initial consonants (e.g., Grammont, 1960). More current research has however shown that consonants that surface in liaison environments (e.g. /n/ in *un air*) are systematically shorter than the same consonant in initial position (e.g. /n/ in *un nerf*). (See for example Gaskell et al., 2002; Spinelli et al., 2003).

Spinelli et al. (2003) hypothesized that native listeners exploit these "subtle but reliable" durational cues in French to mark word boundaries and that this durational variation facilitates access to representations in the mental lexicon (p. 248). They suggested that these differences are robust enough to "bias interpretation in the correct direction" (p. 250) in cases of ambiguity, however this suggestion remains conjectural as this study did not directly demonstrate that duration was guiding participants' responses in the priming tasks.

One way to investigate the use of duration as a segmentation cue in spoken French by native speakers and late learners is to manipulate this single acoustic factor, while holding all other acoustic factors in the signal constant. To this end, the current study employs a forced-choice identification task which utilizes sequences in which the pivotal consonants in ambiguous environments of liaison (i.e. /n/ in [œ̃.nɛʁ], *un air* or *un nerf*) are instrumentally shortened and lengthened while the rest of the utterance remains unaltered. In this way it can be determined whether the durational variation of the pivotal consonants represents a sufficient acoustic cue for segmentation.

3. METHOD

3.1. Participants

The control group consisted of 18 native speakers (NS) of French (15 female, 3 male) ranging in age from 19-54 years (mean: 30.2 years). The experimental group consisted of 18 native speakers of English (11 female, 7 male; mean age: 42.2 yrs, range: 26-71) all of whom met a minimum immersion requirement of five years in France or a French-speaking country at the time of testing (mean residency: 13.8 yrs; range: 5 – 44 yrs). Mean age of arrival in France for the non-native speaker (NNS) group was 28.4 years (range 18-59 years). Mean age of first exposure to French (e.g., either through classroom instruction or time spent in a French-speaking country) was 17.2 years of age (range 6 – 54 years of age).

3.2. Materials

Of the six consonants that surface in liaison environments in French, /g, n, p, ʁ, t, z/, three, /n, t, z/, were chosen to be investigated in this study because they are the most commonly realized in liaison environments in contemporary spoken French (Durand et al., 2005).

Four vowel-initial words were selected, each preceded by words ending in /n, t, z/, thus triggering liaison and ostensibly homophonous sequences. For example, the word *air* 'melody' [ɛʁ] preceded by *un* [œ̃], the singular masculine indefinite article, yields a phonemic sequence consistent with both *un air* 'a melody' and *un nerf* 'a nerve', [œ̃.nɛʁ]. This process resulted in a total of 12 pairs of ambiguous phrases.

Six native speakers of French (5 female and 1 male) aged 25-32 years old (mean 27.3 years) recorded 432 sentences including these globally ambiguous phrases. The durations of the three segments under investigation were then analyzed in both liaison position (e.g., /n/ in *un air*) and initial position (e.g., /n/ in *un nerf*) using Praat sound-editing software (Boersma & Weenink, 2007).

From this production sample, a set of experimental stimuli was created by enhancing the durational differences between liaison consonants (LC) and initial consonants (IC) through instrumental manipulation. In order to determine which value the duration of the manipulated consonants should take, the distribution of durations from the production sample was examined. Following methodology laid out in Shatzman and McQueen (2006), the factor by which the shortened and lengthened segments were manipulated was the standard deviation (SD) in each respective condition.

A three-step durational continuum of stimuli was created which included (1) a shortened consonant representing a LC, (2) a baseline consonant representing durations intermediate to those of LCs and ICs, and (3) and lengthened consonant representing an IC. For each of the three segments, /n, t, z/, three separate measurements were calculated: The shortened (liaison) version of each token represented the mean duration for all instances of that consonant in the *liaison* environment *minus* one SD from that particular mean. The value for the midpoint of the continuum (baseline version) represented simply the mean duration across all instances (LCs and ICs) of each consonant. Finally, the value for the lengthened (word-initial) version of the consonant represented the mean duration for that consonant in *word-initial* position *plus* one SD from that particular mean.

Tokens were subsequently edited using Praat speech-editing software. The manipulation of these phrases resulted in 36 sequences (12 phrases x 3 manipulated versions) that are therefore phonemically identical in their content but differ as to the precise acoustic phonetic realization of the individual consonants under investigation.

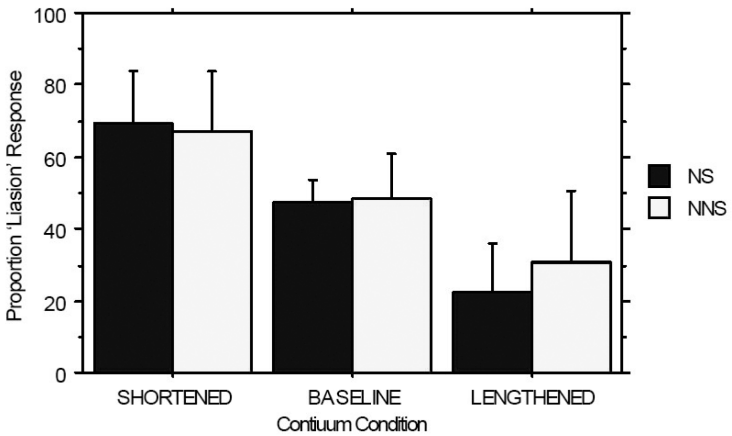
3.3. Procedure

A forced-choice identification task was employed using manipulated tokens taken from the three-step durational continuum of stimuli described above. Each experimental trial had the following structure. Participants heard one of the three manipulated phrases from the durational continuum presented aurally through headphones. Phrases were presented without a carrier frame, thus eliminating any potential priming effects from context. At the offset of the auditory stimulus, two words appeared on the computer screen. The two visual targets consisted of the V-initial and C-initial candidates representing the two possible interpretations of each ambiguous sequence described above. For example, when auditory stimulus is a manipulated version of the sequence [œ̃.nɛʁ] *air* and *nerf* are visual targets. Participants were instructed to indicate which of the two words presented on the screen was present in the phrase they had heard by pressing on the computer keyboard either (1), corresponding to the word on the left of the screen, or (2), corresponding to the word on the right of the screen. There was no delay between the offset of the auditory stimulus and the presentation of the visual targets. Each of the 36 stimuli (i.e. three manipulated versions of each of 12 tokens) was presented randomly six times resulting in a total of 216 trials. Participants completed a training portion consisting of 14 trials before beginning the experimental portion in order to familiarize them with the procedure. Items included in the training portion were not included in the experimental portion. Individual trials were separated by 2000 ms. Visual targets were counter-balanced across participants in order to offset any possible bias toward the left-hand visual target that might occur from reading effects. Half of the participants were presented with the V-initial (liaison) target on the left of the screen and the other half were presented with the C-initial target on the left of the screen. Testing lasted approximately 20 minutes.

3.4. Results

The proportions of V-initial (i.e. 'liaison') responses were calculated for manipulated stimuli in each of the three continuum conditions: the shortened (LC) version, the baseline version, and lengthened (IC) version and are given in Figure 1. A two-way factorial ANOVA compared participant groups and proportions of responses across the three continuum conditions. This analysis revealed a main effect for Continuum Condition: $F(2,102) = 74.30$, $p < .0001$. However, no significant difference between the two Participant Groups was observed: $F(1,102) = 0.73$, n.s. There was no interaction between the two factors: $F(2,102) = 1.256$, n.s.

Figure 1: *Proportion of 'liaison' responses in forced-choice identification across three conditions of durational continuum for NS and NNS participants. Error bars indicate one SD from the mean.*



The above results suggest that the duration of the pivotal consonant alone can indeed modulate the lexical interpretation of ambiguous sequences for both NS and NNS. Shortened consonants elicited significantly more V-initial responses, while lengthened consonants elicited significantly more C-initial responses. In addition, baseline consonants elicited roughly the same proportion of V-initial and C-initial responses, indicating a guessing strategy on the part of participants due to a lack of sufficient acoustic information in the signal.

However, there was a great deal of variation across participants in both groups as evidenced by standard deviations. This brings into question the consistency with which this single acoustic cue is exploited in natural speech and suggest that these durational differences may not represent a consistently robust processing cue in natural speech. However, the fact that both NS and NNS groups responded in the predicted direction demonstrates that segmental duration does have cue value in the processing of liaison environments in spoken French. These results offer strong evidence that durational differences between LCs and ICs are indeed encoded phonologically in both L1 and L2 grammars.

4. L2 LISTENING AND NATIVELIKE PERFORMANCE

Much work in psycho- and applied linguistics seeks to identify and quantify *nativelike* behaviour on linguistic tasks on the part of non-native participants (i.e. non-native behaviour that is indistinguishable from that of native controls). As Birdsong (2009) notes, "referencing learner performance to that of natives provides an easily understood metric of the potential for learner attainment" (p. 408). It is important to note, however, that native performance itself is a measure that must also be empirically established; it is neither uniform nor predictable. Once native performance has been quantified, nativelike behaviour on the part of non-native subjects is usually operationalized as performance that falls either within the actual range of measurements obtained for native controls, or within 1 standard deviation above and below mean native measurements.

However, the quantification of nativelike performance in the current investigation is difficult given the degree of variation among the NS control group. This NS variation resulted in large standard deviations on identification task. For this reason, even more stringent measures of nativelikeness than are usually found in the literature have been employed. Nativelike performance is operationalized here as NNS performance that is at or above native means themselves, as opposed to within 1 standard deviation above or below this mean. Performance on the identification task was operationalized as an average of the proportion of V-initial responses for shortened stimuli and the proportion of C-initial responses for lengthened stimuli. Six NNS participants scored at or above the native mean of 73.50 %, suggesting nativelike sensitivity to allophonic (non-contrastive) durational differences in spoken French.

5. CONCLUSION

The current study has examined the perceptual capacities of highly advanced adult learners of French, which touches upon an area of research that has received little attention to date—namely, the acquisition and exploitation of within-category allophonic variation in L2 processing. Specifically, we have investigated the perceptual capacities of adult learners of L2 French in the exploitation of durational differences that arise between segments produced in word-initial position and segments that surface in liaison. Our results suggest that highly advanced learners of L2 French can develop nativelike sensitivity to allophonic durational variation in environments of liaison in spoken French.

The current results contribute to a growing body of research on the upper limits of L2 phonological processing. Instances of nativelike performance in an L2 have been attested in numerous experimental tasks dealing with L2 domains ranging from morphosyntax to pronunciation (see Birdsong, 2006 for a review). Furthermore, while the acquisition of L2 phonemic *contrasts* has generated an extensive body of work (e.g., Best, 1995; Flege, 1995) much less research focus has been placed on the use of non-contrastive phonetic detail the L2 (for an exception see Darcy et al., 2007). At present, the current study is among the first to demonstrate nativelike attainment with respect to perceptual sensitivity to fine-grained acoustic detail in the L2.

6. REFERENCES

Best, C. (1995). A direct realist view of cross-language speech perception: New Directions in Research and Theory. In Winifred Strange (ed.) *Speech perception and linguistic experience: Theoretical and methodological issues.* (pp. 171–204). Baltimore: York Press.

Birdsong, D. (2006). Age and Second Language Acquisition and Processing: A Selective Overview. *Language Learning*, *56(s1)*, 9-49.

Birdsong, D. (in press). Age and the end state of second language acquisition. In W. Ritchie & T. Bhatia (Eds.), *The new handbook of Second Language Acquisition*. Amsterdam: Elsevier.

Boersma, P. & Weenink, D. (2007). Praat: doing phonetics by computer (Version 4.6.36) [Computer program]. Retrieved May 12, 2007 from http://www.praat.org/

Cutler, A., Mehler, J., Norris, D.G. & Segui, J. (1989). Limits of bilingualism. *Nature*, *340*(6230)*,* 229-230.

Cutler, A., McQueen, J. & Suomi, K. (1997). Finnish vowel harmony and speech segmentation. *Max Planck Institute Annual Report 1997*, 9.

Cutler, A. & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(1)*,* 113-121.

Cutler, A. (2001). Listening to a second language through the ears of a first. *Interpreting*, *5*(1)*,* 1-23.

Darcy, I., Peperkamp, S. & Dupoux, E. (2007). Plasticity in compensation for phonological variation: the case of late second language learners. In J. Cole & J. Hualde (Eds.) *Laboratory Phonology* , 9, Mouton de Gruyter, Berlin.

Dupoux, E., Pallier, C., Sebastián-Gallés, N. & Mehler, J. (1997). A destressing 'deafness' in French? *Journal of Memory and Language*, *36*(3), 406-421.

Durand, J., Laks, B. & Lyche, C. (2005). Un corpus numérisé pour la phonologie du français. In G. Williams (Ed.), *La linguistique de corpus* (pp. 205–217). Rennes: Presses Universitaires de Rennes

Flege, J. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 233-273). Timonium, MD: York Press.

Gaskell, G., Spinelli, E. & Meunier, F. (2002). Perception of Resyllabification in French. *Memory and Cognition. 30*(5), 798-810.

Golato, P. (2002). Word parsing by late-learning French-English bilinguals. *Applied Psycholinguistics*, *23*(3), 417-446.

Grammont, M. (1960). *Traité pratique de prononciation française*. Paris: Delagrave.

Mehler, J., Dommergues, J.Y., Frauenfelder, U. & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior*, *20*(3), 298-305.

Nguyen, N., Wauquier, S., Lancia, L. & Tuller, B. (2007). Detection of liaison consonants in speech processing in French: Experimental data and theoretical implications. In P. Prieto, J. Mascaró & M.-J. Solé (Eds.), *Segmental and prosodic issues in Romance phonology* (pp. 3–23). Amsterdam: John Benjamins.

Sebastian-Galles, N., Dupoux, E., Segui, J. & Mehler, J. (1992) Contrasting syllabic effects in Catalan and Spanish: The role of Stress. *Journal of Memory and Language*, *31*(1)*,* 18-32.

Shatzman, K. & McQueen, J. (2006). Segment duration as a cue to word. boundaries in spoken-word recognition. *Perception and Psychophysics*. *68*(1), 1-16.

Spinelli, E., McQueen, J. & Cutler, A. (2003). Processing Resyllabified Words in French. *Journal of Memory and Language*, *48*(2)*,* 233-254.

Strange, W. (1995). Cross-language studies of speech perception: A historical review. In Strange, W. (Ed.) *Speech perception and linguistic experience: Issues in cross-language speech research*. (pp 3-45). Timonium, MD: York Press.

Weber, A. (2001). Language-specific listening: The case of phonetic sequences. Doctoral Dissertation. University of Nijmegen (MPI Series in Psycholinguisitcs, 16.)

Learning minimally different words in a third language: L2 proficiency as a crucial predictor of accuracy in an L3 word learning task

Ellen Simon[1], Paola Escudero[2], Mirjam Broersma[3]

[1]Ghent University, [2]University of Amsterdam, [3]Max Planck Institute for Psycholinguistics, Nijmegen

Ellen.Simon@UGent.be, escudero@uva.nl, mirjam@mirjambroersma.nl

ABSTRACT

This study examines the effect of proficiency in the L2 (English) and L3 (Dutch) on word learning in the L3. Learners were 92 L1 Spanish speakers with differing proficiencies in L2 and L3, and 20 native speakers of Dutch. The learners were divided into basic and advanced English and Dutch proficiency groups according to their scores on general listening comprehension language tests. Participants were trained and subsequently tested on the mapping between pseudo-words and pictures of non-objects. The analysis revealed that, surprisingly, English proficiency but not Dutch proficiency affected word learning in Dutch. We argue that the expansion of the vowel inventory during L2 learning facilitates L3 word learning.

Keywords: third language acquisition, word learning, minimal pairs, proficiency.

1. INTRODUCTION

The present study sets out to examine native and non-native listeners' learning of minimally different words in a third language (L3), which is defined here as the language acquired after the first (L1) and second language (L2), but which may also be the fourth or fifth language (see Hammarberg 2001:22). Specifically, this study aims to get insight into the effect of L2 and L3 proficiency on L3 word learning.

The perception and identification of L2 sounds which are not contrastive in learners' L1 is known to be highly problematic and has received ample attention in previous research (see the collection of studies in Strange 1995 and Bohn and Munro 2007 for an overview). Well known examples are the problematic perception of the English /r/-/l/ contrast by native speakers of Chinese and Japanese (e.g. Aoyama et al. 2004, Goto 1971) and that of the English /ɛ/-/æ/ contrast by native speakers of Dutch (e.g. Broersma 2005a, Escudero and Simon 2008, Schouten 1975). Inaccurate perception also entails inaccurate recognition of minimally different words. Japanese learners of English have, for instance, been shown to confuse minimal word pairs like *light-write* (Cutler and Otake 2004) and native speakers of Dutch have been reported to experience difficulty with minimal pairs like *flesh-flash* (Broersma 2005b).

Besides the difficulty that learners experience with the perception and recognition of sound contrasts in the L2 which are absent in the L1, it has also been shown that bilinguals cannot separate the lexicons of their two languages (Escudero, to appear). This holds even for highly proficient sequential bilinguals. This has implications for L2 word recognition, since it means that L2 learners listening to the L2 also activate words from their L1 (Marian et al. 2003, Schulpen et al. 2003, Weber and Cutler 2004). In L3 word recognition, the situation is even more complex, since there is cross-linguistic interaction between three instead of two languages. Dijkstra and Van Hell (2003) report on a word recognition experiment with trilingual Dutch-English-French speakers, who were asked to associate L1 Dutch words which did or did not have cognate status with L2 English or L3 French words. The results revealed that the participants were faster in associating L1 words that were cognates with their L2 English and L3 German translations, suggesting that L3 can have a cross-linguistic influence on listeners' L1, even when learners are not aware that their L3 plays a role in the task at hand. However, it is as yet unclear what the role of the two previously-learned languages, i.e. L1 and L2, is in L3 word recognition.

A number of previous studies have investigated the effect of L2 proficiency on L2 acquisition, but provide contradictory evidence: while some studies have shown that experience in the L2 positively correlates with L2 perception and production (Flege 1991, Flege et al. 1997), others did not find a facilitative effect of L2 experience (Cebrian, 2003, 2006; Escudero et al. 2009). The situation is again more complex in L3 acquisition, as both proficiency and experience in the L2 and L3 have to be taken into account.

The present study aims to get insight into the role of learners' L2 and L3 proficiencies in the acquisition of novel L3 words. To that end, we conducted a word learning task in which L1 Spanish speakers with differing proficiencies in L2 English and L3 Dutch performed a Dutch word learning task with novel words.

2. METHOD

2.1. Participants

In total, 92 native speakers of Spanish and a control group of 20 native speakers of Dutch participated in the study. The Spanish-speaking participants came from Spain or a variety of Latin American countries and resided in the Netherlands at the time of testing. All participants performed a general listening comprehension test in Dutch and English prior to testing (DIALANG, www.dialang.org, Alderson and Huhta, 2005). On the basis of the scores for this test, participants were divided into five groups according to their listening proficiency in English and Dutch. Table 1 presents the five groups and the number of participants in each group.

Table 1: Average and sd (between brackets) for each of the five proficiency groups of: N= number of participants, AT= age at testing, AoA=age of arrival, LoR= length of residence. (D = Dutch, E = English).

Group	Language proficiency	N	AT	AoA	LoR
1	D native speakers	20	21.00 (2.6)	--	--
2	Basic D, Basic E	19	33.32 (7.6)	30.84 (7.1)	2.49 (2.7)
3	Basic D, Advanced E	10	30.60 (3.6)	29.30 (3.6)	1.49 (1.1)
4	Advanced D, Basic E	40	39.02 (8)	31.90 (7.3)	7.04 (5.7)
5	Advanced D, Bdvanced E	23	34.04 (8)	28.39 (7)	5.90 (4.1)

2.2. Stimuli

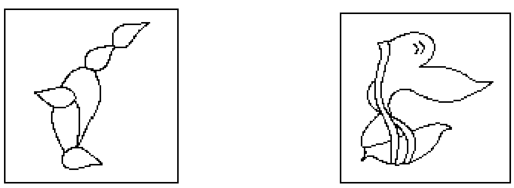
The stimuli consisted of 12 Dutch pseudo-words, six of which were minimally different from each other and six of which were completely different.

The minimally different items were of the form /pVx/, with one of six Dutch vowels, yielding the words /pɪx/, /pix/, /pʏx/, /pyx/, /pɑx/, /pax/.

The other items were disyllabic pseudo-words taken or adapted from Shatzman and McQueen (2006), namely /ˈbe:ptu/, /ˈfo:mpəl/, /ˈjɔmto:/, /ˈkɛstə/, /ˈsurkɛt/, /ˈtœykfɔm/.

Each of the twelve pseudo-words was paired to a line drawing of a pseudo-object (Shatzman and McQueen 2006) (see Figure 1 for examples).

Figure 1. Two examples of pseudo-objects (Shatzman & McQueen 2006)



2.3. Design

The experiment consisted of two parts: a training phase and a testing phase.

During the training phase, participants were first presented with a visual stimulus together with an auditory stimulus (‘This is an X’). Next, they were presented with the same visual target stimulus together with a visual distracter stimulus, and they were asked to click on the target stimulus (‘Click on the X’). The total number of trials in the training phase was 72 (12 items * 6 trials as target).

During the test phase, two visual stimuli were presented and participants were asked to click on the drawing which matched the auditory stimulus (‘Click on the X’). Items were presented either together with a drawing of a minimally different item (‘Minimal pair condition’, e.g. a picture of a /pɪx/ presented with one of a /pix/, with the instruction ‘Click on the /pɪx/), or with a drawing of a completely different item (‘Non-minimal pair condition’, e.g. a picture of a /pɪx/ presented together with one of a /ˈbe:ptu/, with the instruction ‘Click on the /ˈbe:ptu/).

2.4. Procedure

Participants were tested one at a time in a quiet room. For the training phase, they were told that they were going to be taught new Dutch words. For the testing phase, they were informed that they were going to be tested on their recognition of the newly learnt Dutch words. Each phase started with a number of practice trials, after which participants could ask questions. There was a short break between training and test phase, in which instructions were provided.

3. RESULTS

The average percentage correct for each participant was higher than 70% (range: 70%-100%) and all results were therefore included in the statistical analyses.

Repeated measures ANOVAs were done on the Spanish listeners' results with Pair Type (minimal pair and non-minimal pair) as within-subjects variable and Dutch and English proficiency (advanced and basic) as between-subjects variable. In the first ANOVA, the dependent variable was the percentage of correct responses, in the second one it was the RT. The analyses revealed that minimal pairs had a lower percentage correct than non-minimal pairs (percentage correct: $F(1,88)=177.53$, $p<0.001$; RT: $F(1,88)=172.23$, $p<0.001$). Surprisingly, learners with advanced Dutch proficiency did not have a higher percentage correct than those with basic Dutch proficiency ($F(1,88)=0.011$, $p=.918$). However, learners with advanced English proficiency had a higher percentage correct than those with basic English proficiency ($F(1, 88)=10.297$, $p<0.01$).

Regarding percentage correct, there was a significant three-way interaction between Pair Type, Dutch Proficiency, and English Proficiency ($F(1,88)=5.40$, $p<0.05$). In order to investigate this three-way interaction, the advanced and basic English proficiency groups were compared with two t-tests, for minimal pairs and non-minimal pairs separately. Learners with advanced Dutch proficiency did not have a significantly higher percentage correct than learners with basic Dutch proficiency, neither for the minimal pairs ($t(91)=1.35$, $p=0.249$) nor for the non-minimal pairs ($t(91)<1$). Crucially, however, learners with advanced English proficiency did have a higher percentage correct than learners with basic English proficiency, both for the minimal pairs ($t(91)=6.58$, $p<0.05$) and for the non-minimal pairs ($t(91)=4.73$, $p<0.05$).

In order to further explore the three-way interaction for percentage correct, each of the four groups of Spanish learners (with advanced-advanced, advanced-basic, basic-advanced, and basic-basic proficiency in Dutch and English, respectively) and the Dutch native listeners were compared. First, a repeated measures ANOVA on percentage correct with Pair Type as within-subjects factor and Group (see Table 1 for the 5 groups) as between-subjects factor showed a significant interaction between Pair Type and Group ($F(4,107)=7.41$, $p<0.001$). Separate one-way ANOVAs for minimal and non-minimal pairs showed a significant effect of Group for the minimal pairs ($F(4,111)=9.74$, $p<0.001$) but not for the non-minimal pairs

(F(4,111)=2.12, p=0.083). Therefore, further investigations were done with the minimal pairs only. Table 2 presents the results of Bonferroni-corrected comparisons between the five groups.

Table 2: Bonferroni-corrected post-hoc tests comparing the five listener groups, for mean difference in percentage correct for minimal pairs. (Positive and negative values refer to the group on the top row, having a higher or lower percentage correct, respectively, than the group in the left column). (D = Dutch, E = English).

	D native speakers	Advanced D, Advanced E	Advanced D, Basic E	Basic D, Advanced E	Basic D, Basic E
D native speakers	-				
Advanced D, Advanced E	+7, $p<0.05$	-			
Advanced D, Basic E	+8.9, $p<0.01$	+1.8, $p=1.0$	-		
Basic D, Advanced E	+3.8, $p=1$	-3.2, $p=1.0$	-5.1, $p=0.519$	-	
Basic D, Basic E	+13.7, $p<0.01$	+6.7, $p<0.05$	+4.9, $p=0.185$	+9.9, $p<0.01$	-

The results in Table 2 show that the Dutch native listeners’ percentage correct for the minimal pairs was significantly higher than that of all other groups, except for the learners with basic Dutch and advanced English proficiency, whose accuracy did not significantly differ from the native listeners' percentage correct (possibly due to low statistical power, as that learner group contained only 10 subjects; see Table 1). As for the learner groups, Bonferroni-corrected comparisons suggest again that English proficiency was a more important predictor of accuracy on the task than Dutch proficiency: groups that differed only in Dutch proficiency (i.e., *advanced Dutch* and advanced English versus *basic Dutch* and advanced English, and *advanced Dutch* and basic English versus *basic Dutch* and basic English) did not exhibit a significant difference in accuracy. By contrast, some groups that differed only in English proficiency had significantly different accuracies: learners with basic Dutch and *advanced English* proficiency had a higher percentage correct than learners with basic Dutch and *basic English* proficiency. These results confirm that the source of the difference in percentage correct lies in the learners’ level of English proficiency rather than in their Dutch proficiency.

4. DISCUSSION AND CONCLUSIONS

The most important finding in the present study is that proficiency in the learners’ L2, English, was a better predictor of their accuracy in learning minimally different Dutch words than their proficiency in their L3, Dutch. Here we address a number of possible explanations for this surprising result.

One potential explanation why learners with advanced English proficiency performed better at the Dutch word learning task than learners with advanced Dutch proficiency could lie in learners’ different levels of proficiency in English (L2) and Dutch (L3). Specifically, if learners’ average level of English proficiency were higher than their average level of Dutch proficiency, this would explain why proficiency in English had more influence than proficiency in Dutch. We tested this hypothesis with an independent samples t-test comparing the Dialang scores (from 1 to 6) of the learners who had high English proficiency with those who had high Dutch proficiency. Crucially, we found a significant difference in the opposite direction, i.e. proficiency in

Dutch was higher than proficiency in English (Dutch advanced mean (N=63): 5.62, English advanced mean (N=33): 4.67, $t(94)=6.953$, $p<0.001$). In other words, the advanced Dutch and English learners were not more proficient in English than in Dutch.

A second explanation related to the learners' level of proficiency in the two languages could be a difference in the ranges of proficiency scores: the difference between English advanced and basic might be larger than that between Dutch advanced and basic. However, an independent samples t-test comparing the Dialang scores that were grouped as basic Dutch proficiency with the scores that were grouped as basic English proficiency again reveals a result in the opposite direction: the basic Dutch scores were on average lower than the basic English scores (Dutch basic mean: (N=29): 1.448, English basic mean (N=59): 2.017, $t(66.84$[1]$)=-3.159$, $p<0.01$). Further, as shown above, the advanced Dutch scores were on average higher than the advanced English scores. Thus, learners with advanced Dutch had higher Dialang scores than learners with advanced English, while learners with basic Dutch had lower scores than learners with basic English. Consequently, the difference between advanced and basic learners was larger in Dutch than in English, which contradicts the hypothesis that a higher English proficiency could account for the fact this language was the best predictor of L3 word learning accuracy.

Thirdly, general second language acquisition constraints such as the age factor, the order of acquisition and foreign language learning abilities (see, among others, Mayo and Lecumberri 2003, Singleton and Ryan, 2004) are likely to have contributed to the present results. Specifically, the fact that the learners in this study had acquired English earlier in life than Dutch could potentially account for the greater influence of English proficiency compared to Dutch proficiency. Whereas the Dialang scores showed that the learners were not more proficient in their earliest acquired foreign language, English, than in their L3, Dutch (general listening proficiency in Dutch was higher than in English), it could still be the case that an earlier acquired language affects L3 learning more than a later acquired language.

Finally, the most likely explanation for the greater influence of English proficiency compared to Dutch proficiency is that English is comparable to Dutch in terms of the size of the vowel inventory. Specifically, the English vowel inventory is considerably larger than the Spanish one and hence more similar in size to the Dutch inventory. The expansion of the vowel inventory during the acquisition of English may have benefited word learning in Dutch. Even though Dutch and English vowels are not the same, this similarity in vowel inventory size between English and Dutch may be the key to the facilitative effect of English on the learning of Dutch words: learners who have acquired a second language with a large vowel inventory would have an advantage when learning a third language with a similarly large inventory. Similarly, Mattock et al. (2010) showed that bilingual French-English infants learned minimally different words faster than monolingual children when the words differed in phoneme contrasts that were contained in both of the bilingual infants' two languages. Possibly, the support of an L2 with a similar sound system is required for language proficiency to have an effect on vowel perception and word recognition, which might explain why in our study Dutch proficiency on its own could not predict learning accuracy. The hypothesis that vowel inventory expansion affects L3 learning also implies that L2 learners of a language with a small vowel inventory would not have the same advantage when learning an L3 with a large inventory. For instance, Spanish learners of Dutch would not have an advantage when learning Dutch words had they learned Basque or any other language with a small vowel inventory as a second language, instead of English. This was confirmed by Gonzalez Ardeo (2001) and Gallardo del Puerto (2007), who showed that Spanish-Basque bilinguals did not show an advantage over monolinguals when learning English vowels. If knowledge of a language with a small vowel inventory does not facilitate the learning process, the general conclusion can be that L2 language learning by itself does not necessarily or automatically facilitate L3 learning.

5. REFERENCES

Alderson, J.C., Huhta, A. 2005. The development of a suite of computer-based diagnostic tests based on the Common European Framework. *Language Testing* 22. 301-320.

[1] Corrected for unequal variance between groups: Levene's test for equality of variance was significant ($p<0.045$).

Aoyama K., Flege J.E., Cuion S.G., Akahane-Yamade R., Yamada T. 2004. Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics* 32. 233–250.

Bohn, O.-S., Munro, M. J. (eds.) 2007. *Language experience in second language speech learning: In honor of James Emil Flege*. Amsterdam: John Benjamins.

Broersma, M. 2002. Comprehension of non-native speech: Inaccurate phoneme processing and activation of lexical competitors [oral]. Proceedings of the 7th International Conference on Spoken Language Processing, Center for Spoken Language Research, University of Colorado, Boulder (CD-ROM). 261-264

Broersma, M. 2005a. Perception of familiar contrasts in unfamiliar positions. *Journal of the Acoustical Society of America* 117. 3890-3901.

Broersma, M. 2005b. Phonetic and lexical processing in a second language. Nijmegen, The Netherlands: PhD dissertation, Radboud University Nijmegen.

Cebrian, J. 2003. Input and experience in the perception of an L2 temporal and spectral contrast. In D. Recasens, M.J. Solé, J. Romero (eds.), Proceedings of the 15th International Congress of the Phonetics Sciences. Barcelona, Spain: Universitat Autònoma de Barcelona/Causal Productions. 2297-2300.

Cebrian, J. 2006. Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics* 34. 372-387.

Cutler, A., Otake, T. 2004. Pseudo-homophony in non-native listening. Paper presented to the 75th meeting of the Acoustical Society of America, New York.

Dijkstra, T., van Hell, J. G. 2003. Testing the Language Mode hypothesis using trilinguals. *International Journal of Bilingual Education and Bilingualism* 6(1). 2-16.

Escudero, P. to appear. Speech processing in bilingual and multilingual listeners: The interrelation between sound perception and word recognition. In Cohn, A.C., Fougeron, C., Huffman, M. (eds.), *Oxford Handbook of Laboratory Phonology*. Oxford: Oxford University Press.

Escudero, P., Simon, E. 2008. Perception of two English front vowels by L1 Dutch speakers with different regional accent. *Journal of the Acoustical Society of America* 123(5). 3076.

Flege, J. E. 1991. Perception and production: the relevance of phonetic input to L2 phonological learning. In Huebner, T. , Ferguson, C. A. (eds.), *Cross Currents in Second Language Acquisition and Linguistic Theory*. 249-290.

Flege, J. E., Bohn, O.-S., Jang, S. 1997. Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics* 25. 437-470.

Goto, H. 1971. Auditory perception by normal Japanese adults of the sounds "l" and "r". *Neuropsychologia* 9. 317-323.

Hammarberg, B. 2001. Roles of L1 and L2 in L3 production and acquisition. In Cenoz, J., Hufeisen, B. and Jessner, U. (eds.), *Cross-linguistic influence in third language acquisition. Psycholinguistic perspectives.* Clevedon: Multilingual Matters. 21-41.

Marian, V., Spivey, M., Hirsch, J. 2003. Shared and separate systems in bilingual language processing: Converging evidence from eyetracking and brain imaging. *Brain and Language* 86. 70-82.

Mattock, K., Polka, L., Rvachew, S., Krehm, M. 2010. The first steps in word learning are easier when the shoes fit: Comparing monolingual and bilingual infants. *Developmental Science* 13. 229-243.

Mayo, M., Lecumberri, M. (eds.). (2003). *Age and the acquisition of English as a foreign language*. Clevedon: Multilingual Matters.

Shatzman, K.B., McQueen, J. M. 2006. Prosodic knowledge affects the recognition of newly acquired words. *Psychological Science* 17. 372-377.

Schouten, M. E. H. (1975). *Native-Language Interference in the Perception of Second-Language Vowels: An Investigation of Certain Aspects of the Acquisition of a Second Language.* Unpublished doctoral dissertation, Utrecht University, The Netherlands.

Schulpen, B., Dijkstra, T., Schriefers, H. J., Hasper, M. 2003. Recognition of interlingual homophones in bilingual auditory word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 29. 1155-1178.

Singleton, D., Ryan, L. 2004. *The Age Factor.* Clevedon: Multilingual Matters.

Strange, W. (ed.) 1995. *Speech Perception and Linguistic Experience*. Baltimore: York Press.

Weber, A., Cutler, A. 2004. Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language* 50. 1-25.

Acquiring the high vowel contrast in Quebec French: How assibilation helps

Laura Catharine Smith and Wendy Baker

Brigham Young University
laurasmith@byu.edu and wendy_baker@byu.edu

ABSTRACT

In Quebec French (QF), /t/ and /d/ are assibilated to [t^s] and [d^z] before /i/ and /y/, but not before /u/. Since the /y/-/u/ contrast is known to be difficult for English speakers learning French as a second language (L2), we examine whether L2 learners of French who have acquired the assibilation rule have any advantage in producing and perceiving the French /i/-/y/-/u/ contrast over L2 learners who produce less or no assibilation in their L2 French. Results demonstrate that L2 learners who are strong assibilators are better at producing vowels similarly to native QF speakers than weak assibilators, but in perception, L2 learners who produce strong assibilation had no statistically significant advantage over L2 learners who are weak assibilators in being able to discriminate or identify the French high vowels. We conclude that production of assibilation in L2 Quebec French helps learners in production, though not perception, further providing insight into the relationship between L2 perception and production.

Keywords: Quebec French, Second Language Acquisition, Assibilation, Perception, Production

1. INTRODUCTION

The question of how speech perception and production are related is one of the major questions in second language acquisition (L2) phonological research (Flege et al. 1999). Many researchers assume that accurate perception leads to accurate production (i.e., Flege 1995), although other researchers have been unable to find a clear link (Zampini and Green 2001) or have argued that production is more independent of perception than previously thought (Smith 2001). In this paper, we examine the relationship between L2 perception and production by investigating whether an additional cue in the target dialect of a given L2, namely assibilation of /t/ and /d/ to [t^s] and [d^z] before the high front vowels /i/ and /y/ in Quebec French (QF) (cf. Walker 1984), can help learners of French acquire both the perception and production of the contrast between the high vowels /i/, /y/ and /u/. The potential impact of assibilation on acquiring this contrast is particularly noteworthy since the contrast between /y/ and /u/ is well known to be difficult for native speakers of English learning French. While this external cue is characteristic of Quebec French in additionally marking the contrast between the front and back high vowels, it is not explicitly taught to L2 learners. This then leads us to ask whether native English speakers learning French as an L2 in Quebec come to acquire assibilation themselves. If they do, does this assibilation lead to a more native-like pronunciation of QF, and/or a superior perceptual ability in contrasting between the French high vowels than those QF learners who have not acquired alveolar stop assibilation?

The current study seeks to address these issues, namely whether having acquired the QF assibilation rule facilitates native-like production and perception (discrimination and identification). To this end, we pose the following research questions at the heart of this study:

1. Do QF learners who have acquired the QF assibilation rule produce the vowels /i/, /y/, and /u/ in a more native-like way than QF learners who assibilate less or not at all?
2. Do QF learners who most frequently assibilate /t/ and /d/ before /i/ and /u/ better discriminate between /i/, /y/, and /u/ than QF learners who produce assibilation less or not at all?
3. Are QF learners who most frequently assibilate also better able to identify the vowels /i/, /y/, and /u/ than non- or less frequent assibilating QF learners?

2. EXPERIMENT 1: PRODUCTION

In the first experiment, we examine whether QF learners who produce higher degrees of assibilation concomitantly produce more native-like productions of /i/, /y/, and /u/. In other words, does the acquisition of the QF assibilation rule in production translate into an advantage in native-like pronunciation?

2.1 Methodology

Twenty participants took part in this study. Participants were all native speakers of North American English who had spent 22 months living in Quebec, Canada during which they were exposed to Quebec French. Participants were assigned to one of two groups based on the percent of times they produced assibilation on /t/ and /d/ before /i/ and /y/ in 22 tokens (an additional 10 /u/-tokens were used to control for incorrect assibilation of /t/ and /d/ before /u/; any incorrectly assibilated tokens were deducted from the assibilation score). Participants in the strong assibilation (SA) group produced assibilation 89-100% of the time, while Weak Assibilators (WA) produced assibilated tokens 0-72% of the time.

Table 1: Demographics of Participants

Groups of QF learners	CA	AOA	LOR	YRS
Strong assibilators (SA) (n=10)	22.25	18.33	22 months	6.5
Weak/non-assibilators (WA) (n=10)	21.61	18.71	22 months	4.0

CA=Current age at testing; AOA=age of arrival in target dialect; LOR= Length of residence in target dialect; YRS=years learning language

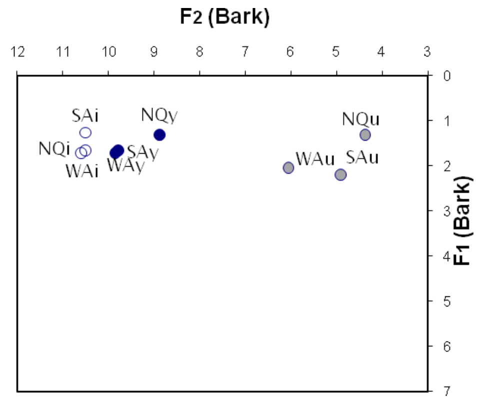
All participants produced six words for each of the three vowels in the carrier phrase "Je dis le mot___" ('I say the word___'). Five native speakers of Quebec French (average age: 26) also produced these same vowels for comparison purposes. Half of the words contained the vowels in a phonetic context where assibilation occurs (after alveolar stop consonants) and half were produced in other contexts. The words were extracted from the carrier phrase and each vowel's fundamental frequency and first three formants were measured. These measurements were normalised to the Bark scale using the following formula: $B = 26.81 / (1+(1960 /F)) -0.53$, where F = the formant (or fundamental frequency) of each vowel measurement (Syrdal and Gopal 1986). We compared the vowel productions of the two groups of QF learners (strong assibilators and weak assibilators) to the productions of the native QF speakers in terms of vowel frontedness (F1), height (F2), and lip rounding (F3).

2.2 Results

We examined whether the strong assibilators were more likely than the weak assibilators to produce the three French vowels in native-like manner, i.e., like a QF native speaker. We hypothesised that the strong assibilator group, because they have already picked up on the extra acoustic cue distinguishing between the French vowels /i/-/y/ and /u/, would produce the vowels more accurately than the low assibilator group.

We tested this hypothesis by comparing the production accuracy of the two learner groups to the native speakers. We ran a series of two-way (group x vowel) ANOVAs on vowel height, frontedness, and rounding. The results of these analyses revealed that the two learner groups did not differ from each other or from the native Quebec French speakers in their production of vowel height ($F(2,24)=1.67$, $p=.197$) nor for lip rounding ($F(2,24)=979$, $p=.381$), but they did differ in terms of vowel frontedness ($F(2,24)=7.63$, $p=.001$). Tukey post-hoc tests revealed that the strong assibilator group produced the vowels in terms of frontedness similarly to the native Quebec French speakers, while the weak assibilator differed from the other two groups (SA group and native QF speakers). This can be seen in Figure 1 below.

Figure 1: Production of French /i/, /y/, and /u/ by native Quebec French speakers (NQ), the strong assibilator learner group (SA) and weak assibilator (WA) learner group



2.3 Discussion

The results of the production task revealed that the strong assibilator group was more accurate in its production of the French vowels than the weak assibilator group in terms of frontedness of the three vowels. Indeed, the weak assibilator group not only produced French /i/-/y/ more closely together, but they also fronted their French /u/ more forward than the strong assibilator and native Quebec French speaker groups. It may be that the lack of assibilation of /t/ and /d/ before /u/ had the effect of drawing attention to the more back production of /u/ in contrast with /i/ and /y/, a noteworthy observation helping the learners produce a more native-like production of /u/.

3. EXPERIMENT 2: DISCRIMINATION

In this second experiment, we examine whether QF learners who have acquired the assibilation rule of QF, namely the strong assibilator group, are better able to accurately discriminate between the French vowels /i/, /y/ and /u/ than the weak assibilator group.

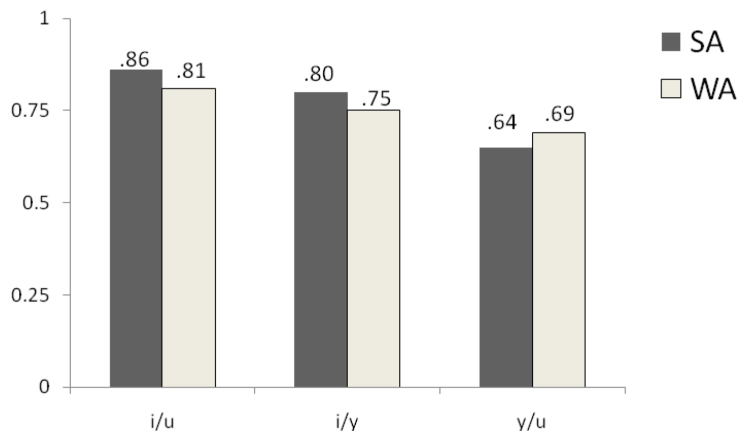
3.1. Methodology

The stimuli for this experiment, produced by native speakers of Quebec French (average age: 22), were tokens of the French words *dit* (/di/), *doux* (/du/), and *du* (/dy/). We chose these words because the native QF speakers produced these words with assibilation before the high front vowels, providing the opportunity to determine whether the two groups differ in their ability to use this cue to discriminate between the vowels, especially French /y/ and /u/. Participants heard pairs of these tokens presented randomly by E-Prime and were asked to determine whether the two tokens they heard were either the same vowel (/i/-/i/, /y/-/y/, /u/-/u/) or different vowels (/i/-/u/, /i/-/y/, /y/-/u/). For each of the vowel-pairs, listeners heard 4 same tokens (2 of one vowel and 2 of another vowel) and 4 different tokens for a total of 36 tokens (3 vowel pairs x 4 same and 8 different vowel pairs) altogether.

3.2. Results

The number of correct discriminations of both same and different vowel pairs was calculated for each vowel pair (i/u, i/y, y/u) for each participant. The correct discriminations were converted into A' scores, which takes into account response bias (see Snodgrass et al. 1985 for a discussion of this measure). A score of 1.0 is a perfect discrimination score, whereas a score of .5 is chance performance. (See Figure 2 below.) To determine whether the strong assibilators were more accurate in their perception the French vowels than the weak assibilators, we submitted the A' scores for each vowel pair to a two-way (group x vowel pair) ANOVA. The results of this experiment revealed a significant effect of vowel ($F(2,19)= 3.66$, $p=.03$),but no effect of group ($F(1,19)=.113$, $p=.739$), nor a group x vowel interaction ($F(2,1)=.316$, $p=.731$). In other words, the two groups did not differ from each other in their discrimination of any of the three vowel pairs, /i/-/y/, /i/-/u/, /y/-/u/, although both groups were more accurate at discriminating /i/-/u/ than /i/-/y/ and /y/-/u/ vowel pairs. In addition, they were more accurate at discriminating /i/-/y/ than /y/-/u/.

Figure 2: Discrimination accuracy of French vowel pairs /i-u/, /i-y/, and /y-u/ by the strong assibilator learner group (SA) and weak assibilator (WA) learner group



3.3. Discussion

The results of this experiment suggest that the two learner groups, those that produce a strong degree of assibilation when producing high front French vowels (modeled after native Quebec French speakers) and those who do not do so, did not differ in their discrimination of the three French vowels /i/, /u/, and /y/. Such findings suggest that producing the extra acoustic cue to distinguishing French /y/ and /u/ did not change the perception abilities of the learners.

4. EXPERIMENT 3: IDENTIFICATION

In this final experiment we investigate whether QF learners who produce more assibilation have any advantage at identifying the vowels /i/, /y/, and /u/. If assibilation does play a role in identification, we would expect that QF learners who produce more assibilation, i.e., the strong assibilator group, would in turn be able to more accurately identify the vowels than the weak assibilator group. We examinated identification as well as discrimination to see how accurately both learner groups were able to identify these vowels in phonetic contexts where assibilation does not occur (in all other contexts except after alveolar stop consonants).

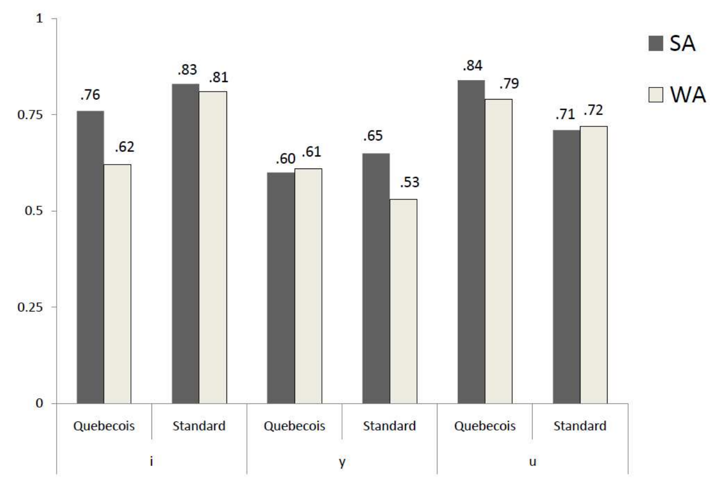
4.1. Methodology

The same participants that participated in the previous experiments also participated in the identification task. Listeners heard tokens of French vowels /i/, /y/, and /u/ in CVC, CV, and V contexts. The tokens were produced both by native Quebec and native European French speakers (average age: 24). Tokens produced by speakers of both QF and standard French were used to determine if the two learner groups differed in their ability to generalize perception abilities to a standard dialect.

4.2. Results

The number of correct identifications for each vowel for each participant was calculated and submitted to a multifactorial (group x dialect x vowel) ANOVA. The results of this analysis revealed a significant effect of vowel ($F(2,19)=3.02$, $p=.05$), but no effect of group ($F(1,19)=.033$, $p=.968$), nor dialect ($F(1,19)=.552$, p=.459) nor any other significant interactions. Tukey post-hoc tests revealed that both groups were less accurate in their identification of /y/ than the other two vowels for both dialects. (See Figure 3.)

Figure 3: Identification of French /i/, /y/, and /u/ produced by native Quebec French speakers (Quebecois) and Standard French speakers by the strong assibilator learner group (SA) and weak assibilator (WA) learner group



4.3. Discussion

The results of the identification task revealed that, although the strong assibilator (SA) group typically correctly identified the French vowels more accurately than the weak assibilator (WA) group, these differences did not reach statistical significance. Moreover, while, again, the strong assibilator group typically identified the vowels spoken by the Standard French speakers more accurately than the weak assibilator group, these differences also did not reach statistical significance.

5. GENERAL DISCUSSION

Based on the results from the three experiments, we can now answer our research questions. First, we found that L2 learners of Quebec French who had acquired the QF assibilation rule, namely the group of strong assibilators, produced the vowels /i/, /y/ and /u/ in a more native-like manner, i.e., more like a native QF speaker, in terms of frontedness than members of the weak assibilator group. Although the two QF learner groups did not differ in terms of vowel height (F1) or lip rounding (F3), the difference between the two groups for frontedness corresponds well to the presence or lack of assibilation at the heart of the study. Simply, the occurrence of assibilation is directly related to the frontedness or backness of the vowel: the high front vowels, /i/ and /y/, trigger assibilation, but the back vowel /u/ does not. That the strong assibilator group correctly applied the QF assibilation rule 89% of the time or more suggests a higher awareness of the difference between these vowels in terms of frontedness, particularly for the positioning of /u/, in comparison to the weak assibilator group.

The difference found between the SA and WA groups in production, however, was not reflected for either discrimination or identification of the vowels. Consequently, we must answer both the second and third research questions in the negative: L2 learners who were strong assibilators (SA) did not have any advantage in discriminating or identifying /i/, /y/, or /u/ over their weak assibilator counterparts (WA). Although the SA group did tend to perform better than the WA group, differences were not statistically significant.

These results suggest a difference in the role of QF assibilation in L2 perception and production of /i/, /y/, and /u/ by learners of Quebec French. While the production of assibilation by L2 learners did impact their production, it did not impact their perception. This mismatch may be explained as follows. All learners in this study, regardless of the extent to which they assibilated before high front vowels, would have been exposed to input where assibilated [t^s] and [d^z] occurred before the high front vowels. This similar input may then have resulted in a similar ability to perceive this contrast. Indeed, in Baker and Smith (Under Review), QF learners who had been exposed to this more salient assibilation cue in their target dialect were better able to produce and perceive the French vowels /i/, /y/, and /u/ than L2 learners of European French. What the present study suggests is that exposure to these more salient acoustic cues has a similar effect on perception for all L2 learners exposed to assibilation, whether or not they come to acquire or use the rule themselves in their productions. This may be due to the common exposure to this external cue.

Differences between the two groups in production, however, could be argued to stem from a difference in how the assibilation rule is internalised for L2 learners. In other words, if L2 learners come to *produce* assibilation, then assibilation impacts other related aspects of their *production*, namely the production of the frontedness of the high vowels. By noticing that assibilation is grouped with front vowels, but not back vowels, these learners may be more able to accurately produce both assibilation and the respective high vowels for which assibilation serves as an additional acoustic cue. Only if the cue is internalised for a learner's own production, can it impact that learner's production. In that way, assibilation plays a separate role in perception and production for L2 learners of Quebec French.

If L2 learners do internalise this cue for production, then, as these results suggest, it gives them yet one additional advantage to produce the high vowels more like native French speakers. These findings when placed within the larger context thus provide insights into the way by which salient acoustic cues can play a role in L2 perception and production, while highlighting the asymmetry in the relationship between production and perception.

6. REFERENCES

Baker, W., Smith, L. Under review. The impact of L2 dialect on learning French vowels: Native English speakers learning Quebecois and European French.

Coughlin, M. Kewley-Port, D., Humes, L. E. 1998. The relation between identification and discrimination of vowels in young and elderly listeners. *Journal of the Acoustical Society of America* 104(6). 3597-3607.

Escudero, P., Boersma, P. 2004. Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition* 26. 551-585.

Flege, J.E. 1995. Second language speech learning: Theory, findings, and problems. In Strange, W. (ed). *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press, 229–273.

Flege, J. E., MacKay, I. R. A., Meador, D. 1999. Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America* 106. 2978–2987

Grosse, C., Hameyer, K. 1979. Dialect and Reading Interferences in Second Language Perception and Production. *Die Unterrichtspraxis*, 12. 52-60.

O'Brien, M.G., Smith, L.C. To Appear. Impact of first language dialect on the production of German vowels. *International Review of Applied Linguistics*.

Smith, L.C. 2001. L2 acquisition of English liquids: Evidence for production independent from perception. In Bonch-Bruevich, X, Crawford,W.J., Hellermann, J., Higgins, C. Nguyen, H. (Eds.), *The Past, Present, and Future of Second Language Research: Selected Proceedings of the 2000 Second Language Research Forum.* Somerville, MA: Cascadilla Press, 3-22.

Snodgrass, J.G., Levy-Berger, G., Haydon, M. 1985. *Human experimental psychology.* New York, NY: Oxford University Press.

Syrdal, A.K., Gopal, H. S. 1986. A perceptual model of vowel recognition based on auditory representation of American English vowels. *Journal of the Acoustical Society of America* 79. 1086-1100.

Teasdale, A. 1997. On the differential substitution of English [θ]: A phonetic approach. *Calgary working papers in linguistics.* 19, 71-85.

Walker, D. C. 1984. *The Pronunciation of Canadian French.* Ottawa: University of Ottawa Press.

Zampini, M., Green, D. 2001. The voicing contrast in English and Spanish: The relationship between perception and production. In Nicol, J. (ed.), *One Mind, Two Languages*, Oxford: Blackwell, 23–48..

L2 Dialect on Learning German Vowels: The Case of Northern German and Austrian Dialects

Laura Catharine Smith and Wendy Baker

Brigham Young University
laurasmith@byu.edu; wendy_baker@byu.edu

ABSTRACT

This study examines how the dialect of a second language (L2) affects how accurately L2 is perceived and produced. Specifically, we examined differences between the production and perception of German vowels /i/, /y/, and /u/ by learners of either Austrian German (AG) or Northern German (NG). Vowels across these dialects differ due to salience of cues to the /i/-/y/-/u/ contrast: (more) derounding of /y/ for AG versus NG leading to loss of an acoustic cue marking /i/-/y/, but a potentially enhanced acoustic cue for the /y/-/u/ contrast. As a result of these differences, both dialects have opposing cues by which to contrast /i/, /u/ and /y/. Results indicate that AG learners are at times more accurate than NG learners in their perception and production of these German vowels, suggesting that L2 dialect exposure impacts L2 phonological learning.

Keywords: German, dialects, vowels, second language

1. INTRODUCTION

Recent studies in second language (L2) research have shown that the L2 dialect to which one is exposed affects a learner's ability to accurately produce and perceive L2 sounds. For instance, L2 learners of Standard British English differ from learners of Scottish English in their perception and production of English vowels (Escudero and Boersma 2004). The effect of the L2 dialect on L2 phonological learning seems especially salient if one dialect contains acoustic cues that help in distinguishing between difficult L2 sound contrasts. Baker and Smith (under review), for example, found that learners of Quebecois French (QF) were more accurate than learners of European French (EF) at perceiving and producing French /i/, /y/, and /u/. QF learners' increased accuracy may have been caused by two factors: QF has an extra acoustic cue (assibilation of alveolar consonants before high front vowels) that distinguishes between French /y/-/u/. Moreover, QF /y/ and /u/ vowels are also more separated acoustically in the vowel space than are EF vowels.

The current study expands this research by examining whether a similar effect of L2 dialect occurs in other languages and dialects. In particular, we compare differences in how native English learners of German perceive and produce German /i/, /y/, and /u/ after exposure to one of two German dialects, Austrian (AG) and Northern German (NG). While speakers of Northern German in major centres such as Hanover, Frankfurt and Hamburg tend towards a more Standard German pronunciation, speakers of Austrian and Southern German (grouped together here as AG) tend to use more dialect in their production and show greater variation in production of /i/, /y/, and /u/. First, many AG speakers tend to deround /y/ such that it is (nearly) merged with /i/ even when speaking a more "standard" German variety (cf. Russ 1990; Moosmüller 1987). A potential consequence for L2 learners of AG, is that they hear few if any clear exemplars by which to establish a new category for /y/ in comparison with Standard German /i/, although this loss of rounding may actually lead to an additional cue by which learners can contrast /y/ with the back rounded vowel /u/, a contrast well known to be difficult for English speakers learning German. The complex relationship between these vowels is further complicated by the fact that /i/ (and its derounded /y/ counterpart) and /u/ are often diphthongised by AG dialect speakers to [iɐ] and [uɐ] respectively (Russ 1990). These differences in Austrian and Northern German dialects thus allow us to determine whether native English speakers exposed to these dialects exhibit differences in the L2 perception and production accuracy of these German vowels.

Thus, the current study examines the following research questions:

1. Are learners more likely to accurately discriminate and identify vowels produced in the dialect to which they were exposed as opposed to the other dialect?

2. Do AG learners identify and discriminate German vowels more or less accurately than NG learners? In other words, does loss of the rounding cue between /i/-/y/ in AG negatively impact AG learners of this contrast in comparison with NG learners exposed to fully contrasted /i/-/y/? Or does derounding serve as an extra acoustic cue for AG learners learning to distinguish /y/-/u/?
3. Are AG learners also more or less likely than NG learners to accurately produce these German vowels like the target dialect to which they were exposed?

The first two research questions were addressed in the first two experiments of this study, the first of which was a discrimination and the second of which was an identification task. The final research question was addressed in the third experiment where participants were asked to produce the three vowels in 18 words.

2. EXPERIMENT 1

In this first experiment, we examined whether the dialect to which learners were exposed influences how accurately learners are able to discriminate the German vowels /i/, /y/, and /u/.

2.1. Methodology

All participants were native English speakers who were exposed to German either in Austria or Northern Germany and who had spent at least 16 months immersed in the target dialect. For demographic information, see Table 1. The “Austrian group” spent 16+ months in both Southern Germany and Austria (Central and Southern Bavarian dialect area). However, since many of them spent more time in Austria than in Southern Germany, we used the “Austrian dialect” as the second dialect in this study. The “Northern group” had spent all of their time in Northern Germany, particularly in the Hamburg and/or Frankfurt areas. All participants had had similar language training focusing on Standard German prior to their immersion in the target dialect.

Table 1: Demographic Information of Participants

	CA	AOA	LOR
Learners of AG (LAG) (n = 10) (8 males, 2 females)	23.1	19.4	20.8 months
Learners of NG (LNG) (n = 10) (8 males, 2 females)	23.4	19.4	24 months

CA=Current age; AOA=age of arrival in target dialect;
LOR=length of residence

As part of the study, we examined three German vowels: /i/, /y/, and /u/ in CV syllables, /di/, /dy/, and /du/. The stimuli used in this experiment were spoken by native female speakers (average age: 24) of either AG or NG. Participants heard two tokens and were asked to determine whether the two tokens contained either the same vowel (/i/-/i/, /y/-/y/, /u/-/u/) or different vowels (/i/-/y/, /i/-/u/, /u/-/y/) for each of the 3 vowel pair combinations. In each trial, participants either heard both tokens spoken by native AG speakers or by native NG speakers. This allowed us to determine whether the two learner groups were better able to discriminate these vowels when produced in either Northern or Austrian German. For each of the vowel-pairs, listeners heard 8 same tokens (4 of one vowel and 4 of another vowel) and 8 different tokens for a total of 92 tokens (3 vowel pairs x two dialects x 4 same and 8 different vowel pairs) altogether.

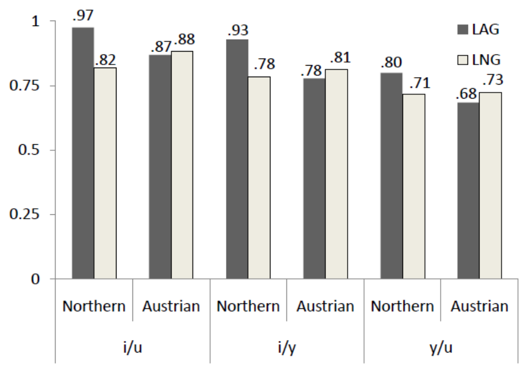
Our first research question asked whether learners are better able to discriminate vowels spoken by speakers of the dialect to which they were exposed. That is, are AG learners more accurate at discriminating vowels spoken by native AG speakers and NG learners more accurate at discriminating vowels spoken by the native NG speakers? Our second research question was whether AG learners differed from NG learners in their ability to discriminate the German vowels regardless of the speaker’s dialect due to the derounding of /y/ in AG which eliminates one cue between /i/ and /y/ while potentially enhancing the /y/-/u/ contrast? These two research questions are answered below.

2.2. Results

The number of correct times each participant indicated that the two vowels in each vowel pair were the "same" or "different" was calculated. For each participant's responses, we calculated A', which controls for response bias (cf. Snodgrass et al. 1985 for a discussion of this measure). A score of .5 indicates chance performance. The results of the analysis are shown in Figure 1. Our first analysis examined our first research question, whether the learner groups were more accurate at discriminating vowels spoken in the dialect to which they were exposed. A two-way (vowel x dialect) ANOVA comparing the AG learners' discrimination accuracy of the NG and AG vowels revealed they were more accurate in discriminating between NG than AG vowels ($F(1,19) = 5.29$, $p < .03$). By contrast, the NG learners' discrimination accuracy of the NG and AG vowels revealed they discriminated vowels produced in both dialects similarly ($F(1,19) = .333$, $p = .566$).

To answer our second research question, whether AG learners differ from NG learners in discriminating the vowels, we ran a two way (group x vowel) ANOVA on the two learner groups' accuracy of the three vowel pairs first for the AG vowels. We found a significant effect of vowel pair ($F(2,19) = 4.94$, $p < .01$), but no significant effect of group ($F(1,19) = .369$, $p = .546$), nor a significant group x vowel interaction ($F(2,1) = .027$, $p = .974$). In other words, neither learner group outperformed the other with the AG vowel pairs, i/u, i/y, and u/y. By contrast, a similar analysis comparing accuracy of discrimination of vowel pairs for the NG vowels revealed a significant effect of group ($F(1,19) = 6.27$, $p = .016$), but no significant effect of vowel, ($F(2,19) = 2.64$, $p = .08$), nor a significant group x vowel interaction ($F(2,1) = .189$, $p = .828$). In this case, the AG learners outperformed the NG learners in perceiving the difference between all three vowel pairs when spoken by NG speakers.

Figure 1: The A' discrimination scores for Learners of Austrian German (LAG) in black bars and Northern German (LNG) in grey bars for both Northern and Austrian German dialects



2.3. Discussion

The results indicate that AG learners were equally accurate as the NG learners at discriminating between vowels spoken by AG speakers and more accurate at discriminating the vowels spoken by the NG speakers. Exposure to derounded /y/ did not hurt discrimination of /i/-/y/ by AG learners. These results suggest that the dialect to which learners were exposed did impact how well vowels were discriminated, although, surprisingly, the AG learners more accurately discriminated *Northern German* vowels than NG learners.

3. EXPERIMENT 2

In this next experiment, we investigated whether the dialect to which learners were exposed influences how accurately learners are able to identify the German vowels /i/, /y/, and /u/ in a variety of phonetic contexts. While discrimination and identification scores are often highly correlated, discrimination of difficult contrasts is often easier for listeners than identification (cf Coughlin, et al. 1998).

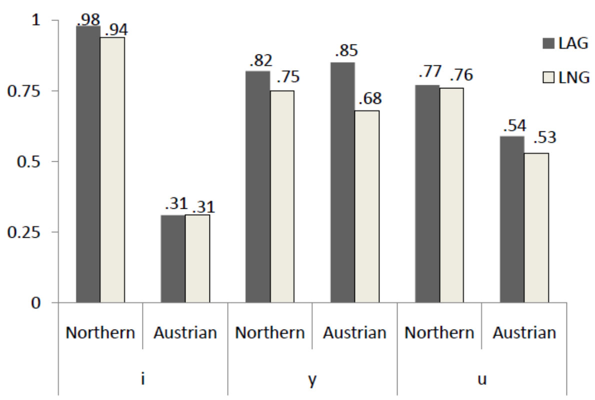
3.1. Methodology

The same learners who participated in experiment 1 participated in this experiment as well. The stimuli, spoken again by native AG or NG female speakers (average age: 24) were either CVC (i.e., *lies*), CV (i.e., sie), or single vowel (i.e., /u/) tokens. The AG tokens displayed diphthongization and derounding characteristic of AG dialects. Participants heard the tokens via headphones presented randomly using the presentation software E-Prime. As they heard the word, participants saw the three German "words" *die, dü, du* on the computer screen corresponding to the high vowels /i/, /y/ and /u/, respectively. Participants were asked to press the key corresponding to the word which contained the vowel they thought they heard.

3.2. Results

We tallied the number of correct identifications for each participant separately for the tokens produced by the native AG and NG speakers (see Figure 2). Visual inspection of the data suggested that the AG learner group was more accurate than the NG learner group at identifying the vowels when produced by the native NG speakers. However, an examination as to whether the two learner groups were more accurate at identifying vowels spoken in the dialect to which they were exposed, revealed that both learner groups were more accurate at identifying vowels spoken in the Northern dialect ($F(2,19) = 48.81$, $p < .0001$). To test whether the AG learners were more accurate than the NG learners at identifying the vowels in either dialect, we first examined how accurately the learners identified vowels spoken by the native NG speakers by submitting the number of correct identifications for each vowel by each participant to a two-way (group x vowel) ANOVA. This analysis revealed no significant effect of group ($F(1,19) = 1.224$, $p = .273$), nor a group x vowel interaction ($F(2,1)=.182$, $p=.835$). A similar analysis examining the learners' accuracy of the AG vowel tokens revealed a similar effect, with no significant effect of group ($F(1,19)=.413$, $p = .523$) nor a group x vowel interaction ($F(2,1)=.110$, $p = .896$). In other words, neither learner group was more accurate at perceiving either the NG or AG vowels.

Figure 2: Identification accuracy of NG and AG vowels by learners of AG (light grey bars) and NG (dark grey bars)



3.3. Discussion

The results of this study determined that neither learner group was more accurate at identifying either Austrian or the Northern German vowels. Moreover, both groups were more accurate at identifying the Northern than Austrian vowels. Surprisingly, both groups were equally poor at identifying the Austrian /i/-tokens. A closer review showed these tokens differed substantially from more typical German /i/ production: AG tokens were strongly diphthongised and produced with a more neutral, if not slightly rounded lip position in contrast with the strongly spread lip productions of Standard German also making it difficult for the researchers to correctly identify as well. The lack of strong spread lip position may have led many participants to judge these tokens as "not /i/" resulting in low accuracy scores for /i/ by both groups.

4. EXPERIMENT 3

In the final experiment, we examined whether AG and NG learners differed in their vowel production.

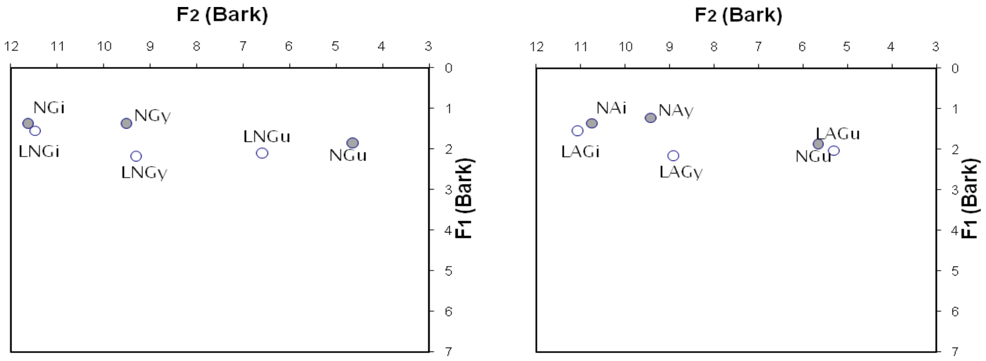
4.1. Methodology

Participants were asked to produce the German vowels in 18 words (6 for each vowel) in the carrier phrase, "Ich sage das Wort__" ('I say the word'). We also asked 4 native AG and 4 native NG speakers to do the same for comparison purposes. For each word spoken, we measured F0 and the first three formants of participants' vowel productions. We normalised them using the Bark Scale (Syrdal and Gopal 1986) by using the following formula: $B = 26.81 / (1+(1960 /F)) - 0.53$, where F = the formant (or fundamental frequency) of each vowel measurement. We compared the learners' productions of the High German vowels, /i/, /y/, /u/ against the productions of the 4 native Northern (NG) and 4 native Austrian (NA) German speakers. Again, we hypothesised that learners exposed to AG would be more likely to produce a contrast between German /y/ and /u/ than would NG learners who lack the additional acoustic cue of /y/ derounding.

4.2. German Production Results

Figure 3 shows the vowel productions of the AG and NG learners and the native AG and NG speakers. Statistical analyses were run on the vowel height, frontedness, and rounding by comparing the native speaker productions to the learners' productions in a series of ANOVAs for each vowel based on vowel height, frontedness, and rounding. The results of these analyses revealed that both learner groups produced German /i/ similarly to the native speakers' productions in terms of height, frontedness, and rounding (all *F's* < 1.176, all *p's* > .34. These analyses also revealed that both learner groups produced German /y/ slightly lower in the vowel space than did the native speaker groups ($F(3,27) = 7.47$, $p < .001$). Finally, analyses also revealed that, while the learners of AG produced German /u/ similarly to native speakers' productions, the NG learners produced the vowel slightly more forward ($F(3,27) = 4.48$, $p < .01$), indeed even more fronted than the AG learners. Thus AG learners did produce at least /u/ more "native-like" than the NG learners.

Figure 3: Left: Vowel Productions of Native Northern German (NG) and learners of NG (LN); Right: Vowel Productions of Native Austrian German (NA) and learners of AG (LA)



4.3. Discussion

The main purpose of this task was to ascertain whether the dialect to which learners were exposed influenced the productions of /i/, /y/ and /u/. Results demonstrated that, while both learner groups differed from the native speakers in how they produced German /y/, the AG learners, but not the NG learners, were able to produce German /u/ similarly to native German speakers. This finding may also be enhanced by the fronted /u/ productions by AG native speakers in contrast to the more back productions of the NG native speakers.

5. GENERAL DISCUSSION

In this study we found that the dialect to which a learner is exposed in part affects how accurately L2 vowels are perceived and produced. AG learners were more accurate than NG learners at discriminating and (to some degree) identifying vowels produced by NG speakers. Likewise, AG learners were also more accurate than learners of NG at producing German /u/. However, the two learner groups did not differ in how accurately they produced German /y/ despite the additional derounding acoustic cue enhancing the /y/-/u/ contrast AG learners had been exposed to, nor did they differ in how accurately they discriminated and identified AG vowels.

The AG learners appear to be more accurate at perceiving and producing (Standard) German vowels than the NG learners. This may be because they were exposed to a dialect that may highlight differences between the very difficult German /y/-/u/ contrast. This exposure may have helped them to identify differences not only between these two vowels, but also helped them to perceive that the German /u/ is much further back than the English /u/. By contrast, the NG learners produced less of a difference between German /y/ and /u/, and produced a more fronted German /u/ similar to English /u/ (Ash 2007). Since native English speakers appear to focus on lip rounding when distinguishing between front and back vowels (Schultheiss 2008), the lesser degree (or lack) of lip rounding for /y/ may have led the learners of AG to not confuse German /y/ and /u/. This would also explain the failure of participants to accurately identify AG /i/-tokens which lacked a strong unrounded production typical of German /i/. In other words, the dialect to which the learners were exposed did in fact affect the learning of these German vowels.

One reason that both groups were more accurate at NG than AG vowels may be that both groups were exposed to Northern German in classes taken prior to departure and in any subsequent German classes afterwards. Moreover, AG learners, while exposed to Austrian German, were also exposed to other (Southern) German varieties as well. It may be that one of the benefits of being exposed to non-standard dialects such as Austrian is the increased amount of variability to which learners are exposed. Previous research has verified that talker variability and phonetic context variability may be important in learning L2 sound contrasts (i.e., Bradlow and Pisoni 1999). In other words, this extra variability may have actually helped, not hindered, learners in perceiving vowels spoken in more standard varieties (akin to NG tokens).

These findings suggest, importantly, that dialect may play an important role in L2 acquisition and suggest that theories of L2 acquisition should take these differences into account. Such findings also have implications for second language teaching. For example, training with L2 some dialects may help native English speakers more accurately learn L2 sounds. Moreover, such training may help learners more easily generalise perception abilities to other dialects. Further research will hopefully inform us of the implications of being exposed to one dialect than another in second language acquisition.

6. REFERENCES

Ash, S. 1996. Freedom of movement: /uw/-fronting in the Midwest. In: Arnold, J., Blake, R., Davidson, B., Schwenter, S. And J. Solomon (eds.), *Sociolinguistic variation: Data, theory, and analysis: Selected papers from NWAV 23 at Stanford*. 3–25. Stanford: Center Study Language and Information.

Baker, W., Smith, L. Under review. The impact of L2 dialect on learning French vowels: Native English speakers learning Quebecois and European French.

Bradlow, A. R., Pisoni, D. B. 1999. Recognition of spoken words by native and non-native listeners: Talker-, listener- and item-related factors. *Journal of the Acoustical Society of America*, 106 (4) , 2074-2085.

Coughlin, M. Kewley-Port, D., Humes, L. E. 1998. The relation between identification and discrimination of vowels in young and elderly listeners. *Journal of the Acoustical Society of America* 104(6). 3597-3607.

Escudero, P., Boersma, P. 2004. Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition* 26. 551-585.

Moosmüller, S. 1987. S*oziophonologische Variation im gegenwärtigen Wiener Deutsch. Eine empirische Untersuchung.* Stuttgart: Franz Steiner.

Russ, C. V. J. 1990. *The dialects of modern German: a linguistic survey.* London: Routledge.

Schultheiss, L. 2008. *Cross-language perception of German vowels by speakers of American English*. MA Thesis, Brigham Young University.

Snodgrass, J.G., Levy-Berger, G., Haydon, M. 1985. *Human experimental psychology.* New York, NY: Oxford University Press.

Syrdal, A.K., Gopal, H. S. 1986. A perceptual model of vowel recognition based on auditory representation of American English vowels. *Journal of the Acoustical Society of America* 79. 1086-1100.

Speech Production and Speech Perception Findings for Native German Speakers Learning English as a Second Language

Bruce L. Smith and Rachel Hayes-Harb
University of Utah
Salt Lake City, UT 84112
U.S.A.
bruce.smith@hsc.utah.edu and hayes-harb@linguistics.utah.edu

ABSTRACT

Although it is reasonable to assume that a relationship between speech production and perception exists for second language (L2) learners, most studies have not found as close a relationship between these two domains as might be expected. Group data regarding this issue can also be misleading when compared with findings among individual speaker-listeners. The present study attempted to evaluate the production-perception relationship by comparing listeners' perceptual accuracy in judging final consonant voicing with the same speakers' acoustic characteristics when producing this voicing contrast. Fifteen native speakers of German and 15 native speakers of American English participated in a listening experiment. Their speech was also recorded as they produced a variety of English target words that contrasted in final consonant voicing, which were then analyzed acoustically. Correlation analyses determined that several acoustic measures of the German subjects' speech were moderately correlated with their perceptual accuracy.

Keywords: L2 speech production, speech perception

1. INTRODUCTION

Many studies have investigated the issue of how speech perception and production are "linked," but the nature of this relationship remains relatively unclear. The nature of the production/perception relationship is even less well understood when considering second language learners' abilities. Although it seems reasonable to assume that some type of relationship between speech production and perception exists for second language learners, carefully conducted studies do not necessarily find as close of a relationship between these two abilities as might be expected. For example, Bradlow, Pisoni, Akahane-Yamada, and Tohkura (1997: 2306-2307) "examined the rank-order correlation between improvement in perception and production across all subjects to test the hypothesis that subjects who show the most perceptual learning also show the most improvement in production. However, …there is no such correlation...." More specifically, they found that although perceptual learning seemed to result in a general transfer that was associated with improved production, the production/perception relationship within individual subjects did not hold in many instances. More recently, Hattori and Iverson (2009) also found that native Japanese speakers learning English as a second language did not demonstrate a strong relationship between perception and production. One important observation by Bradlow et al. (1997) and Hattori and Iverson (2009) was that considerable variation existed when examining the perception and production performance of individual subjects. Despite general improvements that their groups of subjects demonstrated in production of the /l/-/r/ contrast subsequent to perceptual training, individual subjects showed many different responses to such training. For example, some subjects showed little or no change as a result of the training they received, whereas others with similar pre-training perception "profiles" changed considerably in their production patterns after training. In addition, based on an investigation of native Mandarin listeners' performance in making voiced/voiceless judgments about non-native (English) contrasts produced by native English vs. native Mandarin speakers, Hayes-Harb, Smith, Bent, and Bradlow (2008) determined that the nature of subjects' perception performance was partly a function of who was listening and who was talking, as well as the L2 proficiency of both the listeners and the speakers. However, because the native Mandarin and native English speakers whose production patterns were examined were not the same subjects for a given language as those whose perception patterns were assessed, it was not possible to make direct comparisons of perception and production within either language group. A primary purpose of the present study was thus to examine perception and production by the same group of speakers/listeners to determine if specific patterns could be found that would provide evidence of a clear relationship between subjects' production and perception. To this end, several production and perception patterns were evaluated individually, followed by an attempt to integrate the production and perception findings. The primary question of interest was whether subjects with more native-like English production patterns would also show more accurate perception of the word-final voiced-voiceless contrast. Based on the existing literature (e.g., Bradlow et al. 1997; Hattori and Iverson 2009), it was anticipated that only a modest relationship would be observed.

2. SUBJECTS AND PROCEDURES

In the present study we examined the relationship between listening and production performance by native and non-native speakers of English by conducting a more detailed analysis of listening task data previously reported in Smith et al. (2009). We also compared these results to those of a production task conducted with the same subjects that has not been reported previously. Both the listening (section 2.1) and production (section 2.2) tasks were conducted during a single hour-long session with the same group of 30 subjects.

2.1 Listening Task

Smith et al. (2009; Experiment 2) examined the intelligibility of native English and German-accented English speech to 15 native English and 15 native German listeners. Subjects heard 3 randomized repetitions of 8 English words ("cob, cop, cub, cup, lied, light, toad, tote") that had been produced by each of 6 native English and 6 native German speakers. The listeners made voiced/voiceless (VD/VL) judgments of the final consonants in the target words (8 words x 3 repetitions x 12 speakers).

2.2 Production Task

Following the listening task (Smith et al. 2009), the same native English and native German speakers produced a set of English sentences. The acoustic properties of these productions were of interest. Subjects read a list of sentences that contained a variety of target words embedded in a carrier phrase in final position of the sentence and in sentence internal, but non-pre-pausal, position. The target words were "cob, cop, cub, cup, lied, light, log, lock, toad, tote," i.e., a set of five minimal pairs that differed in voicing of the final consonant. Several other non-minimal pair words were also produced in the same sentences to help reduce speakers' awareness of the minimal pairs.

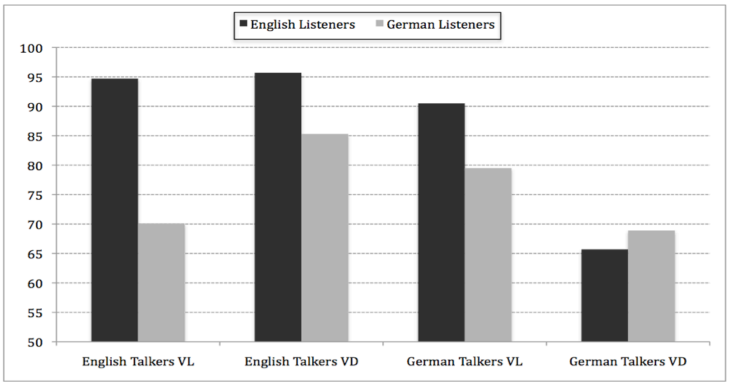
3. RESULTS AND DISCUSSION

3.1 Speech Perception Findings

As reported by Smith et al. (2009), the native English subjects averaged 95% accuracy (range: 82-98%) in judging the voicing of word-final consonant productions by other native English speakers but only 77% accuracy (range: 69-81%) judging productions by the German subjects (paired $t=30.370$; $df=14$; $p<.0001$). The German subjects averaged 78% accuracy (range: 62-92%) judging productions by the English speakers and 74% accuracy (range=64-87%) judging the English productions of other Germans (paired $t=3.106$; $df=14$; $p < .01$). The German and English listeners, as groups, both showed significant correlations ($r=0.87$; $p < .0001$, for both groups) when assessing their accuracy judging consonant voicing for German compared with English speakers, i.e., both groups of listeners performed at approximately the same level of accuracy when listening to voicing patterns produced by native English and German speakers. In addition to considering the general accuracy of listeners' perceptual performance (Smith et al. 2009), it is also of interest to compare their accuracy for voiced vs. voiceless stop productions. As shown in Figure 1, for example, the

English subjects were equally accurate in making judgments of voiced (96%) vs. voiceless (95%) targets produced by other native English speakers (paired t=1.258; df=14; NS). However, this was not the case when they were listening to the German speakers. The English listeners were 90% accurate judging voiceless stop productions of the German speakers, but only 66% accurate judging the Germans' voiced stop targets (paired t=8.669; df=14; p < .0001), which is shown by the third and fourth black bars in Figure 1. Although the difference is smaller, the German listeners were also significantly more accurate judging voiceless (79%) vs. voiced (69%) stop targets produced by German subjects (paired t=2.915; df=14; p < .02), i.e., the third and fourth gray bars. In contrast, the German listeners were more accurate judging voiced (85%) vs. voiceless (70%) targets produced by the English speakers (paired t=5.672; df=14; p < .0001), shown by the first and second gray bars.

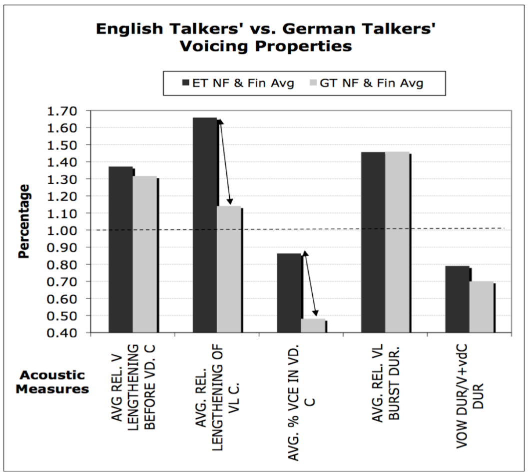
Figure 1. A comparison of English and German listeners' perceptual accuracy for voiced (VD) vs. voiceless (VL) stop targets produced by English vs. German speakers.



3.2 Speech Production Findings

Figure 2 shows general findings regarding production patterns observed for the English and German speakers. Temporal-acoustic comparisons that were made included: (1) relative vowel lengthening preceding final voiced vs. final voiceless stops, (2) relative lengthening of final voiceless vs. voiced stop closure duration, (3) proportion of glottal pulsing during final voiced stop closure (i.e., voiced targets), (4) relative lengthening of voiceless vs. voiced stop consonant release bursts, and (5) relative vowel lengthening before final voiced vs. voiceless stops in vowel + consonant "rhymes" (i.e., essentially the same as the first measure above except that it compares the ratio of vowel duration relative to the combined vowel and final consonant when the consonant target is voiced vs. voiceless). To provide as representative a sample as possible, the various measures were determined on the basis of productions in both sentence final and non-final position that were averaged together. As can be seen in Figure 2 (indicated by the arrows), the two greatest duration differences between the English and German speakers were for relative lengthening of voiceless compared to voiced stop closure duration and proportion of "voicing" (i.e., glottal vibrations) during voiced stop consonant targets. More specifically, the English speakers' voiceless stop closure durations were approximately 68% longer than their voiced closure durations. This is shown in Figure 2 as a value of 1.68, where a value of 1.00 would indicate no difference in duration for voiceless vs. voiced stop closures. In contrast, the German speakers showed a difference of only about 18% (1.18) for voiceless vs. voiced closure durations (Welch-corrected, unpaired t=4.542; df=21; p < .001). With regard to the proportion of voicing that occurs in voiced consonants, a value of 1.00 indicates that the entire closure interval (100%) manifested glottal vibrations, whereas values less than 1.00 indicate that only a portion of the closure interval was voiced. As can be seen in this figure, the English speakers manifested voicing (i.e., glottal pulsing) during approximately 87% of the closure interval compared to about 50% by the German speakers (Welch-corrected unpaired t=7.291; df=24; p < .0001). The other comparisons that were made (e.g., vowel lengthening preceding voiced vs. voiceless stops and voiceless vs. voiced burst duration) showed little or no difference between the two groups.

Figure 2. A comparison of temporal patterns for the English speakers (ET) vs. the German speakers (GT) for non-final (NF) and final (Fin) position words averaged together. The horizontal dashed line at 1.0 indicates no difference between two temporal measures being compared (e.g., vowels preceding voiced vs. voiceless stops).

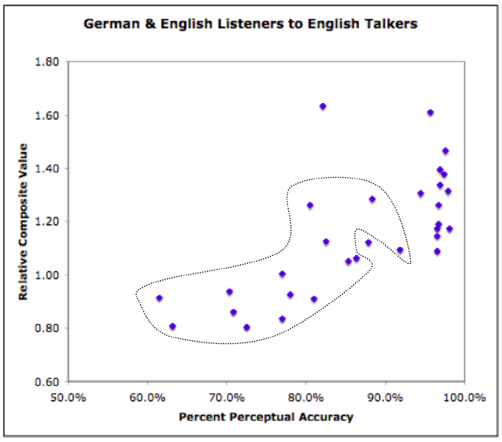


An additional production-based measure was a "composite" voicing calculation for each of the 15 German and 15 English subjects. This composite measure consisted of an average value for each speaker's relative vowel lengthening preceding voiced vs. voiceless stops, relative voiceless vs. voiced consonant closure duration, and relative amount of voicing during voiced consonant closure. The average composite voicing value for the English speakers was 1.31 compared with a value of .99 for the German speakers (Welch corrected, unpaired t=5.353, df=27, $p < .0001$). It was also determined that there was quite limited overlap among the native English and German speakers for this composite measure. That is, only about one-third of the German speakers show a composite value within or reasonably close to the range shown by the English speakers, with only two of the German speakers well within the native range. Furthermore, it was found that each of the 15 English speakers showed some amount of composite lengthening, from approximately 10% to 65% (i.e., 1.10-1.65). In contrast, six of the German speakers showed composite lengthening (ranging from just over 1.00 to not quite 1.30), whereas the other nine subjects showed composite "shortening" (by as much as approximately 20%). Overall, the range of values shown by the German speakers (composite values of .80-1.30, a range of .5) was similar in magnitude to the range shown by the English speakers (composite values of 1.10-1.65, a range of .55).

3.3 Speech Production and Perception Interactions

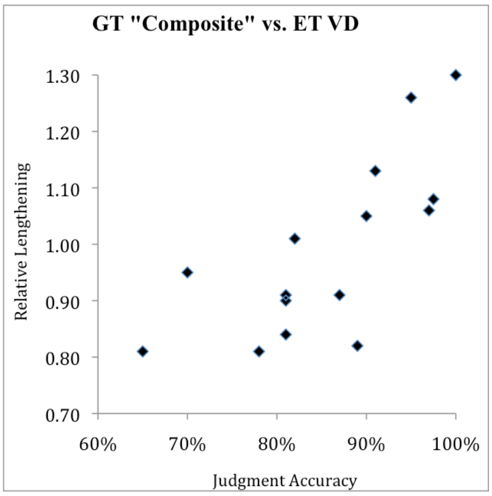
Because the same subjects served as both speakers and listeners and there was a substantial range of performance in both domains across the 30 subjects, it was possible to compare their production and perception performance to determine what relationship might be observed between these two domains. For example, Figure 3 shows results for the 15 German subjects and the 15 English subjects in terms of perceptual accuracy judging final consonant voicing (when listening to native English speakers) vs. their relative "composite" production value for voicing of final stops (averaged across final and non-final positions). The 15 subjects enclosed within the dotted line are the German speakers/listeners; the English speakers/listeners are those 15 outside the dotted line. The production/perception correlation for these 30 subjects was 0.70 ($p < .0001$). As can also be seen, however, the English subjects showed a considerable range of production performance for their composite voicing values, but only a limited range of accuracy for their final consonant voicing judgments (due to a "ceiling effect"). Because of possible impact this perceptual distribution might have, a correlation analysis was conducted on the production and perception data for only the 15 German speakers/listeners. Despite the smaller number of subjects, however, a very similar correlation was obtained ($r=0.69$, $p < .01$).

Figure 3. A comparison of the 15 German subjects' and the 15 English subjects' perceptual accuracy judging final consonant voicing vs. their relative "composite" production value for voicing of final stops.



Because of the phonological differences between English and German for voicing of final stops, separate comparisons were also made for the "composite" production measure relative to perceptual accuracy for voiced vs. voiceless targets. Perceptual accuracy for voiced stop targets was significantly correlated with the composite production measure for the German speakers/listeners when judging German speakers ($r=0.67$; $p < .01$) and when judging native English speakers ($r=0.76$; $p < .001$; see Figure 4).

Figure 4. A comparison of the native German (GT) speakers' production composite measure of voicing vs. their listening accuracy for voiced (VD) targets produced by native speakers of English ($r=0.76$, $p < .001$).



In contrast to voiced target productions, however, correlations for the German listeners were not significant when judging voiceless productions by the English ($r=0.44$; ns) or the German speakers ($r=0.19$; NS). Similarly, comparisons between the production composite measure vs. the voiced and voiceless targets in the listening task were not significant for the native English speakers/listeners ($r=-0.16$; ns and $r=-0.49$; ns respectively). In addition to the correlation findings for the composite measure of voicing compared to the accuracy of perceptual judgments of final stop voicing, several other statistically significant correlations were found between the subjects' individual acoustic measures and their perceptual accuracy. Overall, vowel duration before voiced vs. voiceless stops and voiceless stop closure duration relative to voiced stop closure duration were the two individual acoustic measures that tended to show the greatest number of significant correlations with perceptual accuracy. Neither proportion of voicing during consonant closure nor relative release burst duration showed a significant correlation for the German listeners to the English or German speakers' productions. In general, however, the composite measure tended to have more significant correlations than any of the individual acoustic parameters relative to the accuracy of the German listeners' perceptual judgments. There were no significant correlations for the native English subjects with any of the English or German speakers' acoustic characteristics of final voiced vs. voiceless stops.

4. CONCLUSION

A number of differences were observed for both production and perception performance when comparing native German subjects to native English subjects. Correlation analyses showed that relative vowel duration prior to voiced vs. voiceless stops and relative consonant closure duration for voiceless vs. voiced stops, as well as a "composite" measure of several acoustic parameters, were all modestly correlated with the native German subjects' perceptual accuracy. Other acoustic measures such as relative burst release durations and proportion of "voicing" during voiced stop targets did not correlate with the accuracy of perceptual judgments made by the native German listeners. No significant correlations between perceptual accuracy and any of the acoustic measures were observed for the 15 native English speakers. This may be partly the result of a ceiling effect; that is, perceptual accuracy was quite high for most of the native English subjects, which limited the range of performance and thus affected the likelihood of significant correlations between production and perception. Differences between the English and German subjects in both production and perception were also observed when comparing voiced vs. voiceless stops targets. For example, Figure 2 showed that when speaking English, the native Germans had certain acoustic patterns (e.g., vowel lengthening before voiced vs. voiceless stop targets) that were quite similar to patterns shown by the native English speakers. However, other acoustic measures (e.g., voiceless vs. voiced consonant closure duration and proportion of voicing during voiced targets) produced by the German speakers were quite different from the patterns shown by native English speakers. In terms of perception, the German listeners were more accurate judging voiceless vs. voiced stops produced by other German speakers, but they were more accurate judging voiced vs. voiceless stops produced by native English speakers (Figure 1). It is not clear why the German listeners would perform differently for German vs. English speech, but it may have to do with the fact that the German speakers did not produce some of the acoustic cues associated with voicing (e.g., closure duration differences between voiced and voiceless stops). This would presumably make their voiced targets harder to judge as accurately as voiceless targets. Why they were more accurate with voiced than voiceless targets produced by the native English speakers (vs. being equally accurate with both) is unclear. The significant correlation between the composite acoustic measure of their speech and the perceptual judgments of the German speakers may reflect various factors that affect both production and perception, such as length of time the L2 has been studied, amount of time spent in the L2 country, degree of motivation, natural linguistic abilities, etc. In contrast, the lack of significant correlations between production and perception for the native English speakers listening to other native English speakers is not surprising since all listeners were quite accurate in their perceptual judgments. Furthermore, in contrast to the L2 speakers, the range of performance in the native English speakers' acoustic characteristics associated with their voiced and voiceless stops presumably has nothing to do with their speaking abilities, since they are native speakers, but rather may relate to factors such as dialect differences and possibly "inherent" differences among speakers. In contrast, dialect differences may influence L2 performance less than factors such as length of residence, age of arrival, motivation to learn, and so forth.

5. REFERENCES

Bradlow, A., Pisoni, D., Akahane-Yamada, R., Tohkura, Y. 1997. Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *J. of the Acoustical Society of America* 101, 2299-2310.

Hattori, K., Iverson, P. 2009. English /r/ - /l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy. *J. of the Acoustical Society of America* 125. 469-479.

Hayes-Harb, R., Smith, B., Bent, T., Bradlow, A. 2008. The Interlanguage speech intelligibility benefit for native speakers of Mandarin: Production and perception of English word-final voicing contrasts. *J. of Phonetics* 36. 664-679.

Smith, B., Hayes-Harb, R., Bruss, M., Harker, A. 2009. Production and perception of voicing and devoicing in similar German and English word pairs by native speakers of German, *J. of Phonetics*, 37, 257-275.

Categorizing Mandarin Tones into Listeners' Native Prosodic Categories: The Role of Phonetic Properties

Connie K. So[1], and *Catherine T. Best*[1,2]

[1] MARCS Auditory Laboratories, University of Western Sydney, Australia
[2] Haskins Laboratories, New Haven, CT, USA
c.so@uws.edu.au, c.best@uws.edu.au

ABSTRACT

This study examined whether native speakers of non-tone languages (Australian English, and French) were able to perceive foreign Mandarin tones in a sentence environment, according to their native prosodic categories. Results found that both English and French speakers were able to perceptually categorize foreign tones into their intonational categories (i-Categories), and that categorizations were based on the contextual phonetic similarities of the pitch contours they perceived between Mandarin tones and their native i-Categories. Results also showed that French speakers, but not English speakers, were able to detect the fine-detailed phonetic feature differences between Tone 3 and Tone 4 (low (falling) tone vs. high-falling tone). The findings support the new assumption of PAM for suprasegmentals (So & Best, 2008) that non-native prosodic categories (e.g., lexical tones) will be assimilated to the categories of listeners' native prosodic system (e.g., intonation). In addition, rhythmic differences among languages may also contribute to perception of non-native tones.

Keywords: Lexical tone perception, Cross language perception, Perceptual Assimilation Model (PAM), Phonetic influence.

1. INTRODUCTION

Studies have shown that listeners can assimilate non-native tones to the categories of their native prosodic systems, such as tone, pitch-accent, and intonation (So, 2006; So & Best, 2008; So & Best, 2010), in ways that appear consistent with the assumptions of the Perceptual Assimilation Model (Best, 1995, PAM). This raises an important question as to how adults perceive non-native lexical tones. Do they perceive foreign tones according to the pitch patterns of the intonational categories (i-Categories) in their native prosodic system (e.g., rising pitch patterns for questions)? A recent study (So & Best, 2008) has demonstrated that native English (NE) listeners can perceive non-native tones (on individual single words) in terms of their i-Categories. In general, Mandarin Tone 1 (*High level*) is perceived as *Flat Pitch*, Tone 2 (*mid-rising*) as *Question*, Tone 3 (*falling-rising*) as *Uncertainty* (Some NE listeners perceived it as *Question*), and Tone 4 (*high falling*) as *Statement*. The findings supported the assumption that non-native tonal categories (e.g., lexical tones) will be assimilated to the categories of listeners' native prosodic system. The study also suggested that NE listeners assimilated the phonetic properties of Mandarin tones (e.g., pitch patterns) to those of English i-Categories, when both substantially share similar phonetic features.

However, how do listeners perceive foreign tones when they are embedded in a sentential environment? Do they perceive the foreign tones according to their prosodic categories of their native languages? In addition, it is well documented that the effects of tonal coarticulation (anticipation and carryover) will be involved in connected speech (Xu, 1994, 1997). Do the contextually-varying phonetic characteristics (e.g., rising and falling pitch patterns) of foreign tones in sentences affect how listeners from non-tonal languages perceive them? Further, since languages can be classified as stress-timed and syllable-timed languages according to their rhythmic, it would be important and of theoretical interest to know whether there is any difference in the perceptual assimilations between native speakers of non-tonal languages with different rhythmic properties. Rhythmicity is an intrinsic characteristic of a language's prosodic system but refers to temporal (timing) rather than spectral patterning (F0) of syllables in the language.

To answer the above questions, this study examined the perception of Mandarin tones by native speakers of two non-tonal languages – English, a stress timed or stress accented language (Beckman, 1986), and French, a syllable timed language without an accent system (Fox, 2000). The new assumption of PAM for suprasegmentals (So & Best, 2008) was tested by investigating how native speakers of these non-tone language groups perceived Mandarin tones in a sentence frame according to their native intonational categories (*Flat pitch, Question, Statement, and Exclamation*). Since the tones were in a sentential environment, listeners' categorizations should be based on the *contextual phonetic similarities* of the pitch contours they perceived between Mandarin tones (in the sentential form) and their native i-Categories. Thus, it was predicted that they would perceive Tone 1 as *Statement* (this level tone might be perceived as a tone with a slight falling movement) rather than *Flat Pitch*, Tone 2 as *Question*, Tone 3 (a low falling/low tone) as *Statement*, and Tone 4 as *Exclamation* (a falling tone with a greater falling pitch movement with a steeper slope) rather than *Statement*.

2. METHOD

2.1. Participants

Thirty Australian NE speakers (18 - 24 years of age) and thirty native French (NF) speakers (21 - 37 years of age) were recruited as participants. They were all either undergraduate students at the University of Western Sydney, who received course credits after they complete the experiment, or residents living in Sydney at the time of the experiment who received AUD $40 for their participation. They had neither learned Mandarin nor received formal musical training, as previous studies have shown that listeners with musical training outperformed those without such training in both production and perception tasks with non-native tones (Alexander, Wong, & Bradlow, 2005; Burnham & Brooker, 2002; Gottfried & Riester, 2000). Before they performed the experiment, they all passed a pure-tone hearing screening (250- 8000 Hz at 25 dB HL).

2.2. Stimuli

The stimuli for this study were produced by three native Mandarin speakers (mean age: 24 years). They were asked to produce the four Mandarin tones on the syllable /fu/ in a statement frame. (in Chinese PinYin: *xia4 yi1 ge4 shi4 X zi4*, where the number indicates the tone on the word, and X indicates the target word; the English gloss is "the next one is the X word"). The syllable /fu/ was selected because its pronunciation is close to the one for the English word, *fool,* and similar to that of the French word, *fou*, which means "crazy". Five tokens of each target word (/fu/ with each of the four tones) were produced by each speaker. Among them, 3 samples per tone word per speaker were verified perceptually by another three native Mandarin speakers (mean age: 27.7 years) to ensure the selected stimuli were intelligible to native Mandarin speakers. All of the perceptual stimuli were correctly identified by the native speakers. Note that, Tone 3 (rising-falling) is produced as a low level or a low falling tone in connected speech

2.3. Procedure

Participants were asked to categorize randomized individual presentations of 72 trials of these stimulus sentences (3 speakers x 4 tones x 3 tokens per tone x 2 repetitions) into four English/French i-Categories -- *Flat pitch* (a level pitch without movement), *Question* (a rising pitch contour), *Statement* (a falling contour), and *Exclamation* (it involves a greater falling pitch movement and steeper slope). In this study, the same i-categories were provided to the both the NE and NF groups as these i-categories (e.g., *Question* and *Statement*) are common in both languages, and share similar pitch contours (Hardison, 2004; Ladd, 1996; Post, 2002). During the experiment, the stimulus statements were presented individually from a PC screen, on which five buttons were provided, corresponding to the four i-Categories and a 5th button labelled as Unknown. Listeners were instructed to select *Unknown* when they could not identify a tone into any i-

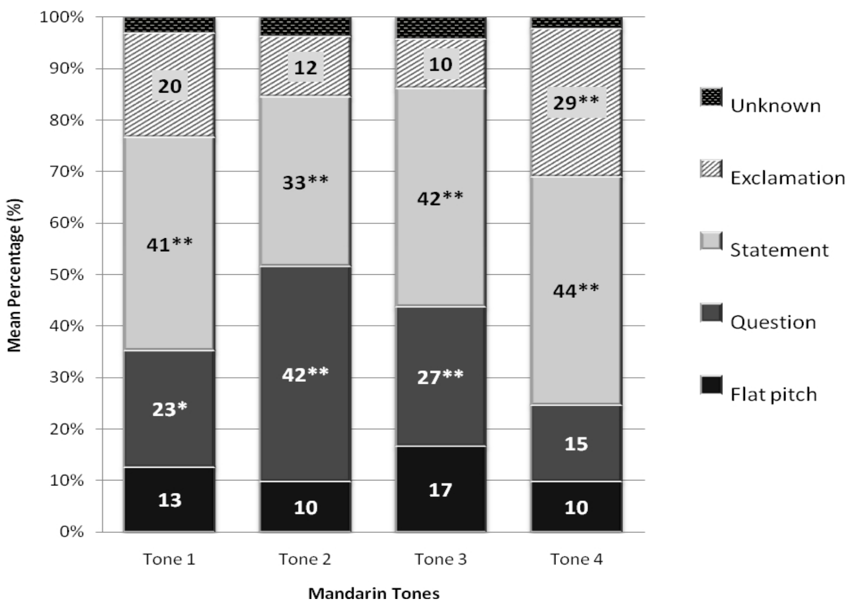
Category. (Note that although the labels were in English (i.e., no French translations), all French listeners were also fluent in English, and careful instructions and a trial block that consisted of 8 samples had been given to both NE and NF listeners before the experiment to ensure they all understood that they were asked to categorize the perceived target tones into the i-categories of their own native language.)

3. RESULTS

3.1. Tonal categorization of Native English (NE) speakers

NE listeners' tonal categorizations for each tone (in %) are shown in Figure 1. Individual t-tests were carried out to test each i-Category mean for each target tone, against chance level (20%). The results confirmed that the means of the following i-Categories for their target tones were all significantly above the chance level (20%): *Question* [$t(26) = 1.880$, $p < .05$] and *Statement* for Tone 1 [$t(29) = 7.880$, $p < .01$], *Question* [$t(29) = 5.799$, $p < .01$] and *Statement* [$t(28) = 4.950$, $p < .01$] for Tone 2, *Question* and *Statement* for Tone 3 [$t(28) = 2.738$ and 6.538, *ps* <.01], and *Statement* and *Exclamation* for Tone 4 [$t(28) = 9.029$ and 4.194, *ps* <.01]. A Chi-square test revealed a significant association between Tones (4) and i-Categories (5), χ^2 (12) = 165.794, $p < 0.001$. A further mixed design 2-way ANOVA[1] (Tone x i-Category) found no significant effect of Tone (*n.s.*), but a significant effect of i-Category [$F(3,389) = 51.952$, $p < .001$] on listeners' mean assimilation percentage (%). The Tone x i-Category interaction was also significant [$F(9,389) = 7.722$, $p < .001$].

Figure 1: Listeners' tonal categorizations for each tone (in %). The total number of responses for each tone category was 540. Categories that have 5% or less are not labelled. The symbols * ($p<.05$) and ** ($p<.01$) show that the mean of the i-Category is above the chance level (20%).

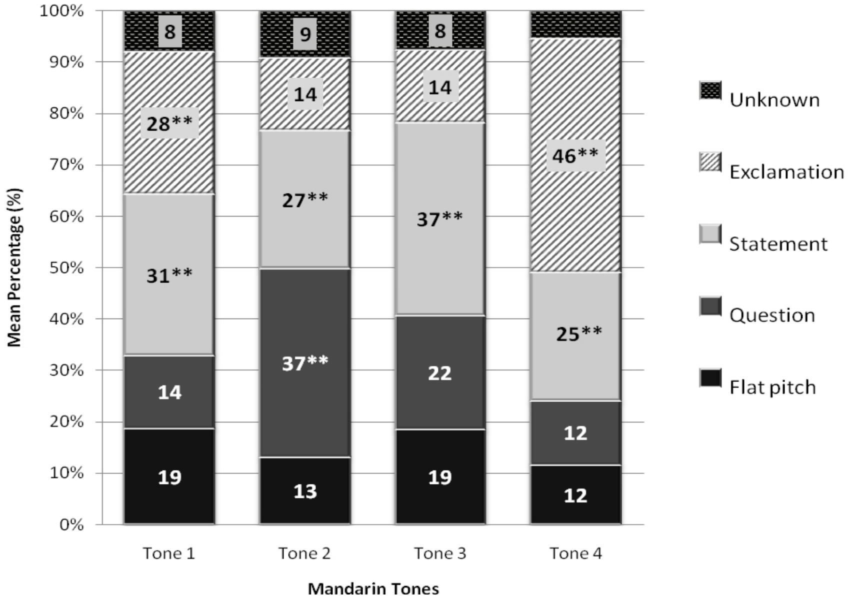


Individual 1-way ANOVAs for the four tones were carried out to investigate the i-Category effect for each tone target. It was found that the i-Category effect was significant for each tone: Tone 1 [$F(3,105) = 21.615$, $p < .0001$], Tone 2 [$F(3,92) = 13.077$, $p < .001$], Tone 3 [$F(3,95) = 13.65$, $p < .001$], and Tone 4 [$F(3, 97) = 30.005$, $p < .001$]. Post-hoc HSD Tukey tests further indicated the following results for each tone. For Tone 1, the mean percentage (%) of *Statement* assimilations (41%) was significantly greater than each of the other counterparts: *Flat Pitch* (13%), *Question* (23%), and *Exclamation* (20%) assimilations (*ps* < .05). For Tone 2, the mean percentage of *Question* assimilations (42%) was significantly greater than those of *Flat Pitch* (10%), and *Exclamation* (12%) assimilations (*ps* < .01), but did not differ significantly from that of *Statement* assimilations (33%; *n.s.*), which was selected significantly more often than *Flat Pitch* and *Exclamation* assimilations (*ps* < .01). For Tone 3, the mean percentage of *Statement* assimilations (42%) was significantly greater than those of *Flat Pitch* (17%), *Question* (27%), and *Exclamation* (10%) assimilations (*ps* < .01). In addition, the mean percentage of *Question* assimilations was significantly greater than that of *Exclamation* assimilations ($p < .05$). For Tone 4, the mean percentage of *Statement* assimilations (44%) was significantly greater than those of *Flat Pitch* (10%), *Question* (15%), and *Exclamation* (29%) assimilations (*ps* < .01). The mean percentage of *Exclamation* assimilations was significantly greater than those of *Flat Pitch*, and *Question* assimilations (*ps* < .01).

3.2. Tonal categorization of Native French (NF) speakers

NF listeners' tonal categorizations for each tone (in %) are shown in Figure 2. Individual t-tests were carried out to test each i-Category mean for each target tone, against chance of 20%. The results confirmed the means of the following i-Categories for their target tones were all significantly above the chance level (20%): *Exclamation* [$t(27) = 3.691$, $p <.001$] and *Statement* for Tone 1 [$t(29) = 4.198$, $p <.001$], *Question* [$t(29) = 4.286$, $p <.001$] and *Statement* [$t(28) = 3.270$, $p <.001$] for Tone 2, *Statement* for Tone 3 [$t(29) = 4.957$, $p <.001$], and *Statement* [$t(26) = 3.463$, $p <.001$] and *Exclamation* [$t(29) = 6.461$, $p <.001$] for Tone 4. A Chi-square test revealed a significant association between Tones (4) and i-Categories (5), χ^2 (12) = 262.22, $p < 0.001$. A further mixed design 2-way ANOVA[2] (Tone x i-Category) found no significant effect of Tone (*n.s.*), but a significant effect of i-Category [$F(3, 403) = 13.045$, $p <.001$] on listeners' mean assimilation percentage (%). The Tone x i-Category interaction was also significant [$F(3,403) = 10.176$, $p <.001$].

Figure 2: Listeners' tonal categorizations for each tone (in %). The total number of responses for each tone category was 540. Categories that have 5% or less are not labelled. The symbol ** ($p<.01$) shows that the mean of the i-Category is above the chance level (20%).



Individual 1-way ANOVAs for the four tones were carried out to investigate the i-Category effect for each tone target. It was found that the i-Category effect was significant for each tone: Tone 1 [$F(3,102) = 5.037$, $p <.001$], Tone 2 [$F(3,102) = 10.327$, $p <.001$], Tone 3 [$F(3, 102) = 7.132$, $p < .001$], and Tone 4 [$F(3, 95) = 20.742$, $p < .001$]. Post-hoc HSD Tukey tests further indicated the following results for each tone. For Tone 1, the mean percentage (%) of *Statement* assimilations (31%) was significantly greater than *Question* (14%; $p <0.01$) and *Flat Pitch* assimilations (19%; $p <0.05$). In addition, the mean of *Exclamation* assimilations (28%) was also significantly greater than that of *Question* assimilations ($p <0.05$). For Tone 2, the mean percentage of *Question* assimilations (37%) was significantly greater than those of *Flat Pitch* (13%), and

Exclamation (14%) assimilations (*ps* < .001), but did not differ significantly from that of *Statement* assimilations (27%; *n.s.*). The mean % of *Statement* assimilations was significantly greater than that of *Flat Pitch* assimilations ($p<0.01$). For Tone 3, the mean percentage of *Statement* assimilations (37%) was significantly greater than those of *Flat Pitch* (19%), *Question* (22%), and *Exclamation* (14%) assimilations (*ps* < .01). For Tone 4, the mean percentage of *Exclamation* assimilations (46%) was significantly greater than those of *Flat Pitch* (12%), *Question* (12%), and *Statement* (25%) assimilations (*ps* < .001). The mean percentage of *Statement* assimilations was significantly greater than that of *Flat Pitch* assimilations ($p < .01$).

4. DISCUSSION

The results, as expected, clearly showed that both NE and NF speakers were able to categorize Mandarin tones (embedded in a sentence frame) into their native prosodic categories. Their selections depended on *contextual phonetic similarities* between the pitch contours of the prosodic categories of Mandarin and those of English and of French. For Australian English speakers, Tone 1 was more perceived as *Statement*, Tone 2 was more perceived as *Question*. However, NEs also assimilated Tone 2 to their *Statement* i-category, which might reflect their perception of the overall falling pitch movement of the target word within the sentence frame (will be discussed later in this section). Tone 3 was perceived more as *Statement,* but it was sometimes perceived as *Question* (will be discussed later in this section). Tone 4 was perceived mainly as *Statement*, although the pitch contour of Tone 4 involves a greater falling movement and steeper slope (-162.09 Hz from its maximum pitch, on average) than that of Tone 3 (-51.64 Hz, on average). NEs were clearly able to perceive the falling pitch movement, but failed to perceive the fine-gained phonetic difference in rate/extent of F0 decline. Similarly, NFs were able to assimilate Mandarin tones into their i-Categories. NFs perceived Tone 1 as both *Statement* and *Exclamation*. This may due to the fact that both share the falling pitch feature (either with the *Statement* i-category or with the sentence environment), and the involvement of the high pitch of Tone 1 together of the descending (falling) pitch direction (due to the sentence frame) might obscure their perception. NFs perceived Tone 2 primarily as *Question*, then *Statement* (will be discussed later in this section). They assimilated Tone 3 to *Statement*, and Tone 4 to *Exclamation,* and this pattern is different from that of the NEs.

Listeners' categorizations of the Mandarin tones in a sentence environment were affected by both the overall descending pitch tendency of the sentence frame, and the contextually-varying phonetic characteristics (e.g., rising and falling pitch contours) of the target word's foreign tones within the sentences. The former characteristic is clearly exerted some perceptual influences on NEs and NFs, and caused them sometimes to perceive Tone 2, the rising tone, as their *Statement* category (NE: 33%; NF: 27%). The latter characteristic, however, actually involves tonal coarticulation effects, both anticipation and carry-over effects (Xu, 1994, 1997), which might also obscure listeners' categorizations of the target tones to some extent. For example, listeners sometimes perceived Tone 3 as they had perceived Tone 2. Tone 3 is generally produced as a low level/falling in a sentence environment, and the production of the word following the target word, *zi*, involves a higher pitch of the onset, because it has a high falling tone, Tone 4). Thus, the anticipatory coarticulation effect might create an illusion of a rising pitch pattern on the target word to the listeners. This explanation may also apply to NEs' perception of Tone 1, which 23% of time was perceived in the same way as they had perceived Tone 2, a rising tone.

The rhythmic properties of different language classes also appear to be influencing the perception of non-native tones by non-tone language speakers. Native speakers of French, a syllable timed language, (Fox, 2000) appear to be able to perceive the fine-gained tonal phonetic features or feature changes across the target words in the sentence context better than the native speakers of English, a stress timed language, (Fox, 2000), because Mandarin is also a syllable timed language (Chen, Robb, Gilbert, & Lerman, 2001; Smit, 2004). That is, both French and Mandarin, but not English, syllables tend to maintain a regular timing interval (syllable duration). These similar rhythmic properties might help NFs locate the pitch contour of the target word better during the perception of the whole sentence. As a result, they were better able to perceive Tone 4 as *Exclamation*, produced a category involves a greater falling movement and steeper slope, than English listeners were.

5. CONCLUSION

The results of the present study indicated that both NE and NF listeners assimilated non-native tones to their native intonational categories (Tone 1 and Tone 3 as *Statement*, Tone 2 as *Question*, and Tone 4 as *Statement* for NE speakers but as *Exclamation* for NF speakers) that share phonetic similarities with those of Mandarin tones in a sentence environment. However, their perception appears to be affected to some extent by the effect of tonal coarticulation in connected speech. In additional, while NE listeners were unable to detect the fine phonetic difference between *Statement* for Tone 3 (involving a slight falling pitch pattern) and *Exclamation* for Tone 4 (involving a dramatic falling pitch pattern), NF listeners were better able to perceive the phonetic difference between Mandarin Tone 3 and Tone 4, which they perceived as similar to two native i-Categories, whereas NE listeners perceived them as exemplars of a single native i-Category, *Statement*. The difference may be attributed to differences in the rhythmic properties of their native languages. Thus, the overall results further affirm the assumption that non-tone listeners (NE and NF) assimilate non-native prosodic categories (e.g., tones) to their native prosodic categories based on the (contextual) phonetic similarities they perceive. Moreover, they add an important novel observation: not only the spectral (intonational) but also the rhythmic (temporal) properties of the listener's native prosodic system can have significant effects on assimilation of non-native tones to native prosodic categories.

6. ACKNOWLEDGEMENTS

This work is supported by the Australian Research Council (PI: C. So). We thank Arman Abrahamyan and Asako Kafafy for their assistance with data collection.

7. REFERENCES

Alexander, J. A., Wong, P. C. M. and Bradlow, A. R. 2005. Lexical tone perception by musicians and non-musicians. *Proceedings of Interspeech 2005*. Lisbon, Portugal.

Beckman, M. E. 1986. *Stress and non-stress accent*. Dordrecht: Foris Publications.

Best, C. T. 1995. A direct realist view of cross-language speech perception. In Strange, W. (e.d.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, 171-232, York Press.

Burnham, D. and Brooker, R. 2002. Absolute pitch and lexical tones: Tone perception by non-musician, musician, and absolute pitch non-tonal language speakers. In Hansen, J. and Pellom, B. (eds), *The 7 th International Conference on Spoken Language Processing*. Denver , USA, 257-260.

Chen, Y., Robb, M. P., Gilbert, H. R. & Lerman, J. W. 2001. A study of sentence stress production in Mandarin speakers of American English. *Journal of Acoustical Society of America, 109*(4), 1681-1690.

Fox, A. 2000. *Prosodic features and prosodic structure: The phonology of suprasegmentals*. New York: Oxford University Press.

Gottfried, T. L. and Riester, D. N. 2000. Preliminary results from study of pitch perception and Mandarin tone identification. *Journal of the Acoustical Society of America 108*, 2604.

Hardison, D. M. 2004. Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning & Technology 8*, 34-52.

Ladd, D. R. 1996. *Intonational phonology*. Cambridge: Cambridge University Press.

Post, B. 2002. French tonal structures. In Bel, B. and Marlien, I. (eds), *Proceedings of the Speech*. 583-586.

Smit, A. B. 2004. *Articulation and phonology resource guide for school-age children and adults*. NY: Thomson Delmar Learning

So, C. K. 2006. *Effect of L1 Prosodic Background and AV Training on Learning Mandarin Tones by Speakers of Cantonese, Japanese, and English*, Ph.D. Thesis, Department of Linguistics, Simon Fraser University.

So, C. K., & Best, C. T. 2008. Do English Speakers Assimilate Mandarin Tones to English Prosodic Categories? In *Proceedings of Interspeech 2008 incorporating SST 2008*. Brisbane, Australia.

So, C. K., & Best, C. T. 2010. Cross-Language Perception of Non-native Tonal Contrasts: Effects of Native Phonological and Phonetic Influences. *Language and Speech, 53*(2).

Xu, Y. 1994. Production and perception of coarticulated tones. The *Journal of Acoustical Society of America, 95*(4), 2240-2253.

Xu, Y. 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics, 25*, 61-83.

NOTES

[1] Analysis was performed without “Unknown” responses, which contributed to 3.33% (72 counts) of total responses (2160 counts)

[2] Analysis was performed without “Unknown” responses, which contributed to 7.5% (162 counts) of total responses (2160 counts)

Function words in read and spontaneous speech produced by L2 and L1 speakers

Helena Spilková & Wim A. van Dommelen

Norwegian University of Science and Technology, Trondheim, Norway
helena.spilkova@ntnu.no, wim.van.dommelen@ntnu.no

ABSTRACT

This paper compares the production of the English function words *of* and *to* spoken by speakers of Czech and Norwegian with native productions. Two different speaking styles were investigated, read and spontaneous speech. Acoustic analysis involved word and segment durations and formant measurements. Measurements revealed different reduction patterns in the two function words. In the word *of* non-natives produced longer durations than natives but read productions were longer than spontaneous tokens for natives and non-natives alike. In contrast, word durations for *to* did not differ between speaker groups and speaking styles. Relative segment durations in both words, however, varied between speaker groups. F1-F0 values for the two words did not differ between groups. Higher F3-F2 values in *of* indicated less fronted vowel quality for the non-natives than for the natives. In this case there was no effect of speaking style. The reverse was found for *to,* where no effect of language was observed but vowels from both natives and non-natives were less fronted in spontaneous speech.

Keywords: speaking style, reduction, function words, Czech, English, Norwegian

1. INTRODUCTION

During the initial period of second language (L2) instruction normally attention will be paid to the L2 sound system, in particular to the sounds that are different from the learner's native language (L1). Efforts will be directed towards the acquisition of specific sounds as they appear in canonical word forms. For example, non-native learners of English are known to encounter problems in producing the vowels in English *pat* (/æ/) vs. *pet* (/ɛ/) (e.g., Swan and Smith 2001). Flege et al. (1997) observed inappropriate duration and spectral contrasts in English /i/ vs. /ɪ/ and /æ/ vs. /ɛ/ produced by German, Mandarin, Spanish and Korean speakers. More advanced learners will have to devote efforts to mastering language-specific reduction rules. Depending on factors like the status of a word as content vs. function word, its position in an utterance, speaking style, etc., native speakers' realizations may show varying degrees of reduction. Examples of such processes are reduction of unstressed vowels in modern Greek (Dauer 1980), reduction and enlarged within-category scatter in spontaneous vs. laboratory speech in Spanish (Harmegnies and Poch-Olivé 1992), reduction of consonants and vowels in read vs. spontaneous Dutch (van Son and Pols 1999), Russian (Bolotova 2003), and Japanese (Nakamura et al. 2008). Previous investigations have produced evidence of L2 learners showing phrase-level effects that differ from what is found for L1 speakers. Native speakers of Spanish in Lowenstein Mairs (1989) produced incorrect patterns of stress assignment in English. Further, Wenk (1985) found improper reduction of vowel quality in francophone English. In a study by Bond and Fokes (1985), native speakers of Thai, Malaysian and Japanese were shown to have insufficient awareness of typical English patterns of word compression due to addition of syllable suffixes. In Gut (2007) English learners of German were shown to have insufficient degrees of vowel reduction. The same was true for German learners of English.

The goal of the present study was to investigate reduction phenomena in L2 speech along two dimensions. Firstly, the realization of English function words by L2 speakers was compared with native production. The second dimension involved speaking style, in that for both native and non-native speakers read speech was compared with spontaneous speech (for the description of the latter type of speech, see section 2.1). In addition, non-native speakers came from two typologically different languages, Norwegian and Czech (cf. Swan and Smith 2001). It was hypothesized that in general non-native speakers of English would reduce function words less than natives. Further, native speakers were expected to exhibit stronger reduction effects in spontaneous vs. read speech than L2 users. Finally, it was postulated that the larger typological distance between English and Czech would cause less native-like productions for Czech than for Norwegian speakers.

2. METHOD

2.1. Speech material

The material used in this study was obtained from several sources. The read material is represented by recordings of non-professional speakers reading transcripts of BBC news texts. Part of the BBC news recordings was recorded in Trondheim and part was provided by the Institute of Phonetics, Charles University in Prague. The spontaneous material consists of spontaneous dialogues in English, elicited using a picture replication task (part of the Kachna corpus; Spilková et al. 2010) and dialogues elicited using a Map Task (White et al. 2010). All the recordings were made in studio environments with a sampling rate of 32 kHz or higher and 16-bit quantization, using a separate channel for each speaker (in dialogues).

The lexical items chosen for analysis were two English function words: *of* and *to*. For both types of material, we aimed to select realization of these words fluently and naturally integrated in surrounding speech, therefore we excluded all cases where a pause, hesitation or another type of disfluency was present in close proximity of the observed word. Attention was also paid to the context and syntactic status of the observed words, where we avoided, e.g., clause-final use of prepositions and strongly lexicalised phrases where a disproportinal reduction could be expected. Five tokens of each word per speaker and speaking style were selected (incidentally less than five for a few speakers with a limited number of suitable items).

2.2. Speakers

The groups of subjects consisted of ten Norwegian speakers, ten Czech speakers and two (native) British English speakers that were recorded in both speaking styles (reading and replication task dialogue). In addition, BBC news recordings of three British speakers and recordings of Map Task dialogue of three (other) British speakers were used. The age of the speakers ranged from 19 to 45 years, and most of the speakers were university students. The speaker pairs in the dialogues were in most cases formed by either classmates or colleagues.

In Norway, the well-established system of English instruction and high exposure to English language (e.g. most movies in English are not dubbed) result in an overall high competence in English in young population. The speakers were therefore selected from university students which guaranteed sufficient proficiency. In Czech Republic, however, such a proficiency standard cannot be generally expected and we had to select speakers from more carefully chosen groups, namely university students of English, and employees in a company using English as the official work language.The dialects of the native English speakers mostly belonged to the Southern English dialect group, one speaker spoke a Northern English dialect.

2.3. Acoustic analysis

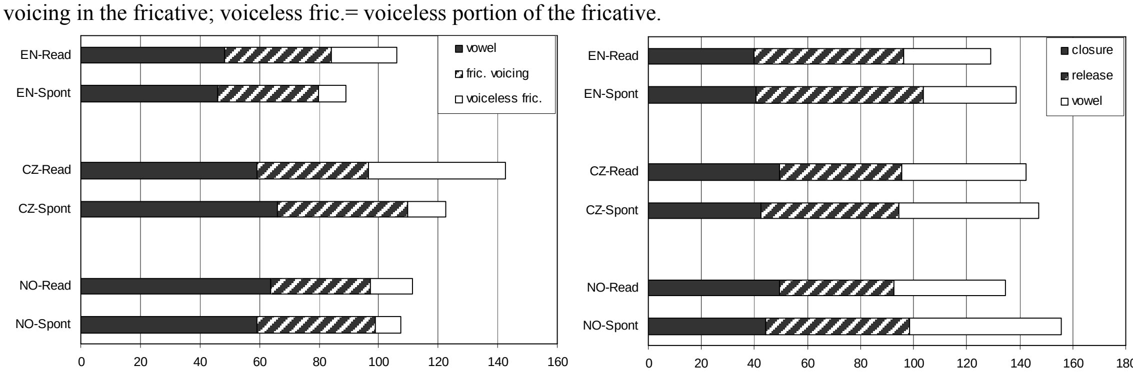
The selected items were segmented using Praat (Boersma and Weenink 2009). Segment durations were obtained for the vowel, the fricative and the voicing in the fricative in *of* and the consonantal closure, release and vowel in *to*. Furthermore, formant values in Bark were measured as means of values obtained from the whole duration of the vowel in the observed item. To be able to eliminate errors in automatic formant tracking, an additional semi-automatic method was used to detect any abrupt jumps between nearby formant measurements. The resulting formant values were used to calculate F1-F0 and F3-F2 values in Bark to reduce the influence of anatomical variation (corresponding to, e.g., gender; cf. Syrdal and Gopal 1986; Adank et al. 2004). The value of F0 necessary for this transformation was measured in the centre of the vowel interval, avoiding the portions with a creaky voice quality where possible.

3. RESULTS

3.1. Temporal organization

In this section we will investigate segmental durations in the function words *of* and *to* spoken by the three groups of speakers. Our measurements showed that generally non-native speakers produced longer durations than natives. Pooled across the two words and speaking styles (read/spontaneous) total word durations were 115ms for the English speakers, 139ms for the Czech and 127ms for the Norwegian speakers. According to an analysis of variance with language, speaking style and word (*of* and *to*) as factors, the effect of language was statistically significant ($F(2, 476)= 6.760$; $p= 0.001$). As could be expected, *of* and *to* had significantly different durations (overall 116ms and 143ms, respectively; $F(1, 476)= 32.713$; $p< 0.001$). Surprisingly, the effect of speaking style turned out to be non-significant ($F< 1$). This result can be explained by the significant interaction found between the factors word and speaking style ($F(1, 476)= 6.704$; $p= 0.010$). While *of* was longer in read speech than in spontaneous speech, the reverse was true for the word *to*. Therefore, the results for the two function words will be dealt with in two separate sections.

Figures 1a , 1b : Mean segment durations (in ms) in the words *of* (on the left) and *to* (on the right) for different language groups (EN= English, CZ= Czech, NO= Norwegian) and speaking styles (Read=read, Spont=spontaneous). fric. voicing= voicing in the fricative; voiceless fric.= voiceless portion of the fricative.



3.1.1. Segment durations and voicing in the function word of

As can be seen from the results for the word *of* presented in Figure 1a, word durations were longer for both Czech (133ms) and Norwegian speakers (109ms) compared to natives (97ms). In addition, durations were longer in read than in spontaneous speech (123ms vs. 109ms). An analysis of variance revealed that whereas the effects of language as well as speaking style were significant ($F(2, 240)= 9.856$; $p< 0.001$ and $F(1, 240)= 4.298$; $p= 0.039$), the language x speaking style interaction did not reach significance ($F< 1$). Bonferroni-adjusted paired comparisons revealed that only Czech word duration differed significantly from English word duration.

Further, V/C ratios broken down for read and spontaneous speech appeared to differ between all three groups of subjects. Relative vowel durations were on average 47% (read) and 52% (spontaneous) for the English speakers, 43% and 55% for the Czech, and 59% and 56% for the Norwegian speakers, respectively. An analysis of variance showed that the effects of the factors language ($F(2, 240)= 8.464$; $p< 0.001$), speaking style ($F(1, 240)= 4.347$; $p= 0.038$) as well as their interaction ($F(1, 240)= 5.860$; $p= 0.030$) were significant.

The data also allowed us to investigate the degree to which the speaker groups produced the phonologically voiced fricative in *of* with phonetic voicing. Statistical analysis revealed that the effect of both language and speaking style on voicing was significant ($F(2, 232)= 3.472$; $p= 0.033$ and $F(1, 232= 18.355$; $p< 0.001$). Pooled across the two speaking styles, Czech and English subjects had similar amounts of voicing (73% and 75%, respectively), while Norwegian speakers made the fricative more voiced (85%). Phonetic classification of immediately neighbouring segments as voiceless/voiced revealed that the amount of voicing in the fricative correlated with the voicing status of the following segment. The distribution of voiceless/voiced contexts across the three languages was similar. Therefore, the effect of language on the amount of fricative voicing was not an artifact caused by context. Further it appeared that speaking style affected the degree to which the fricative was filled with voicing (69% in read speech vs. 88% in spontaneous speech). For the English and, especially, the Czech speakers this result can be explained by relatively long fricative durations in read speech (read vs. spontaneous: Czech 84ms vs. 56ms; English 58ms vs. 43ms). Norwegian speakers had similar fricative durations for these two conditions (both 48ms).

3.1.2. Segment durations in the function word to

Results for the word *to* are shown in Figure 1b. At first sight, it may seem that in contrast to the previous results word durations were shorter for read than for spontaneous speech. Mean values across language groups were 137ms and 149ms for read vs. spontaneous speech. An analysis of variance with the factors language and speaking style revealed, however, that the effect of the latter failed to reach significance ($F(1, 236)= 2.601$; $p= 0.108$). Moreover, word durations for the three speaker groups did not differ significantly from each other (pooled across speaking styles for English: 134ms; Czech: 145ms; Norwegian: 145ms).

Although word duration differences for the three languages did not differ significantly, there might exist differences in relative vowel and consonant durations. This expectation was confirmed by the data, relative vowel durations being 25% for English, 34% for Czech and 32% for Norwegian. In spite of the relatively small differences, statistical analysis showed that these values differed significantly ($F(2, 236)= 8.277$; $p< 0.001$). According to Bonferroni testing, this result was due to both the Czech and Norwegian values being significantly different from the value for English. There was no significant effect of speaking style ($F< 1$).

For all three speaker groups the release of the consonantal closure constituted a considerable part of the total consonant duration (pooled over all conditions 54%). This amount did not vary with speaking style for the English speakers (read: 61%; spontaneous: 62%) and was somewhat smaller in read vs. spontaneous speech condition for the other groups (Czech: 48% and 54%; Norwegian: 47% and 59%, for read and spontaneous, respectively). The influence of both language and speaking style appeared to be significant ($F(2, 236)= 8.621$; $p< 0.001$ and $F(1, 236)= 11.509$, $p= 0.001$) with no significant interaction ($F(2, 236)= 2.492$; $p= 0.085$). Bonferroni testing showed that the effect of language was due to relative release durations being shorter for both groups of L2 speakers (Czech: 51%; Norwegian: 53%; English: 61%).

3.2. Spectral measures

To study the degree of vowel reduction in read vs. spontaneous speech by natives and non-natives the differences F1-F0 and F3-F2 were taken as measures of vowel height and backness. Pooled across the two words and speaking styles F1-F0 values did not differ between the three groups of speakers ($F(2, 433)= 1.283$; $p= 0.278$) but the factor word had a highly significant effect ($F(1, 433)= 134.694$; $p< 0.001$). F3-F2 differed between language groups ($F(2, 466)= 22.745$; $p< 0.001$) and words ($F(1, 466)= 145.70$; $p<0.001$). Therefore, like for segment durations separate analyses were carried out for the words *of* and *to*.

3.2.1. Vowel reduction in the function word of

F1-F0 and F3-F2 values are presented in Figure 2. For F1-F0 an analysis of variance with the factors language and speaking style showed no significant differences at all (language: $F(2, 225)= 2,.136$; $p= 0.121$; speaking style: $F< 1$). Only Norwegian subjects' F1-F0 values were larger for read than spontaneous speech (3.3 Bark vs. 2.9 Bark; $t(92)= 2.168$; $p= 0.033$). In contrast, for the F3-F2 measure a significant effect of language was found ($F(2, 239)= 32.225$; $p< 0.001$). Bonferroni-adjusted paired comparisons showed that F3-F2 values for Czech (4.6 Bark) as well as Norwegian (4.5 Bark) were different from English (3.3 Bark). This indicates more peripheral vowel qualities for the non-native speakers. Both speaking style (read vs. spontaneous: 4.4 Bark vs. 4.3 Bark) and its interaction with the factor language, however, did not reach statistical significance ($F(1, 239)= 1.558$; $p= 0.213$ and $F(1, 239)= 2.347$; $p= 0.098$, respectively).

3.2.2. Vowel reduction in the function word to

Results for F1-F0 values for the word *to* were similar to what was found for the word *of* (see Figure 3). Neither here, the effects of language and speaking style reached statistical significance ($F< 1$ and $F(1, 208)=$

2.710; p= 0.101, respectively). Also for F3-F2 values no significant effect of language was observed (F< 1). There was, however, a reliable effect of speaking style. All three speaker groups had somewhat lower values for read than for spontaneous speech (averaged across language groups 3.0 Bark vs. 3.2 Bark; F(1, 227)= 8.300; p= 0.004). Although the difference was larger for English (2.7 Bark vs. 3.3 Bark) than for Czech and Norwegian subjects (for both 3.1 Bark vs. 3.2 Bark), the language x speaking style interaction was non-significant (F(1, 227)= 1.602; p= 0.204).

Figure 2: Mean Bark distances (F1-F0, F3-F2)in the function word *of* for different language groups (EN=English, CZ=Czech, NO=Norwegian) and speaking styles (Read=read, Spont=spontaneous).

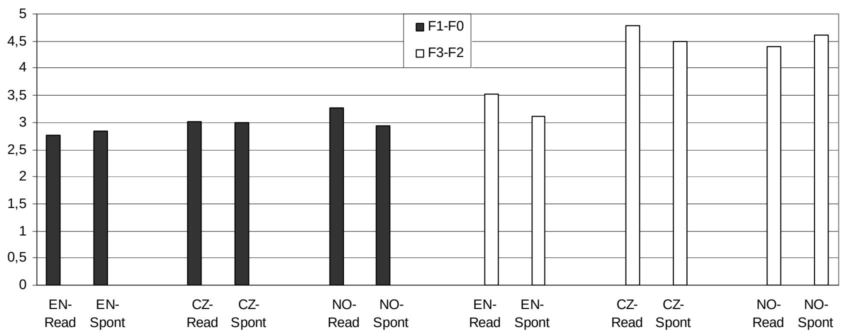
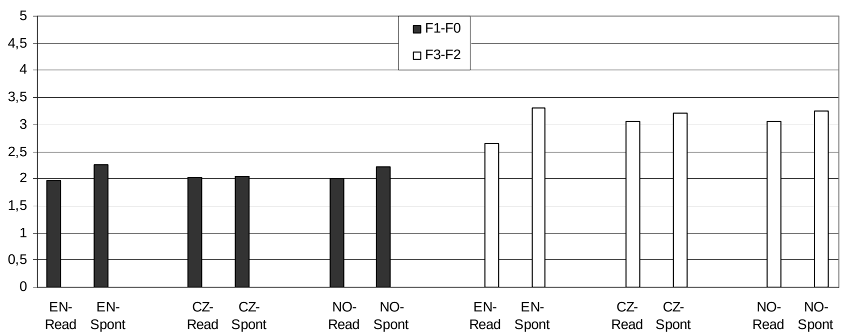


Figure 3: Mean Bark distances (F1-F0, F3-F2)in the function word *to* for different language groups (EN=English, CZ=Czech, NO=Norwegian) and speaking styles (Read=read, Spont=spontaneous).



4. SUMMARY AND DISCUSSION

The results of the present study only partly confirmed the experimental hypotheses formulated at the outset. In general, effects in the temporal domain were stronger than in the spectral domain. For the two function words investigated (*of*, *to*) different reduction patterns were observed. Non-natives produced longer durations in the word *of* than natives. In congruence with the postulated influence of typological distance, this effect was stronger for the speakers of Czech compared to Norwegian. Read productions were longer than spontaneous tokens for natives and non-natives alike. Further, internal syllable structure (i.e., V/C ratio) differed for natives vs. Norwegian but not Czech speakers. Only the Norwegians had a larger percentage of voicing in the fricative which was explained by their relatively short fricative durations. In contrast to what was found for *of*, word durations for *to* did not differ between speaker groups and speaking styles. V/C ratio, however, was higher and the proportion of the release within the plosive was lower for non-natives than natives. As to the spectral measures, no significant effects of language and speaking style were found for the F1-F0 measure. As was witnessed by higher F3-F2 values in the preposition *of* for the non-natives, these L2 groups had less fronted vowels than the English speakers. For the word *to* no such effects were found. In this case, however, lower F3-F2 values were observed in read compared to spontaneous tokens.

The present observation of longer non-native *of* durations is in line with the findings by Gut (2007). In her investigation longer syllable durations were found in non-native vs. native English and German. In contrast to the present results, she did not find significant effects of speaking styles (free speech, story retelling, and reading) for native English and native as well as non-native German. Only her non-native speakers of English showed stronger reduction in free speech compared to reading and story retellings. While the absence of a speaking style effect on *to* durations again is in congruence with Gut (2007), the present absence of a native/non-native difference for the word *to* is at odds with her results. The lack of consistent results, also as to spectral vowel quality, might at least partly be explained by individually different reduction strategies. For example, in Dauer's 1980 study idiosyncratic factors outweighed differences in reading vs. spontaneous speech. Speakers in Laan (1997) had different approaches to speaking styles. Similarly, large between-speaker variability was observed for speakers of Russian, Finnish and Dutch in de Silva et al. (2003). Only the speakers of Russian had consistently shorter speech sound durations in read vs. spontaneous speech. Also, in Lavoie (2002) only two out of five speakers had reliably longer rime durations in *for* following strong vs. weak syllables. In conclusion, all these results suggest that in second language learning idiosyncratic factors are highly relevant and that in teaching particular attention has to be paid to individual behaviour.

5. ACKNOWLEDGEMENTS

The authors were supported by the EU Marie Curie Research Training Network *Sound to Sense*. We would like to thank Radek Skarnitzl for providing the read Czech material and to Laurence White for making available the recordings of native English speakers' Map Task.

6. REFERENCES

Adank, P., Smits, R., van Hout, R. 2004. A comparison of vowel normalization procedures for language variation research. *Journal of the Acoustical Society of America* 116. 3099-3107.

Boersma, P., Weenink, D. 2009. Praat: Doing phonetics by computer (Version 5.1.06) [Computer program]. Retrieved May 11, 2009, from http://www.praat.org/.

Bolotova, O. 2003. On some acoustic features of spontaneous speech and reading in Russian (quantitative and qualitative comparison methods). *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, 3-9 August 2003 (1). 913-916.

Bond, Z.S., Fokes, J. 1985. Non-native patterns of English syllable timing. *Journal of Phonetics* 13. 407-420.

Dauer, R.M. 1980. The reduction of unstressed high vowel in modern Greek. *Journal of the International Phonetic Association* 10(1). 17-27.

de Silva, V., Iivonen, A., Bondarko, L.V., Pols, L.C.W. 2003. Common and language dependent phonetic differences between read and spontaneous speech in Russian, Finnish and Dutch. *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, 3-9 August 2003 (3). 2977-2980.

Flege, J.E., Bohn, O.-S., Jang, S. 1997. Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics* 25. 437-470.

Gut, U. 2007. Learner corpora in second language prosody research. In: Trouvain, J., Gut, U. (eds.), *Non-native prosody: Phonetic description and teaching practice.* Berlin/New York: Mouton de Gruyter. 145-167.

Harmegnies, B., Poch-Olivé, D. 1992. A study of style-induced vowel variability: Laboratory versus spontaneous speech in Spanish. *Speech Communication* 11. 429-437.

Laan, G.P.M. 1997. The contribution of intonation, segmental durations, and spectral features to the perception of a spontaneous and a read speaking style. *Speech Communication* 22. 43-65.

Lavoie, L. 2002. Some influences on the realization of *for* and *four* in American English. *Journal of the International Phonetic Association* 32(2). 175-202.

Lowenstein Mairs, J. 1989. Stress assignment in interlanguage phonology: an analysis of the stress system of Spanish speakers learning English. In: Gass, S.M., Schachter, J. (eds.), *Linguistic perspectives on second language acquisition.* Cambridge: Cambridge University Press.260-283.

Nakamura, M., Iwano, K., Furui, S. 2008. Differences between acoustic characteristics of spontaneous and read speech and their effects on speech recognition performance. *Computer Speech and Language* 22. 171-184.

Spilková, H., Brenner, D., Öttl, A., Vondřička, P., van Dommelen, W.A., Ernestus, M. 2010. The Kachna L1/L2 Picture Replication Corpus. *International Conference on Language Resources and Evaluation, 2010*, Malta. (to appear)

Syrdal, A.K., Gopal, H.S. 1986. A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America* 79. 1086-1100.

Swan, M., Smith, B. (eds.). 2001. *Learner English: a teacher's guide to interference and other problems.* (2nd ed.) Cambridge: Cambridge University Press.

van Son, R.J.J.H., Pols, L.C.W. 1999. Acoustic description of consonant reduction. *Speech Communication* 28. 125-140.

Wenk, B.J. 1985. Speech rhythms in second language acquisition. *Language and Speech* 28(2). 157-175.

White, L., Wiget, L., Rauch, O., Mattys, S.L. 2010. Segmentation cues in spontaneous and read speech. *Proceedings of the 5th Conference on Speech Prosody*, 2010, Chicago. (to appear)

The L2 Acquisition of Japanese Length Contrasts

Fumiko Shimizu Summerell

University of Calgary
summerel@ucalgary.ca

ABSTRACT

The acquisition of vocalic and consonantal length contrasts in Japanese is one of the problematic areas of phonological acquisition for native English speakers (Enomoto 1992; Han 1992; Toda 1997, 2003). This study seeks to address the following two questions: (1) whether or not native English speakers can acquire Japanese length contrasts; and (2) what property of the grammar is responsible for this success or failure. Fifty-six subjects (19 beginners, 13 intermediate, 12 advanced, 12 native Japanese speakers) performed an auditory discrimination task and a picture identification task. The results showed that all groups performed at near ceiling capacity in discriminating two sounds acoustically in the auditory task. In the picture task, however, significant differences were found between the beginners and native Japanese speakers for single versus geminate consonant contrasts and long vowel versus geminate consonant contrasts. The results seem to support the hypothesis proposed by Mah and Archibald (2003) and Archibald (2004) that English speakers can acquire Japanese length contrasts by redeploying the mora licensing properties in their L1 grammar.

Keywords: mora licensing, geminate

1. INTRODUCTION

1.1. The problem

One of the main issues in the study of second language (L2) is the acquisition of L2 phonology. Acquiring the sound system of the target language involves learning the segmental inventory, syllable structure, and suprasegmental features such as stress, rhythm, and intonation of the language, which may differ from the L1, the learner's first language (Archibald 1995; Major 2001; Riney and Flege 1998). Various researchers have reported difficulties for native English speakers (L1 English) in acquiring Japanese vocalic and consonantal length contrasts (Enomoto 1992; Han 1992; Toda 1997, 2003).

1.2. Research Questions

In this research, native English speakers' perception of non-native contrasts, namely short and long contrasts in both vowels and consonants, will be investigated in an attempt to address the following two questions: (1) whether or not native English speakers can acquire Japanese length contrasts; and (2) what property of the grammar is responsible for this success or failure.

The main aim of this study is to investigate the acquisition of novel contrasts by native English speakers, so the focus is on examining English and Japanese phonemes and syllable structures. Analyses of Japanese pitch and intonation patterns are, therefore, limited in detail. In addition, the language examined is restricted to modern standard Japanese, often referred to as the Tokyo dialect.

2. BACKGROUND

2.1. Phonemic inventory of Japanese and English

Japanese has durational contrasts in both vowels and consonants and these differences are phonemically distinctive. For vocalic length contrast, for example, *i* with a short vowel means 'stomach' and *ii* with a long vowel means 'good'. English has a length contrast between lax vowels, which are monomoraic, and tense vowels, which are bimoraic. In English while the lax vowel in a word, for example [ɪ] in *tin,* is monomoraic, the tense vowel [i] in *teen* is bimoraic (Hayes 1989). English also has a vowel quality contrast in lax vowels (e.g., *tin*, [ɪ]) versus tense vowels (e.g., *teen*, [i]), whereas Japanese does not have a quality contrast. In short while English has both quality and quantity contrasts in vowels, Japanese only has quantity contrasts.

For consonantal contrasts, Japanese has phonemic differences between single and double (or "geminate") consonants. For example, *kite* 'wear' with a single consonant contrasts with *kitte* 'stamp' with a geminate, and *saka* 'slope' and *sakka* 'writer' also contrast. In English there is no phonemic consonantal contrast based on length. Although English has a phenomenon referred to as "ambisyllabicity" (as in *pony* and *black cat*), in which the intervocalic consonants belong to two syllables, these consonants are not referred to as geminates as there are no minimal pairs (Kahn 1976).

2.2. Theoretical models of L2 phonological acquisition

In this section, the Speech Learning Model (SLM) and the Feature Geometry Model (FGM) are discussed. Both theories are used to examine the L2 phonological acquisition in relation to the phonetics or phonology of the learner's L1, and are employed in this research.

2.2.1. Speech Learning Model (SLM): Flege 1992, 1995

Flege's Speech Learning Model (SLM) is designed to account for differences in segmental perception by native speakers and nonnative speakers of a language. Flege (1995) argues that the cause of L2 perception difficulties for adult L2 learners is the incorrect mapping of L2 phones onto the existing L1 categories.

The SLM distinguishes three types of L2 phones, namely "same', "new", and "similar". *Same* phones are assumed to be identical with L1 phones and are easily acquired, as in English and German /m/ (Bohn and Flege 1990). *New* phones are not identified with any L1 phones and allow the learner to establish a new category that will eventually be acquired; an example would be the acquisition of English /r/ by native German speakers (Bohn and Flege 1990). *Similar* phones are not exactly the same as L1, but L2 learners do not perceive the difference. The formation of a new category is likely to "be blocked by the mechanism of equivalence classification" (Flege 1995: 239). That is, although L2 learners may detect a particular phone as being not exactly the same as an L1 phone, they would categorize it as being the same as the L1 phone. It is, therefore, problematic for the learner to acquire this contrast. For example, the English liquids /r/ and /l/ would likely be identified as "similar" and are mapped onto the same category /ɾ/ by L1 Japanese, since Japanese has one liquid (Flege et al. 1996).

McAllister et al. (2002) investigated the acquisition of the Swedish vocalic length contrasts by native Estonian (L1 Estonian), Spanish (L1 Spanish), and English speakers. Swedish and Estonian have durational contrasts in both vowels and consonants, but durational distinctions are not phonemically relevant in Spanish and English (McAllister et al. 2002). The results revealed that L1 Estonian subjects obtained significantly higher perceptual accuracy rates than L1 Spanish and L1 English subjects ($p<0.05$). They argued that L1 Estonian were aided by the presence of durational distinctions in their L1 and were able to acquire the Swedish durational contrasts. The study also noted that the mean accuracy rate for L1 English was significantly higher than that of L1 Spanish ($p<0.05$) for the mid vowels, which utilize durational contrasts. They explained that since English vowels have quantity contrast, L1 English were able to use the durational cues while Spanish vowels have quality contrasts only L1 Spanish were unable to do so. In short, their findings indicated that the phonetic structure of the L1 would influence the L2 speech acquisition. Therefore, if there were a mismatch to signal contrast in the L1 and L2, the formation of a new L2 would be blocked. On the other hand, learning would be facilitated through reattunement of the existing L1 phonetic structure if the contrast were present in the L1.

To summarize, the SLM hypothesizes the main reason for an L2 learner's difficulty to acquire certain contrasts is the incorrect mapping of L2 phones onto existing L1 inventory. The property of grammar responsible for the difficulty is the L2 learners' inability to establish a new phonetic category in long-term memory.

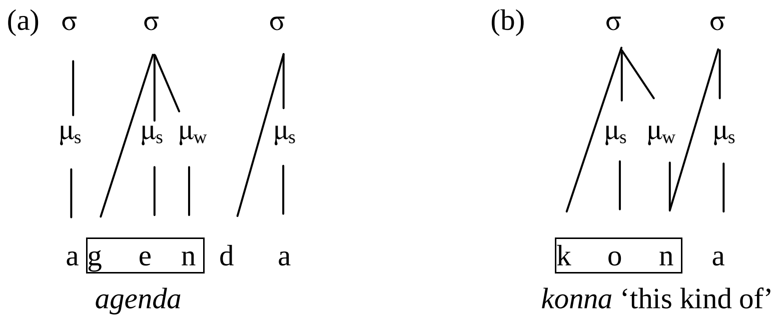
2.2.2. Feature Geometry Model (FGM): Brown 1997, 2000

Brown's Feature Geometry Model (FGM) is based on a generative phonology approach, which views a phoneme as an abstract entity built out of features. Brown argues that it is the L1 feature geometry that determines the L2 learner's perception, not the segments.

According to Brown, if the relevant feature is available, then the native grammar will facilitate perception, whereas it will block the filtering process if the feature is absent in the learner's L1. Brown (1997) investigated the acquisition of the English lateral approximant, /l/, versus the central approximant, /r/, by native Japanese and Mandarin Chinese (L1 Chinese) learners of English. Although neither Japanese nor Chinese have the /l/ and /r/ distinction, her study demonstrated that L1 Japanese were unable to discriminate the /l/ and /r/ contrast phonetically and phonologically, but L1 Chinese were able to do so with very high accuracy rates. A possible explanation for the difference of the performance between L1 Chinese and L1 Japanese, according to Brown, is the different feature inventories in their L1s. Since Chinese has a phonological feature responsible for these two contrasts (i.e., the feature [coronal] to distinguish the alveolar /s/ from the retroflex /ʂ/), Chinese learners of English were able to establish a new phonological category to build a new representation for the contrast. On the other hand, Japanese lacks the feature separating the contrast; therefore, the perception of /l/ and /r/ is mapped onto one phonemic category and the acquisition of the non-native segment would not be successful. Brown argues that the distinctive feature in the L1 is the crucial factor in the acquisition of non-native phonemic contrasts, rather than the L1 phonemic inventory.

Mah and Archibald (2003) and Archibald (2004) extended Brown's hypothesis, which operates at the featural level, to the moraic level. They reported a study on the production ability of Japanese vocalic and consonantal length contrasts by a single L1 English subject. The subject was able to produce the long vowels and consonants that were significantly longer than their short counterparts. Following Zec's (1995) claim, in which she argued that the mora projected by a vowel is strong (μ_s) while the mora projected by a consonant is weak (μ_w) as illustrated in Figure 1, Mah and Archibald (2003) and Archibald (2004) made the following proposal: Since coda consonants are licensed by a weak mora in English and geminate consonants are licensed by a weak mora in Japanese, L1 English can build new phonological representations by the redeployment of this weak mora licensing property from their L1. According to them, English speakers will be able to utilize this existing mora licensing properties to acquire vocalic and consonantal length contrasts in Japanese.

Figure 1: Moraic representation of English word *agenda* and Japanese word *konna* 'this type of'



To summarize, according to the FGM, it is the redeployment of existing phonological structure that plays a crucial role in the acquisition of L2 phonology and affects L2 perception. Contrary to the SLM, which focuses on phonetic discerning ability, FGM hypothesizes that the learner cannot add a new category to his or her inventory.

2.3. The present study

The present study seeks to expand upon Flege (1992, 1995) and Brown's (1997, 2000) studies by investigating the acquisition of durational contrasts. As noted above, Flege operates at the segmental level and Brown operates at the featural level, and this study examines the contrasts at the prosodic level.

On the basis of what I have discussed above, three hypotheses for my present perception study can be summarized as follows:

Hypothesis 1: L1 English will be unable to acquire either vocalic or consonantal contrasts. L1 English will perceive both short and long vowels and single and geminate consonants as "similar", since there are no contrasts in their L1.

Hypothesis 2: L1 English will be able to acquire vocalic length contrasts, since English vowels have quantity contrasts. However, since English lacks consonantal length contrasts, L1 English will perceive both single and geminate consonants as "similar", and the formation of a "new" category will likely be hindered.

Hypothesis 3: L1 English will be able to acquire both vocalic and consonantal contrasts by redeploying the mora licensing properties from their L1 grammar.

Table 1 illustrates the predictions of my study on the basis of the above hypotheses.

Table 1: Predictions

Hypothesis	Model	proposed by	vowel	consonant
H 1	SLM-Flege	Flege	No	No
H 2	SLM-McAllister et al.	Flege/McAllister et al.	Yes	No
H 3	Mora Licensing	Mah & Archibald	Yes	Yes

3. EXPERIMENTS

In this section, the present study is summarized. It begins with an overview of the subjects and method, followed by the results and discussion of each experiment.

3.1. Subjects

Fifty six adults participated in this study. Subjects were divided into four groups according to their proficiency levels in Japanese: Beginners (n=19), Intermediate (n=13), Advanced (n=12) and Control (n=12). All reported having normal hearing. The subjects in the Beginners' group were all university or college students who were in the first year of a Japanese program. Length of learning Japanese was less than 150 hours. The Intermediate group consisted of students who were in the second, third or fourth years of a Japanese program at the university level. Length of learning Japanese varied from 160 hours to 445 hours. All of Advanced subjects had lived in Japan for one to ten years and spoke fluent Japanese. The Control group included twelve native Japanese speakers.

3.2. Method

The subjects performed the following three tasks: (1) an AXB auditory discrimination task (Task 1); (2) a lexical knowledge task (Task 2); and (3) a picture identification task (Task 3).

The purpose of Task 1, an AXB auditory discrimination task. was to investigate whether or not native English speakers were able to distinguish acoustic signals of natural stimuli that are not phonemic in their L1; therefore, stimuli recorded by a female standard Japanese speaker were used instead of synthesized tokens. A 250 millisecond (msec) inter-stimulus interval (ISI) was inserted between words (i.e., word A + 250 msec ISI + word X + 250 msec ISI + word B) in order to cause perception to operate at the auditory level instead of the phonemic level (Werker and Logan 1985). Subjects listened to 144 triads over the headphones, and responded if the stimulus in the middle (X) is the same as the fist one (A) or the last one (B) by pressing the A and B keys respectively on the computer keyboard.

Task 2, a lexical knowledge task, was designed to confirm that the lexical items were stored in each learner's long-term memory, in preparation for Task 3. Before the experiment, the subjects were given a list of 39 words, which consisted of 30 test tokens and 9 foils, and were asked to familiarize themselves with the words. Subjects were presented with a picture and four choices of words on the computer screen and were asked to choose the correct lexicon for the picture by pressing A, B, C or D on the keyboard.

Task 3, a picture identification task, investigated whether the learners were able to tap the aural information to access to lexical knowledge. The test words for Task 2 were also used in Task 3. Since subjects who scored 100% accuracy rate proceeded to Task 3, any auditory errors would have been due to incorrect perception, but not the lack of lexical knowledge. Two drawings were paired and appeared on the computer screen while the aural stimuli were presented through the headphone. The aural stimulus was placed in the following sentence frame in order to avoid the list effect: *Watashi-wa ima__ to iimashita*. 'I said __ just now.' After hearing the phrase that included a test token, a subject would press the key, A or B, which he or she believed to correspond to the correct token.

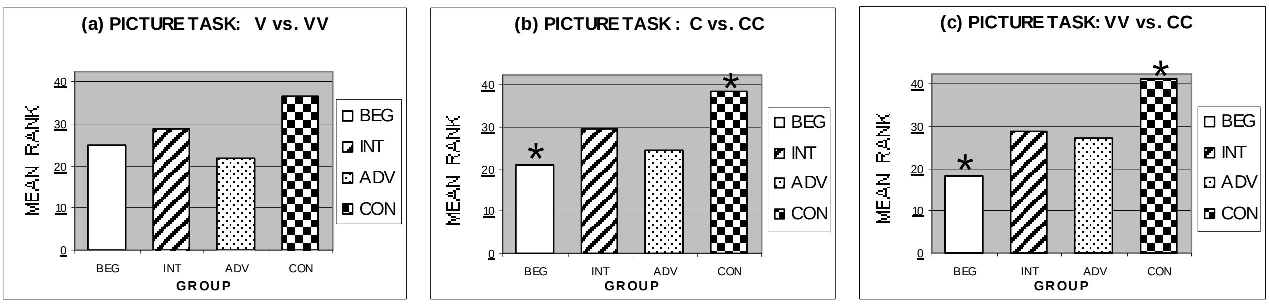
3.3. Results

The results of Task 1 indicated that both L1 English and L1 Japanese were able to discriminate vocalic and consonantal contrasts equally well. All groups had higher accuracy rates of over 90%. No statistical differences were found between L1 English or L1 Japanese, and among the four groups.

As mentioned, the score of 100% accuracy rate was needed for Task 2 to proceed to Task 3. Only one subject in the Beginners' group was unable to reach the perfect score and discontinued the study.

Figure 2 illustrates the results of Task 3. As seen in Figure 2 (a), the vocalic length contrasts (V vs. VV) the mean rank of the Control group (CON) is much higher than the L1 English groups; however, the difference revealed no statistical significance ($p>0.05$). Significant differences were found for the consonantal contrasts (C vs. CC) and long vowel and geminate consonant contrasts (VV vs. CC) between the Beginners (BEG) group and Control group as indicated with an asterisk in the graphs in Figure 2(b) and (c) respectively. Further analysis indicated that for C vs. CC, the results for the stop contrasts (/p, t, k/ vs. /pp, tt, kk/) and the fricative contrasts (/s, ʃ/ vs. /ss, ʃʃ/) were significant ($p<0.05$) but not nasals (/m, n, ŋ/) vs. /mm, nn, ŋŋ/). And for VV vs. CC, only the nasals (VV vs. /mm, nn, ŋŋ/) showed significance ($p<0.05$), not for stops (VV vs. /pp, tt, kk/) and fricatives (VV vs. /ss, ʃʃ/). The results indicated that while the Intermediate (INT) and Advanced (ADV) groups were able to discriminate these contrasts, the Beginners group scored considerably low accuracy rates. In short, the results show that L1 English are able to acquire Japanese length contrasts.

Figure 2: Picture identification task (Task 3) by Beginners (BEG), Intermediate (INT), and Advanced (ADV) and Control (CON) groups



Although the data revealed no significance ($p>0.05$), it is clear from the graphs in Figure 2 that the Control group has much higher accuracy rate than the Intermediate and Advanced groups for discriminating the length contrasts. The errors made by L1 English might be explained by the work of Mah and Archibald (2003) and Archibald (2004) as follows: While the learners in the Intermediate and Advance groups were able to redeploy the mora licensing properties that are present in their L1 to discriminate the length contrasts, the Beginners group was unable to manipulate the properties.

4. CONCLUSION

The purpose of the present study was to investigate whether L1 English were able to discriminate Japanese length contrasts. The results show that L1 English are able to acquire Japanese length contrasts. The present study also aimed to investigate the property of the grammar responsible for the acquisition of these contrasts. The results seem to support Hypothesis 3 proposed by Mah and Archibald (2003) and Archibald (2004). L1 English may be redeploying the mora licensing properties that are present in their L1 to discriminate Japanese length contrasts.

Although the scope of the study was limited in number of subjects and test tokens, the experiments reported in this research open up many avenues for future research. For example, the testing of L2 learners whose L1s utilize length contrasts, such as Swedish and Estonian, and of other learners whose L1s do not utilize length contrasts, such as Spanish and French, may confirm or contradict the mora licensing hypothesis. In addition, collection of data from larger sample and various stimuli would provide valuable information in this field of study. And finally, an investigation of the subjects' pronunciation abilities in addition to their perception may provide some clues for the relationship between perception and production.

5. REFERENCES

Archibald, J. 1995. The acquisition of English stress by speakers of non-accentual languages: lexical storage versus computation of stress. *Linguistics*. 35 (1). 167-181.

Archibald, J. 2004. Interfaces in the prosodic hierarchy: New structures and the phonological parser. *International Journal of Bilingualism*. 8 (1). 29-50.

Bohn, O-S., Flege, J. E. 1990. Interlingual identification and the role of foreign language in L2 vowel perception. *Applied Psycholinguistics*. 11. 303-323.

Brown C. 1997. *The Acquisition of Segmental Structure: Consequences for Speech Perception and Second Language Acquisition*. Doctorial Dissertation. McGill University. Montreal Quebec.

Brown C. 2000. Interrelation between Speech Perception and Phonological Acquisition and Phonological Acquisition from Infant to Adult. In: Archibald, J. (ed.). *Second Language Acquisition and Linguistic Theory*. Malden MA: Blackwell. 4-63.

Enomoto, K. 1992. Interlanguage phonology: the perceptual development of durational contrasts by English-speaking learners of Japanese. *Edinburgh Working Papers in Applied Linguistics*. 3. 25-36.

Flege J.E. 1992. Speech learning in a second language. In: Ferguson, C.A., Menn, L., Stoel-Gammon, C. (eds.) *Phonological development: models research implications*. Timonium MD: York. 565-604.

Flege J. E. 1995. Second language speech learning: theory findings and problems. In: Strange W. (ed.). *Speech perception and linguistic experience: Issues in Cross-language Research*. Baltimore: York Press. 233-277.

Flege, J. E., Takagi, N., Mann, V. 1996. Lexical familiarity and English-language experience affect Japanese adults' perception of /ɹ/ and /l/. *Journal of the Acoustical Society of America*. 99. 1161-1173.

Han, M. S. 1992. The timing control of geminate and single stop consonants in Japanese: a challenge for nonnative speakers. *Phonetica*. 49. 102-127.

Hayes, B. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry*. 20. 253-306.

Kahn, D. 1976. *Syllable-Based Generalizations in English Phonology*. Indiana University Linguistics Club: Bloomington, IN.

McAllister R., Flege J., Piske, P. 2002. The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics*. 30. 229-258.

Mah, J., Archibald J. 2003. Acquisition of L2 length contrasts: *Proceedings of the 6th Generative Approaches to Second Language Acquisition Conference (GASLA 2002)*. In: Liceras, J. M., Zobl, H., Goodluck, H. (eds.). Somerville, MA: Cascadilla Proceedings Project. 208-212.

Major, R. C. 2001. *Foreign accent: the ontogeny and phylogeny of second language phonology*. Nahwah, NJ: L. Erlbaum Associates.

Riney, T. J., Flege, J. E. 1998. Change over time in global foreign accent and liquid identifiability and accuracy. *Studies in Second Language Acquisition*. 20. 213-243.

Toda, T. 1997. Strategies for producing mora timing by non-native speakers of Japanese. *Acquisition of Japanese as a Second Language* 1. Tokyo: Bonjinsha. 57-96.

Toda, T. 2003. *Second Language Speech Perception and Production: Acquisition of Phonological Contrasts in Japanese*. Lanham, MD: University Press of America.

Werker, J., Logan, J. 1985. Cross-language evidence for three factors in speech perception. *Perception and Psychophysics*. 3 (7). 35-44.

Zec, D. 1995. Sonority constraints on syllable structure. *Phonology*. 12. 85-130.

Phonetically difficult words in intermediate learners' English

Jolanta Szpyra-Kozłowska

Department of English, Maria Curie-Skłodowska University, Lublin, Poland
jolanta.szpyra-kozlowska@poczta.umcs.lublin.pl

ABSTRACT

Foreign-accented English abounds in numerous phonetically incorrect words which are stored in learners' memory with phonologically deviant representations, e.g. *Disney* is often pronounced in Polish English as [ˈdʲisnej]. In a recent experimental study (Szpyra-Kozłowska, in press) I have demonstrated that such mispronounced items are highly detrimental to successful communication. Consequently, they deserve to be thoroughly investigated and pedagogically prioritized.

This paper examines the subjective evaluation of phonetically difficult words by intermediate Polish learners of English, in whose speech they are particularly numerous. It attempts to analyse the major sources of word pronunciation errors and identify several others which have so far escaped notice of pronunciation specialists. This is done on the basis of an experiment carried out by the author in which 80 secondary school pupils, all intermediate learners of English, were asked to list those words whose pronunciation they found particularly difficult and to provide comments on the problematic aspects of these items. The obtained data are presented, classified and examined. The analysis hopes to offer some new insights into word pronunciation in the Polish English of intermediate learners and carry useful classroom implications.

Keywords: Phonetically difficult words, intermediate learners' English, Polish English interlanguage.

1. INTRODUCTION

One of the striking features of foreign learners' English is a high number of word mispronunciations of different types. Phonetically difficult lexical items are not only those which contain sounds either absent or different from the L1 inventory (leading to the so-called global errors), but also those which tend to be idiosyncratically stored in learners' memory with incorrect phonological representations and stem from various interference factors (resulting in local errors). As argued by Szpyra-Kozłowska (in press), phonetic instruction should focus on local errors because of serious consequences they have for intelligibility, comprehensibility, accentedness and acceptablility of non-native speakers' English. Proper understanding of the mechanisms that lie behind such serious errors and which contribute to the phonetic difficulty of words is therefore crucial for effective pronunciation training of foreign learners.

In spite of this, the issue of phonetically difficult words is generally ignored in phonetic manuals and pronunciation practice materials, probably due to the fact that the majority of them are addressed to international users and are not concerned with errors which are L1-specific. In this context Sobkowiak's (1999) work on the Phonetic Difficulty Index should be pointed out as a valuable attempt to deal systematically with phonetically problematic words in Polish English. Difficulty ratings of English words were carried out by him on the basis of direct observations and thus represent a teacher's perspective. Our study, which deals with the learners' perception of phonetic difficulty, can be seen as both complementing and verifying Sobkowiak's approach.

In this paper we examine about 250 words collected by means of an experiment and subjectively evaluated as phonetically difficult by intermediate Polish learners of English in order to shed more light on this important aspect of the Polish English interlanguage and to draw some pedagogical implications.

2. EXPERIMENTAL DESIGN

In November 2009, 80 teenagers, aged between 14 and 18, of both sex, representing an intermediate level of proficiency and attending a private secondary (junior and senior) school in Lublin, were asked by their English teacher, during regular language lessons, to note down anonymously several English words whose correct pronunciation they found particularly difficult to remember and which they tended to mispronounce. They were also requested to point to the problematic aspects of these items. The applied procedure yielded about 350 different words. A given item was analysed only when it was found in several pupils' responses. Over 250 such cases are discussed in this paper.[1]

3. RESULTS AND DISCUSSION

In this section we present and discuss the obtained data The collected items are grouped into several categories starting with the most numerous ones and proceeding to the less common types.

3.1. Spelling-based forms

The major culprit responsible for learners' phonetic difficulties is English spelling or, to be more exact, irregularities found among spelling-to-pronunciation rules as well as differences between such correspondences in Polish and English. As interference of English orthography on pronunciation is a well-known issue (see Sobkowiak 1996), in what follows I focus on those cases only which were particularly numerous in the participants' answers.

The majority of difficult words provided by the pupils share multiple phonetic realizations of a given letter or a sequence of letters. A typical comment on such items was as follows: "*I never know whether to pronounce* [s] *or* [z] *in this word.*"

(1) <s> (a) *base*, *basic*, *basis*, *isolate*, *isolation*, *crisis*, *fantasy*, *ecstasy*, *bison*, *philosophy*

(b) *comparison*, *curiosity*, *generosity*, *consist*, *insist*, *increasing*, *releasing*, *inclusive*, *exclusive*

What the items in (1a) and (1b) have in common is that here the letter <s> tends to be pronounced by learners as a voiced fricative though for different reasons; in (1a) because of the presence of [z] in similar Polish words (e.g. *ba*[z]*a*, *kry*[z]*ys*), in (1b) because of some kind of s-voicing rule (operating mostly intervocalically and next to sonorants) that Poles tend to develop. Interestingly, this is not a case of interference from Polish, which allows for both [s] and [z] in such contexts, e.g. *o*[s]*a* 'wasp.'

<c> and <cc>

The pronunciation of the letter <c> or a sequence of two such letters was also problematic for the pupils.

(2) <c> *civil*, *scene*, *cement*, *cycling* <cc> *success*, *accent*, *accelerate*, *accident*

The source of difficulty lies here in the presence of the voiceless dental affricate [ts] in the corresponding Polish words in the first set, e.g. [ts]*ywil*, and the cluster [kts] in the second set, e.g. *su*[kts]*es*, and this type of pronunciation is carried over to the English words under discussion.

<ch>

This digraph was often underlined by the subjects as difficult to pronounce in the following words:

(3) *technique, technology*, *techno (music), chaos, choir*, *echo*, *orchid*

where it tends to be pronounced as the voiceless velar fricative, as in their Polish equivalents.

<ous>, <able>, <ate>, <ace>, <ough>, <augh>

Some suffixes and sequences of letters are also known to cause pronunciation difficulties.

(4) <ous> *dangerous, famous, jealous, nervous, marvelous, continuous*

This suffix is frequently rendered as [ɔws] or [us] in Polish English.

A common problem concerns the pronunciation of <a> in <able>, <ate> and <ace>, often realized by Polish learners as [ej] with a frequent additional side effect of vowel mispronunciation in the form of the incorrect stress placement on such sequences. The following difficult items were provided: :

(5) <able> *capable*, *available*, *valuable*, *comfortable*, *vegetable*

<ate> *delicate*, *certificate*, *ultimate*, *separate* (adj.)

<ace> *surface*, *preface*, *necklace*, *palace*

Two more sequences of letters with multiple phonetic realizations were considered problematic:

(6) <ough> *tough*, *enough*, *brought*, *fought*, *thought*, *although*, *through*, *cough*
<augh> *taught*, *caught*, *laugh*, *draught*

3.2. Phonetic 'false friends'

The second largest group of problematic words contains those lexical items which appear in both languages in an identical or very similar orthographic form. While the majority of them are cognates, this is not always the case as in many instances such items are completely unrelated, as shown in the translations of the Polish forms, e.g. E *ten* – P *ten* 'this, masc.,' E *pan* – P *pan* 'mister,'E *gnat* – P *gnat* 'bone, aug.' Nevertheless, cognates prevail in this group and lead to a strong tendency to be pronounced by learners the way they are in Polish. The items in (7), found in the collected data ,belong to this category:

(7) *taxi*, *karate*, *alibi*, *album*, *model*, *panel*, *atom*, *tandem*, *safari*, *mania*, *horror*, *agent*,
boa, *baobab*, *jaguar*, *giraffe*, *panda*, *contact*, *robot*, *stereo*, *video*, *poster*, *flash*, *slogan*

Some of the participants commented on such cases in the following way: "*They are like Polish words so I pronounce them in the Polish way.*"

A large subgroup of such items comprises proper nouns of various types. They include people's names,

(8) *Daniel*, *Audrey*, *Sara*, *Naomi*, *Adrian*, *Howard*, *Penelope*, *Murphy*, *Turner*,
Ryan, *Brittney Spears*, *Rourke*, *Einstein*, *Graham*, *Spielberg*, *Shakespeare*

as well as geographical terms, e.g.

(9) *Japan*, *Nepal*, *Madrid*, *Edinburgh*, *Geneva*, *Sidney*, *Yale*, *Haiti*, *Arkansas*, *Nebraska*, *Idaho*

The problem of phonetic 'false friends' cannot be underestimated as their number runs into hundreds. As a matter of fact, many items in the previous section, i.e. in (1a), (2) and (3) also belong to this category as the issue of 'false friends' is closely connected with interference from spelling.

3.3. Word stress

As is well-known, word stress belongs to the most difficult areas of English phonetics for Poles, who, with their fixed-stress mother tongue, find the intricacies and irregularities of English stress very hard to master. Since this issue is discussed in more detail in other studies (Sobkowiak 1996, Waniek-Klimczak 2002), here I will confine myself to listing the items which the participants provided with such comments as: "*I keep forgetting how to stress this word correctly.*"

The following bisyllabic items appear in the collected data, with the incorrect stress placement indicated:

(10) (a) *'guitar*, *'hotel*, *'event*, *'technique*, *'alarm*, *'success*, *'Japan*, *'exam*, *'support*, *'suspense*
(b) *e'ffort*, *fe'male*, *fo'reign*, *da'maged*, *ca'pable* (when pronounced as [ɛˈfɔːrt], [fɔˈrejn], etc.)

The examples in (10a) tend to receive penultimate stress in Polish English, partly due the stress pattern in the corresponding Polish words (when they exist) and partly due to the transfer of the Polish penultimate stress rule. The forms in (10b) are particularly interesting since here an opposite tendency can be observed, i.e. that of learners' stressing the ultimate syllable. It seems that they develop some sensitivity to syllable weight and stress the final syllable as they regard it, correctly or incorrectly, as heavy and therefore stress-attracting.

Problematic stress placement in longer words concerns the forms in (11):

(11) (a) *in'dustry*, *o'rigin*, *a'lgebra*, *valu'able*, *avai'lable*, *Janu'ary*, *Febru'ary*, *e'nergy*,
edu'cated, *orga'nizer*, *inte´resting*, *fasci'nating*, *demons'trated*,
(b) *unfor'tunately*, *'successful*, *'September*, *'October*, *'November*, *'interpret*, *e'conomic*,
'detergent, *ar'tificial*, *'variety* (pronounced as [ˈvɛrjetɨ])
(c) *'computer*, *'museum*, *'professor*

In the case of items in (11a) learners frequently shift stress onto the penultimate syllable, in accordance with the Polish rule. The examples in (11b) and (11c) depart from this pattern in a curious way; here the antipenultimate syllable tends to be stressed, which is particularly surprising in the case of the words in (11c) with cognates in Polish stressed, like in English, on the penultimate syllable, i.e. *kom'puter*, *mu'zeum*, *pro'fesor*. In these instances learners appear to develop their own version of the English antepenultimate stress rule which operates in items such as *A'merica* and *'marvelous*, and apply it to the forms in (11b and c).

Another generalization that emerges from our analysis is that intermediate learners take into account only the last three syllables as a possible location of stress and fail to stress properly those items in which it falls somewhere else, as shown in (12), where several commonly misstressed four-syllable items are presented.

(12) *ne'cessary, for'tunately, ca'tegory, con'sequently, ar'bitrary, criti'cizing, illus'trating, cha'racterize*

3.4. Difficult consonant clusters

Polish learners, with their mother tongue abounding in a rich variety of consonant sequences, generally have no major problems with the pronunciation of English clusters, with some exceptions, however. Many participants of our study regard as phonetically difficult those words which contain an interdental fricative in combination with another consonant and provide the following examples,

(13) (a) <th+C> *three, throw, threw, through, throat, thriller, threaten, birthday,*
mathematics, maths, rhythm, truthful, faithful, athletic
(b) <C+th> *sixth, seventh, eighth, ninth, tenth, hundredth, thousandth, although,*
length, strength, month, healthy, wealthy, depth, width, warmth, enthusiasm

Thus, the interdental fricatives, difficult for Poles to learn in any context, are particularly troublesome when they appear next to another consonant. As shown in (13), the order of consonants is irrelevant since in (13a) the spirant appears as the first member of a cluster while in (13b) as the second one. The quality of the other consonant does not seem relevant either; the examples in (13) comprise combinations of 'th' with rhotics, nasals, plosives, laterals and fricatives. It should be pointed out, however, that in my data the most frequently repeated examples involve 'th' followed by /r/.

The issues discussed so far belong to well-known problems not only for Poles, but for learners of other L1's as well. In what follows we focus on lesser known sources of phonetic errors uncovered by our study.

3.5. Longer words

The pronunciation of longer words fails to be addressed by the majority of phonetic manuals, which, in fact, do not recognize it as a phonetic issue as such. Yet, such a category has been isolated by several participants, as shown by their comments like "*this word is difficult because it's long.*" Apparently, longer items are problematic for intermediate learners because of a variety of factors they have to control at the same time: the placement of stress, the articulation of many different new sounds and complex sound sequences.

The words considered difficult by the pupils because of their length are listed in (14)

(14) (a) trisyllables: *excitement, adventure, Australia, picturesque, quotation, weightlessness*
(b) quadrisyllables: *relaxation, astonishing, surprisingly, competition, desperately*
(c) quintisyllables: *encyclopaedia, occasionally, exaggeration, association,*
opportunity, simultaneously, Mediterranean

The provided items contain three, four and five syllables. Evidently, for intermediate learners even words which are three-syllable long may be difficult.[2]

3.6. Liquids

One of the most interesting results of this study is the discovery that the presence of several liquids, i.e. rhotics and laterals, in one word contributes to its considerable difficulty for intermediate learners.

One source of such problems is the failure to master the restrictions on the occurrence of /r/ in RP, commonly taught in Poland, and a frequent case of pronouncing it word-finally and preconsonantally. Such realizations of /r/ result in the presence of several rhotics in one item and the learners' complaints that it is difficult since "*there are too many r's in this word.*" The following examples have been supplied by the pupils:

(15) *murderers, portray, cartridge, appropriate, library*

Nevertheless, these are forms which contain both laterals and rhotics (found in pronunciation and/or in spelling) that were frequently provided with such comments as "*I get my tongue in a twist when I try to say this word.*"

(16) *regularly*, *particularly*, *rarely*, *barely*, *burglary*, *world*, *girlfriend*, *rural*, *elderly*, *literally*, *cellular*

In such cases learners frequently attempt to pronounce all r's present in spelling, which is particularly difficult if this creates two and three-consonant clusters. The problem with liquids is expressed clearly by one pupil, who writes that, "*red lorry, yellow lorry – this is a real tongue twister.*"

It should be pointed out that sequences of liquids are problematic for Poles even when they occur in Polish. For example, *kolorowy* 'colourful,' *laboratorium* 'laboratory' and *Labrador* belong to frequently mispronounced words.

3.7. Alternating forms

The collected data contain words regarded as difficult by the respondents due to the fact that in the roots they contain morphological alternations take place. Since in English such changes are often highly irregular and idiosyncratic, this fact contributes to the perceived difficulty of the items in question. Some examples are presented in (17), with related forms provided in parentheses.

(17) (a) *so**c**iety* (so**c**ial), *nor**th**ern* (nor**th**), *sou**th**ern* (sou**th**), *lon**g**evity* (lon**g**er), *an**x**iety* (an**x**ious)

(b) *c**a**n't* (c**a**n), *var**i**ety* (var**i**ous), *br**ea**the* (br**ea**th), *n**u**merous* (n**u**mber), *w**i**dth* (w**i**de),
*d**e**pth* (d**ee**p), *s**ou**thern* (s**ou**th), *sinc**e**rity* (sinc**e**re), *b**a**the* (b**a**th), *n**a**tural* (n**a**ture)

(17a) contains cases which involve consonant alternations while those in (17b) vowel alternations. It is likely that pupils learn first more frequent words given in parentheses and when faced with less common related items, transfer the pronunciation of the underlined segments from the former to the latter. The degree of difficulty increases due to the fact in the above forms the alternating segments are spelt in the same way.

3.8. High front vowels

For intermediate Polish learners difficult English words are also those which contain two different high front vowels, i.e. [iː] and [ɪ], as in (18),[3]

(18) *reading*, *sleeping*, *feeling*, *dreaming*, *cheating*, *ceiling*, *greeting*, *easy*, *speedy*, *greedy*, *sleepy*

In such instances they tend to employ two [iː] vowels (or rather its shorter and less tense Polish counterpart [i]). Interestingly, when these vowels appear in the reverse order, the pronunciation difficulty of such a sequence decreases, e.g.

(19) *believe*, *receive*, *deceive*, *precede*, *repeat*, *release*

However, when the progressive *–ing* suffix is attached to the verbs in (19), a very difficult sequence of [ɪ] – [iː] – [ɪ] is created, as in (20).

(20) *believing*, *receiving*, *deceiving*, *preceding*, *repeating*, *releasing*

Here the vowels in the final two syllables are usually pronounced as [iː], as in the case of the forms in (18).

Yet another problem with [ɪ] is created by the following words:

(21) *innocent*, *image*, *impression*, *important*, *industry*

In these items the initial vowel was indicated by some participants as difficult to pronounce and usually replaced by them with Polish [i]. Apart from the powerful influence of English spelling, another active factor here seems to be the phonotactic constraint of Polish banning in the word initial position the occurrence of the Polish front centralized vowel [ɨ], which is very close to English [ɪ].

4. CONCLUSION

Over 250 lexical items presented and analysed in this paper provide a rich source of information on the issue of phonetically difficult words for intermediate Polish learners' of English. Since we have dealt here with the pupils' subjective judgements, the emerging picture is far from complete and needs to be supplemented by a direct observation of the learners' performance.

According to the participants of this study, what makes English words difficult to pronounce is the presence of phonetic 'false friends' in both languages, the frequently nonphonemic English spelling, lack of fixed word stress, consonant clusters which include interdentals, length of words (three syllables and more), the occurrence of several liquids, a sequence of high front vowels and irregular morphological alternations in words. It should be added that, apart from the first three factors which are well-known and discussed in the pronunciation literature, the remaining ones have been uncovered by this study.

It is also interesting which lexical items were most frequently listed as troublesome. In (21) 15 such words are provided.

(22) *through*, *valuable*, *particularly*, *regularly*, *certainly*, *sixth*, *foreign*, *gigantic*,
encyclopaedia, *rarely*, *unfortunately*, *Australia*, *early*, *can't*, *vague*

These examples involve the majority of problems discussed in this paper. Thus, three of them are 'false friends' (*encyclopaedia*, *Australia*, *gigantic*), three are difficult because of stress (*valuable*, *foreign*, *unfortunately*), three are longer words (more than four syllables) (*particularly*, *encyclopaedia*, *unfortunately*), two contain clusters of interdentals and other consonants (*through*, *sixth*), three comprise some liquids (*particularly*, *regularly*, *rarely*), one involves an irregular morphological alternation (*can't*) and most are problematic because of spelling-to-sound correspondences. In many cases several factors contribute to the perceived phonetic difficulty of one item. For example, *through* contains not only a cluster of an interdental fricative and a rhotic, but also a complex letter combination, i.e. <ough>. *Valuable* is a problem due to the presence of the suffix <able> as well as the occurrence of two laterals.

There are several important pedagogical implications that stem from this study. First of all, it fully supports claims made by Szpyra-Kozłowska (in press) and Szpyra-Kozłowska and Stasiak (in press) concerning the need to prioritize word pronunciation in phonetic instruction. As evidenced here, there are hundreds of English words perceived as phonetically difficult by learners and many more of which they are not aware. No training which focuses on sounds, usually practiced in simple monosyllabic items or minimal pairs, can eliminate error-prone lexemes. Secondly, phonetic difficulty is closely connected with learners' language proficiency and should be approached in this particular context. This means that we agree with Nation (1990) that the order in which learners are taught vocabulary should take into account the level of phonetic difficulty of English words to reduce part of the learning burden and enhance the effectiveness of this process.

5. REFERENCES

Nation, P. 1990. *Teaching and Learning Vocabulary.* New York: Newbury House.

Sobkowiak, W. 1996. *English Phonetics for Poles.* Poznań: Bene Nati.

Sobkowiak, W. 1999. *Pronunciation in EFL Machine-Readable Dictionaries.* Poznań: Motivex.

Szpyra-Kozłowska, J. (in press). On the irrelevance of sounds and prosody in foreign-accented English. In Waniek-Klimczak (ed.) *Issues in Accents of English 3*. Newcastle: Cambridge Scholars Publishing.

Szpyra-Kozłowska, J. & S. Stasiak (in press). From focus on sounds to focus on words in English pronunciation instruction. In Waniek-Klimczak (ed.) *Issues in Accents of English 3*. Newcastle: Cambridge Scholars Publishing.

Waniek-Klimczak, E. 2002. Akcent wyrazowy w nauczaniu języka angielskiego. In Sobkowiak, W. & E. Waniek-Klimczak (eds) *Dydaktyka Fonetyki Języka Obcego. Zeszyty Naukowe PWSZ w Płocku,* t. III. Płock: Wydawnictwo Naukowe PWSZ w Płocku. 101-114.

NOTES

[1] As the experiment had a written form, in some cases it was not clear what type of pronunciation difficulty some of the provided items represented if no comments were included.

[2] Sobkowiak (1999) includes words which comprise five or more syllables into his Phonetic Difficulty Index.

[3] This problem concerns those learners who make a distinction between two vowels under consideration. It should be added that many Poles tend to replace most occurrences of English [ɪ] with Polish [i].

Accent placement in the absence of heavy syllables: the case of British English learners of Japanese

Becky Taylor

Nagoya University
rebecca.taylor@e.mbox.nagoya-u.ac.jp

ABSTRACT

This paper presents novel experimental production data to establish generalisations about accent patterns produced by British English (BE) learners of Japanese in utterances consisting of a sequence of light syllables. The results show both considerable inter-learner variability and within-learner systematicity. Some sensitivity to the accent types of Standard Japanese was observed, but this was the exception rather than the rule, even in the case of advanced learners. Instead, the learners have the common characteristic that utterances that have the same properties - part of speech, number of mora, presence or absence of a following function word - tend to be produced with the same, or a similar distribution of, accent types. This suggests that the learners are making generalisations about accent over different lexical item *types*, rather than learning the accent of each lexical item. Variation between learners was observed not only in the specific accent types produced, but also the types of stimuli over which generalisations are made. Whether the BE learners are encoding pitch accent in their phonological representation is discussed.

Keywords: Japanese pitch accent, production, British English learners, inter-learner variability, within-learner systematicity

1. INTRODUCTION

Previous research into the production of Japanese accent patterns by English learners of Japanese has shown that the learners tend to produce words unaccented, or to place an accent either two or three mora from the end of a word (Horiguchi 1973, Toki 1980, Yoshimitsu 1981, Yamada 1994). The dominant accent type has been reported to differ between learners (Kuno 1998), and accent types have been reported to vary in repeated words (Toki 1980, Yamada 1994). Heavy syllables in both Japanese and English tend to be accented (Kubozono 2006), and there is some evidence for a relationship between syllable structure and accent placement in the production of English learners of Japanese (Horiguchi 1973). Some learners have, however, been observed instead to increase the amount of unaccenting in words with heavy syllables (Kuno 1998). Other than this, it is not known which words learners produce with which accent type. The current study focuses on words containing only light syllables, in order to determine which factors affect the accent types produced.

2. PROCEDURE

This paper presents novel experimental production data to establish generalisations about accent patterns produced by British English (BE) learners of Japanese in utterances consisting of a sequence of light syllables. Two groups of learners were studied, 13 learners with *less Japanese experience* (less than 450 hours of classroom time and between zero and three months in Japan), and 8 learners with *more Japanese experience* (at least 650 hours of classroom time and at least 10 months in Japan). The 21 BE learners of Japanese read aloud, in pseudo-randomized order, 312 stimuli of 8 *stimuli types* differing in (i) number of mora of the content word (2 or 3), (ii) part of speech (noun or verb) and (iii) presence or absence of a following function word. Each *stimuli type* contained 12 stimuli of each of the possible accent types in Standard Japanese (SJ). The SJ accent types of each stimuli type are shown in table 1. (These are all the possible accent types for these mora numbers and parts of speech). The codes shown (e.g. nc2in) will be used when presenting the results.

Table 2: The eight stimuli types

Stimuli type	SJ accent type			
	initial	medial	final	unaccented
2 mora isolated nouns	nc2in			nc2ta**, nc2un
2 mora nouns before function word	np2in		np2ta	np2un
3 mora isolated nouns	nc3in	nc3me		nc3ta**, nc3un
3 mora nouns before function word	np3in	np3me	np3ta	np3un
2 mora isolated verbs	vco2a			vco2u
2 mora verbs before function word*	vpn2a,vpo2a		vpn2u	vpo2u
3 mora isolated verbs		vco3a		vco3u
3 mora verbs before function word*		vpn3a, vpo3a	vpn3u	vpo3u

*Two function word types to give both final and unaccented SJ **Final accent not realised in isolation

The majority of the stimuli were chosen from the beginners textbook *Minna No Nihongo* (3A Corporation, 1998). The words were displayed to the learners in *kanji* (Chinese characters), with the phonetic reading displayed in the *hiragana* script. An English translation of the key word (not the following function word) was provided. The learners were encouraged to repeat a word if they were very hesitant or if they mis-read the script. In the very few cases where errors remained to the extent that the word was not recognisable (e.g. different number of syllables, more than one segmental error) the productions were discarded, but productions with minor errors (e.g. voicing, non-Japanese liquids) were included in the analysis. The accent types produced were identified by three phonetically-trained native Japanese speakers. Where their responses differed, the majority response (two out of three) was used. If all three responses differed, the data was not used in the analysis.

3. RESULTS

Table 2 shows the accent types produced by each learner for each stimuli type. Subject names beginning with LF or LM are in the less experienced group and MF and MM are in the more experienced group. For each stimuli type, any accent type produced at 20% or higher is included in the table. When there is a dominant accent type at least 30 percentage points higher that any other, it is capitalised. 'Init' (initial), 'med' (medial), 'fin' (final), and 'func' (function word) refer to the position of the accent produced, and 'un' is unaccented.

It can be seen from table 2 that there is considerable inter-learner variation in the accent types produced. Two learners (LM_MT and MM_TG) are (almost) entirely unaccented, three learners (LF_AG, LF_LC and LM_JR) are mainly unaccented but also show some accenting, two learners (LM_JO and MF_PH) have both unaccented and accented with no dominant pattern, nine learners (LF_GC, LF_LD, LF_MB, LM_JE, LM_NB, MF_KW, MM_DT, MM_JB and MM_LH) are mainly accented but also show unaccenting, and five learners (LF_EP, LF_JW, LM_MD, MF_NT and MM_DM) are only accented (at least at the 20% level for each stimuli type). The majority of learners do not produce final accent (at least at the 20% level) but of the five learners that do (LF_EP, LM_JE, LM_JO, LM_MD and MF_PH), one learner (LF_EP) has final accent as a dominant accent type for one stimuli type (2 mora verbs before a function word). Lastly, two learners (LM_JE and MM_JB) have a tendency to place an accent on the function word following 3 mora verbs.

Variation is also seen in whether a stimuli type is produced with one dominant accent type, or two or more accent types. Two learners (LF_AG and LM_JR) have a tendency for one dominant accent type to change to two with the addition of a function word, whereas five learners (LF_LC, LF_MB, MF_KW MM_JB and MM_LH) show the opposite effect, with two or more accent types per stimuli type in isolation, and one dominant accent type before a function word. In this way, which stimuli are produced with which accent types is seen to depend on the entirely on the learner, with none of the above variation seen to correspond to the relative Japanese experience of the learners.

Table 2: The accent types produced by each learner for each of the eight stimuli types

	Stimuli type							
	n2		n3		v2		v3	
Subject	c	p	c	p	c	p	c	p
LF_AG	UN	init, un	UN	init, un	init, UN	init, un	UN	med, un
LF_EP	INIT	init, fin	init,MED	MED	INIT	init, FIN	MED	MED
LF_GC	init, un	INIT	init, med	init,MED	INIT	INIT	init, med	init,MED
LF_JW	INIT	INIT	INIT,med	init, med	INIT	INIT	init, med	MED
LF_LC	init, un	INIT, un	med, un	med, un	init, un	UN	med, UN	UN
LF_LD	INIT	INIT	init, un	init, med	INIT	INIT	MED,un	MED
LF_MB	INIT, un	INIT	init, un	init, med	init, un	INIT	med, un	med, un
LM_JE	INIT	INIT	INIT,med	MED	IN	IN, fin	med, un	med,func
LM_JO	init, un	INIT, un	init,med,un	med, un	init, UN	INIT	UN	fin, un
LM_JR	UN	init, un	UN	med, un	UN	init, un	UN	MED
LM_MD	INIT	init, fin	init, med	MED	INIT	init, fin	MED	MED
LM_MT	UN	UN	UN	UN	UN	UN	UN	UN
LM_NB	init, un	INIT	init,med,un	MED	init, un	INIT	init,med,un	MED
MF_KW	INIT,un	INIT	init,med,un	INIT,med	INIT	INIT	MED, un	MED
MF_NT	INIT	INIT	INIT	init, med	INIT	INIT	init, med	MED
MF_PH	init, UN	init,fin,un	med,un	MED	init, un	init, fin	med, un	MED
MM_DM	INIT	INIT	INIT	init, med	INIT	INIT	init, MED	MED
MM_DT	INIT, un	INIT	INIT, med	init, med	INIT, un	INIT	MED	MED
MM_JB	INIT, un	INIT	init,med,un	init,MED	INIT, un	INIT	med, un	MED,func[1]
MM_LH	init, un	INIT	init, un	init, med	INIT, un	INIT	init,med,un	MED
MM_TG	UN	UN	UN	med, UN	UN	UN	UN	UN

Table 3 summarizes further the accent types produced by each learner. Each stimuli type is separated into groups of 12 words, corresponding to the accent types of the words in standard Japanese: rather than 8 stimuli types, here there are 26. Because some productions were discarded from the analysis, some stimuli types contain between 8 and 11 words, although the majority have the full 12. The results are presented as follows: one or two instances of an accent type are shown in brackets, 3 or more in normal (non-capitalised) letters, and if there is a dominant accent type of at least 3 instances greater than any other, it is shown in capitals. Those stimuli types that have the same distribution of accent types are grouped together, with SJ-like and non SJ-like variation shown as an increase (↑) or decrease (↓) of that accent type[2]. 'All' refers to all the possible accent types for that stimuli type, '-two' and '-three' imply an accent two or three mora from the end of the word and 'init&func' and 'med&func' are used to show that a accent is placed on both the initial/medial mora and the following function word.

The trends shown in table 3 are summarized below.

- LF_AG: Unaccented in isolation, two or three accent types before a function word. Verb patterns depend on type of function word. Initial accent is increased in SJ initially-accented words.
- LF_EP: Isolated words are accented 2 mora from the end. Before a function word, dominant accent type depends on mora number. Some signs of SJ across all stimuli types.

Table 3: The accent types produced by each learner for each SJ-separate stimuli type

<table>
<tr><th>Subject</th><th>Accent types produced</th><th>SJ-like variation</th><th>Non SJ-like variation</th></tr>
<tr><td>LF_AG</td><td><table><tr><td>c</td><td>np2</td><td>np3</td><td>vp3</td><td>v<u>pn</u>2</td><td>v<u>po</u>2</td></tr><tr><td>UN</td><td>init,un</td><td>init,med,un</td><td>med,un</td><td>init,fin</td><td>init,un</td></tr></table></td><td>↑init:nc3in,v<u>co</u>2a,np2in,np3in,v<u>pn</u>2a ↑med:v<u>pn</u>3a ↓med:np3un</td><td>↑init:v<u>co</u>3a</td></tr>
<tr><td>LF_EP</td><td><table><tr><td>c</td><td>p2</td><td>p3</td></tr><tr><td>-TWO</td><td>init,FIN</td><td>(init),MED,(fin)</td></tr></table></td><td>↑init:nc3in,np2in ↓med:v<u>pn</u>3u ↓fin:v<u>pn</u>2a ↑un:nc2ta,v<u>co</u>2u,nc3un</td><td>↑init:np2un</td></tr>
<tr><td>LF_GC</td><td><table><tr><td>nc2</td><td>nvp2</td><td>c3</td><td>p3</td></tr><tr><td>all</td><td>INIT</td><td>init,med</td><td>init,MED</td></tr></table></td><td>↑init:nc3in ↓med:np3in</td><td>↑init:nc2un</td></tr>
<tr><td>LF_JW</td><td><table><tr><td>2</td><td>n3</td><td>vc3</td><td>vp3</td></tr><tr><td>INIT</td><td>INIT,med</td><td>init,med</td><td>(init),MED</td></tr></table></td><td>↓med:nc3in ↑med:np3me</td><td>↓med:v<u>pn</u>3a</td></tr>
<tr><td>LF_LC</td><td><table><tr><td>c2, np2</td><td>c3, np3</td><td>vp</td></tr><tr><td>INIT,un</td><td>(init),med,un</td><td>(-two),UN</td></tr></table></td><td>↓init:np2ta,nc2un ↓init↑un:nc2ta ↑un:nc3un,v<u>co</u>3u ↓un:v<u>pn</u>2a</td><td>↓init:v<u>co</u>2a ↑un:nc3me,v<u>co</u>3a,np3me ↑fin:v<u>pn</u>2a</td></tr>
<tr><td>LF_LD</td><td><table><tr><td>2</td><td>n3</td><td>v<u>o</u>3</td><td>v<u>pn</u>3</td></tr><tr><td>INIT,(un)</td><td>INIT,med,un</td><td>(init),MED,un</td><td>MED</td></tr></table></td><td>↓init:nc3me,nc3un ↓init↑med:np3me ↓med:np3ta</td><td>↓med:v<u>po</u>3a</td></tr>
<tr><td>LF_MB</td><td><table><tr><td>nc2, v<u>o</u></td><td>np2, v<u>pn</u></td><td>nc3</td><td>np3</td></tr><tr><td>-two, un</td><td>-TWO,(fin),(un)</td><td>all</td><td>init,med,(fin),(un)</td></tr></table></td><td>↑init:nc2in,v<u>po</u>2a ↑med:v<u>co</u>3a,np3me ↑un:v<u>co</u>3u,v<u>po</u>3u ↓med:v<u>pn</u>3u,nc3in</td><td>↑init:nc2ta,nc3ta,np3ta ↑med:np3in</td></tr>
<tr><td>LM_JE</td><td><table><tr><td>c2</td><td>c3</td><td>p</td></tr><tr><td>INIT,(un)</td><td>all</td><td>-TWO,(-three),(fin),(func),(two&func),(un)</td></tr></table></td><td>↑init:nc3in ↑med:v<u>co</u>3a ↓init:v<u>pn</u>2u ↓med:np3in,np3ta,v<u>po</u>3u</td><td>↑init:nc3in</td></tr>
<tr><td>LM_JO</td><td><table><tr><td>c</td><td>p2, np3</td><td>vp3</td></tr><tr><td>all,UN</td><td>-TWO,(three),(fin),un</td><td>med,un,(fin),(func)</td></tr></table></td><td>↓un:v<u>co</u>2a,nc3in,nc3me,v<u>po</u>2a,↓med:np3in,np3ta,np3un ↓init:v<u>po</u>2u ↑fin:v<u>pn</u>3u ↑un:v<u>po</u>3u</td><td>↓un:nc2un, ↑un:np3ta,v<u>pn</u>3a</td></tr>
<tr><td>LM_JR</td><td><table><tr><td>c</td><td>np, v<u>po</u>2</td><td>v<u>pn</u>2, vp3</td></tr><tr><td>UN</td><td>-two,un</td><td>-TWO,(fin)</td></tr></table></td><td>↑init:v<u>po</u>2a ↑un:np3un,v<u>po</u>2u,v<u>po</u>3u</td><td>↑un:np2in,np3in</td></tr>
<tr><td>LM_MD</td><td><table><tr><td>c2</td><td>c3</td><td>p2</td><td>p3</td></tr><tr><td>INIT,(un)</td><td>(in),MED</td><td>INIT,fin</td><td>(init),MED,(fin)</td></tr></table></td><td>↑fin:v<u>pn</u>2u ↓init:np2un,<u>vpo2u</u></td><td></td></tr>
<tr><td>LM_MT</td><td><table><tr><td>All</td></tr><tr><td>UN</td></tr></table></td><td>↑init:v<u>pn</u>2a</td><td></td></tr>
<tr><td>LM_NB</td><td><table><tr><td>c</td><td>p</td></tr><tr><td>all</td><td>-TWO,(all),(func),(two&func)</td></tr></table></td><td>↑un:v<u>co</u>2u,nc3ta ↓med:np3ta</td><td>↑un:v<u>co</u>3u</td></tr>
<tr><td>MF_KW</td><td><table><tr><td>2, v3</td><td>n3</td></tr><tr><td>-TWO,(un)</td><td>init,med,un</td></tr></table></td><td>↑init:nc3in,np3in ↑med:np3me ↑un:nc3un ↓med:np3un</td><td>↑init:np3ta</td></tr>
<tr><td>MF_NT</td><td><table><tr><td>2</td><td>nc3</td><td>np3, vc3</td><td>vp3</td></tr><tr><td>INIT</td><td>INIT,(all)</td><td>init,med</td><td>init,MED</td></tr></table></td><td>↑init:np3in, ↑med:np3me</td><td>↑init:np3ta ↑med:np3un ↓med:v<u>po</u>3a</td></tr>
<tr><td>MF_PH</td><td><table><tr><td>nc2</td><td>nc3, np2</td><td>vc</td><td>np3</td></tr><tr><td>init,UN</td><td>all</td><td>-two,un</td><td>MED,(all)</td></tr><tr><td>v<u>po</u>2</td><td>v<u>po</u>3</td><td>v<u>pn</u>2</td><td>v<u>pn</u>3</td></tr><tr><td>init,fin,func,(un)</td><td>med,fin,func</td><td>init,fin</td><td>MED,(fin)</td></tr></table></td><td>↑un:nc3un,np3un ↑init:v<u>po</u>2a ↑med:v<u>po</u>3a</td><td></td></tr>
</table>

MM_DM	2	v3	n3				↑**init**↓**med:**nc3in ↑**init:**np3in ↓**un:**nc3me	↓**med:**vco3a ↑**init:**np3un
	INIT,(un)	(init),MED	init,med,un					
MM_DT	c2	np2,vpn2	vco3,vpn3	nc3	np3	vpo	↓**init:**nc2ta ↑**init:**nc3in ↑**init**↓**med:**np3in ↓**med:** np3ta ↓**med**↑**un:**vpo3u	
	INIT, un	INIT	(init),med	all	init, med	-TWO, func		
MM_JB	n2	nc3	np3	vc	vpn	vpo	↓**init:**nc2ta ↑**un:**nc3un ↓**med:**np3un ↓**med**↑**un:**vco3u	
	INIT, un	all	init,MED,un	-two, un	-TWO, (fin), (un)	-two,func, -two&func		
MM_LH	c2	np2, vp2	c3	np3	vpo,vpn3		↑**init:**nc3in ↓**init:**nc2ta ↑**un:** nc3un,vco3u ↓**med:**vpn3u,vpo3u	↑**init:**vco2a,vpo2a
	INIT, un	INIT, (fin)	all	INIT,med, (un)	-two,(fin),(func), (un)			
MM_TG	c2	n3	vco3	p2	vp3			
	(init), UN	(init), (med),UN	(med), UN	(in),(fin), UN	(med),(fin),UN			

- LF_GC: Only isolated 2 mora nouns are produced unaccented. The addition of a function word changes two accent types to one dominant type. Few variations.
- LF_JW: 2 mora nouns are initially accented with no variation. Accent types of 3 mora nouns vary with part of speech, presence of a function word and also SJ.
- LF_LC: Verbs become unaccented with the addition of a function word, nouns do not. Some SJ-like and non-SJ like variation.
- LF_LD: 2 mora words have dominant initial accent. 3 mora words vary according to part of speech and presence/type of function word. Some minor variation, mostly SJ-like.
- LF_MB: Difference between three mora nouns and verbs. SJ *final* accent nouns show increased *initial* accent. 3 mora verbs show SJ-like variation.
- LM_JE: Isolated words pattern differently according to mora number. An accent two mora from the end of the word dominates before a function word. Most variation is SJ-like.
- LM_JO: Isolated words are unaccented. Accent patterns before a function word depend on mora number and part of speech. Strong SJ-like variation for 3 mora unaccented verbs before a function word.
- LM_JR: Nouns show one type (unaccented) in isolation and 2 types before function word. Verbs pattern differently according to the type of function word. Some SJ-like variation before a function word.
- LM_MD: Dominant accent two mora from end of word, with final accent before function words. Dominant pattern changes to final for 2 mora verbs before a function word, as per SJ.
- LM_MT: Unaccented almost without exception. 5 verbs are accented as SJ.
- LM_NB: Two or three accent types in isolation, one before a function word. Some SJ-like variation.
- MF_KW: 2 mora words have dominant initial accent. 3 mora has initial accent only in nouns. Verbs unaccented only in isolation. SJ-like variation only in 3 mora nouns.
- MF_NT: Initial or medial accent depending on mora number and part of speech. Minor variation, some SJ-like and some non-SJ like.
- MF_PH: Eight different accent patterns, with only minor variation. Variation is SJ-like.
- MM_DM: All 2 mora words pattern the same, but 3 mora words vary with part of speech. Most variation in 3 mora nouns, mostly SJ-like.
- MM_DT: Six different patterns. Three mora nouns show strong SJ-like variation.
- MM_JB: Six patterns, with difference between two types of function words after verbs. SJ-like increase in unaccented words.
- MM_LH: Five patterns. Some minor variation, most SJ-like.
- MM_TG: Dominant unaccenting. Five patterns with no variation.

These results show that although there is considerable between-learner variation, there is also considerable within-learner systematicity. The learners have the common characteristic that stimuli with the same properties - part of speech, number of mora, presence or absence of a function word - tend to be produced with the same accent type (or a similar distribution of accent types) with only minor variations. Many learners showed some sensitivity to the accent types of Standard Japanese in some stimuli types (especially in three mora nouns) but this is the exception rather than the rule, even in the case of advanced learners. This suggests that the learners are making generalisations about accent over different lexical item types, not learning the accent of each lexical item. Variation between learners was observed not only in the specific accent types produced, but also the types of stimuli over which generalisations are being made.

4. DISCUSSION

This paper shows that there is no one accent type that can be considered to be representative of BE learners for any type of stimuli: initial, medial, final, unaccented and all combinations of these are observed. The learners were seen to have the common characteristic that utterances that have the same properties - part of speech, number of mora, presence or absence of a following function word - tended to be produced with the same, or a similar distribution of, accent types. This suggests that the learners are making generalisations about accent over different lexical item *types*, rather than learning the accent of each lexical item. However, when there was deviation from these patterns, the majority of it was in the direction of Standard Japanese. In addition, although all the learners appear to be making generalisations about accent over different types of stimuli, the types of stimuli they generalise over - same part of speech, same number of mora, same presence or absence of a function word - also differed from learner to learner, as well as the generalisations that are made.

An implication of this combination of inter-learner variability and within-learner systematicity could be that the learners are not encoding pitch accent in their phonological representations. This conclusion has been made about French learners of Spanish regarding lexical stress (Dupoux et. al. 1997, 2001, 2007). Whether the absence of pitch accent encoded in the phonological representation could lead to the type of generalisation over stimuli type seen here, why some Standard Japanese-like variation was observed, and why the learners' level of Japanese experience did not lead to a difference in the amount of Standard Japanese-like patterning, are questions for further research.

5. REFERENCES

3A Corporation (1998). *Minna no Nihongo skokyu I, II.*

Dupoux, E., Pallier, C., Sebastian, N., Mehler, J. (1997). *A destressing "deafness" in French?* Journal of Memory and Language, 36, 406-421.

Dupoux, E., Peperkamp, S., Sebastian-Galles, N. (2001). *A robust method to study stress "deafness".* Journal of the Acoustical Society of America 110(3), 1606-1608.

Dupoux, E., Sebastian-Galles, N., Navarrete, E., Peperkamp, S. (2007). *Persistent stress 'deafness': The case of French learners of Spanish. Cognition, 106, 682-706.*

Horiguchi, J. (1973). Eigokokumin ni yoru nihongo no yon-onsetsu meishi no akusento no yosoku to sono jissai (Predicted and actual accents types of 4-mora nouns produced by English speakers). *Nihongo Kyouiku,* 19, 97-112.

Kubozono, H. (2006). Where does loanword prosody come from? A case study of Japanese loanword accent. Lingua, 116, 1140-1170.

Kuno, S. (1998). Bei-eigowasha ni okeru nihongo akusento no seisei (The production of Japanese accent by American English speakers). *On-in Kenkyuu*, 1, 83-90.

Toki. T. (1980). Eigo wo bogo to suru gakushuusha ni okeru akusento no keikou (Accent trends in learners with English as a first language). Amerika, Kanada Juuichi Rengou Nihon Kenkyuu Sentaa Kiyou, 3, 78-96.

Yamada, N. (1994). Nihongo akusento shuutoku no ichidankai: gaikokujin gakushuusha no baai (A stage in acquiring Japanese accent: the case of foreign learners). Nihongo Kyouiku, 83, 108-120.

Yoshimitsu, K. (1981). Gaikokujin gakushuusha no akusento (Foreign learners' accent). Nihongo Kyouiku, 45, 63-75.

NOTES

[1] Some had an accent on both the second mora and the function word.

[2] Differences between the presence or absence of brackets, or a bracketed accent type and zero, are ignored.

The impact of visual cues and lexical knowledge on the perception of a non-native consonant contrast for Colombian adults

Michele Thompson & Valerie Hazan
Speech, Hearing and Phonetic Sciences, UCL, London, UK

m.thompson@ucl.ac.uk, v.hazan@ucl.ac.uk

ABSTRACT

The study investigates the impact of visual cues and lexical knowledge on the identification of a non-native phonemic contrast. Twenty native Colombians were tested on an identification task involving 16 minimal pairs of English words, produced by four English speakers, contrasting in the presence of /b/ or /v/ in initial or medial position. The test was run in three conditions: audiovisual (AV), audio only (A) or visual only (V). Prior to the identification task, their knowledge of the lexical items was evaluated; they were also recorded while reading the words. Mean identification scores were higher for the AV than the A condition, but V and AV scores not differ. Relative to previous /b/-/v/ studies with Peninsular Spanish speakers, Colombians relied more heavily on visual cues in their identification of /b/-/v/. Although there was a trend for identification scores to be higher for known lexical items, this effect was not statistically significant. Finally, production accuracy for the /b/-/v/ contrast was not correlated with perception accuracy, but production tended to be more accurate in speakers with better lexical knowledge. The visual weighting results suggest that the degree of visual bias in speech perception may be 'culture-specific' rather than merely 'language-specific'.

Keywords: visual cues, lexical effects, L2 perception

1. INTRODUCTION

Face-to-face communication is rich in multisensory information, with both auditory and visual cues available to the listener. Weighting of visual cues typically increases where the auditory channel is degraded (e.g., Sumby & Pollack, 1954). As listening in an L2 can be considered as a form of auditory degradation, it is of interest to consider how second language learners use visual cues when perceiving speech?

Investigation of Japanese speakers' ability to identify contrasts through manipulation of the McGurk effect was conducted by Sekiyama and colleagues (Sekiyama et al, 1993, 2003, 2008) and it was established that Japanese speakers showed less of a visual bias than native English speakers. This may in part due to the fact that visual cues are less informative in some languages, as languages that comprise a greater number of visemes may be more salient for speech-reading (Sekiyama & Burnham, 2008; Hazan et al, 2005). Furthermore, there may also be a cultural perspective to consider as individuals may learn to rely more on auditory cues in cultures where it is considered disrespectful to look at the speaker (Sekiyama & Tohkura, 1993). Hazan et al (2006) explored the effect of visual cues on the perception of non-native phonemic contrasts for Spanish and Japanese listeners in audio visual (AV), auditory-only (A) and visual-only (V) conditions. The contrast chosen was the labial-labiodental contrast, which does not have phonemic status in either Spanish or Japanese. For both the Spanish and Japanese groups, superior performance was obtained in the AV condition followed by the A condition and then the V condition. Spanish listeners' performance was superior in all conditions, with significantly better identification in the V condition than the Japanese listeners. Hazan et al. concluded that familiarity with a visual gesture in the native language may impact on the acquisition of the L2 phoneme; indeed, Spanish speakers are exposed to the labiodental gesture in their native language even though it is not contrastive with the labial gesture, whereas Japanese speakers do not.

Other recent research has focused on the question of whether lexical effects influence the ability to discriminate non-native phonetic contrasts. Mora (2005), in a study with Spanish and Catalan learners of English; significantly higher identification scores were obtained for words than non-words for advanced learners of English, suggesting that lexical knowledge is a significant factor in perceiving non-native phonemic contrasts. Hayes-Harb (2007) compared the effect of phonetic training using minimal pairs (lexical information) or statistical learning (phonetic information) in a group of monolingual English speakers and reported that both lexical and statistical learning can contribute to the acquisition of a non-native contrast.

The aim of this study was to investigate how Colombian speakers make use of visual and lexical information in their identification of a non-native consonant contrast. The phonemes chosen were the voiced bilabial plosive /b/ and the voiced labiodental fricative /v/. In Colombian Spanish, /b/ is produced as a voiced bilabial plosive unless it is intervocalic or in the post-nuclear position (Moreno & Marino, 1998) when it is produced as the voiced bilabial fricative /β/. /v/ is often substituted for /b/ in the Spanish vernacular but is never found in phonemic contrast with it. It also exists as an allophone of /f/ before voiced consonants.

It is of particular interest to test Colombians as they are reputed to make greater use of eye contact than speakers from Spain. If a similar weighting of A and V cues is found as in the Spanish cohort in Hazan et al. (2005), this would suggest a language effect on the weighting of visual cues (given their shared L1), whereas if they show a greater weighting of visual cues, this would indicate a cultural influence on the weighting of visual cues in speech perception.

The research questions driving this study were as follows:

1. Do Colombian speakers make significant use of visual cues in disambiguating a non-native contrast?
2. Is the ability to discriminate a word pair related to whether the items are known or unknown?
3. Is the ability to produce the novel contrast related to the Colombian speakers' ability to perceive it?

2. METHODOLOGY

The aim was to examine how native Colombians identify the English consonants /v/ and /b/ in three conditions: audio-visual (AV), audio only (A) and visual only (V) when they are in word-initial and word-medial positions. The experiment comprised four sections: (1) completion of an English language questionnaire, (2) translation of the words used in the experiment to establish lexical knowledge, (3) a speech perception identification test, (4) speech recordings of the words used in the experiment. All participants received a consent form and instructions translated into Spanish.

2.1. Participants

Twenty Colombian speakers (15 women and 5 men) were tested in Bogota. Their level of English proficiency was estimated from their responses to an English language questionnaire and an informal evaluation of their speech by the first author, a trained EFL teacher. 55% of the participants were classified as being of low proficiency and 45% as being of intermediate/advanced proficiency. The participants were aged between 19 and 60 years (M=35.2 years; SD 14.5). All had received some basic level of English teaching at secondary school. Five native speakers of British English (4 women, 1 man) served as the control group. They were aged between 22 – 56 years (M=37 years; SD 13.2).

2.2. Materials

Sixteen minimal pairs contrasting /v/ and /b/ were chosen. The word-initial pairs were: ballet-valet, bat-vat, bend-vend, best-vest, bigger-vigour, boat-vote, bowel-vowel, burble-verbal. The word-medial pairs were: cupboard-covered, dribble-drivel, fibre-fiver, hobble-hovel, marble-marvel, rebel-revel, sabre-savour.

2.2.1. Computer-controlled speech perception task

Four speakers (2 women, 2 men) with a Southern British English accent each recorded 32 words on video for a previous study (Hazan et al, 2005), giving a total of 356 items across the three conditions (unfortunately some of the individual tokens were missing resulting in the loss of 28 tokens). Stimuli for the A condition was generated by stripping out the audio track from the video recording (leaving a blank screen for this part of the task), and the V condition was generated though the removal of the audio channel on the video. The listening experiment was designed using DMDX software (Forster, 2002). Two different orders of presentation of the three modalities were counterbalanced across participants: AV, A, V and A, AV, V. Within each test condition, tokens were randomised across speakers.

2.2.2. Identification of Colombian participants' productions by native English speakers

The eight minimal pairs that were the most reliably produced by Colombian participants were chosen for the identification test with native English speakers. Productions of these 16 words by each participant were digitised and intensity-normalised. They were then presented in an identification test designed using Praat software (Boersma & Weenink, 1997). Ten native English speakers who had no knowledge of Spanish, performed the task. After hearing each word produced by a Colombian speaker, they had to decide whether the word contained a /b/ or a /v/ by clicking on either the B or V label on the screen.

2.3. Test procedure

After completing the English language questionnaire and translating the stimuli, The Colombian participants performed the computer-controlled identification task. This was presented on a laptop with the stimuli presented at a comfortable listening level through headphones. After hearing or seeing each word, they had to click on the label B or V to indicate which consonant they heard. This part of the test was approximately 35 minutes in duration. During the final part of the experiment, a Sanyo Digital Talkbook (ICR-B80NX) was used to record the participants while they read out the test words from a printed sheet.

3. RESULTS

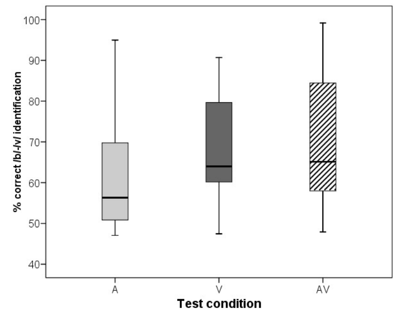
3.1. Identification Task – Control Participants

Mean scores for the five native controls were: A condition: 97.5 % (SD 1.3); V condition: 94.2% (SD 2.0) and AV condition: 99.2% (SD 1.19). T-tests showed a significant difference in performance between AV and V conditions ($t(8) = 4.69$, $p = 0.002$) and A and V conditions ($t(8) = 2.98$, $p = 0.017$), but that the A and AV conditions did not differ ($t(8) = -2.15$, $p = 0.064$). These data suggests that the /b/-/v/ contrast is very salient visually for native speakers.

3.2. Identification Task – Colombian Group Results

The mean scores across the three conditions (see Figure 1) were: A 62.7% (SD 14.9); V 68.1% (SD 13.1); AV 70.0% (SD 16.8). There was a great deal of individual variability in this task: mean identification scores across all three conditions ranged from chance (48.9%) to near ceiling level (94.4%). A repeated-measures ANOVA with within-subjects factors of condition (A, AV and V) and position (word initial, word medial) and between-subjects factor of proficiency indicated a significant main effect of condition ($F(2,36) = 7.013$, $p = 0.003$, $\eta_p^2 = .280$). Paired comparisons with Bonferroni adjustments showed that the condition effect was due to better performance in the AV and V conditions (which did not differ from each other) compared to the A condition. The main effect for position was also significant ($F(1,18) = 26.77$, $p < 0.001$, $\eta_p^2 = .598$) with higher identification rates for the word-initial tokens (M=71.2%) than the word-medial tokens (M=63.5%). There was no significant interaction between condition and position. The effect of language proficiency was not significant: proficiency in English for this group therefore had little impact on the perception of the /b/-/v/ contrast.

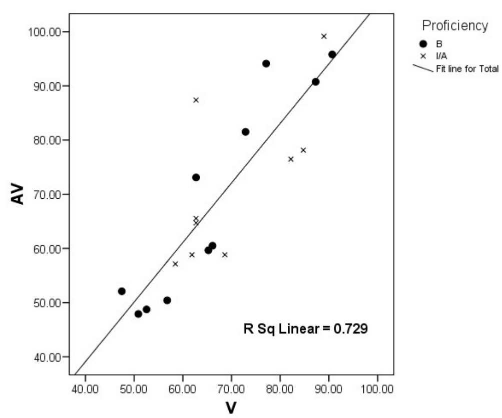
Figure 1: Box plot to show the proportion of correct responses (%) per condition



3.3. Correlations across conditions

Pearson's product-moment correlations were run to see whether identification using auditory cues was correlated with identification when the contrast was cued visually. An R^2 of 0.55 was obtained for the correlation between A and V scores, of 0.69 for the correlation between A and AV scores and of 0.73 for the correlation between V and AV scores. Performance by Colombian participants when only visual cues were presented was therefore strongly correlated with their AV scores. As shown in Figure 2, few participants obtained a higher score in the AV condition than they did when only visual cues were presented. These results demonstrate the participants' strong visual bias.

Figure 2 : Correlation between identification scores in the V and AV conditions for Colombian participants classified in terms of their language proficiency (beginner, Intermediate/Advanced)



3.4. Lexical effects

The word translation task was used to provide a measure of word knowledge per participant. Words that the participants did not know were treated as non-words without lexical representations, requiring bottom-up processing. Words known by the participants were assumed to have some form of lexical representation. By comparing identification scores for known and unknown tokens, it is possible to get a sense of the effect of word knowledge on identification for each participant. There was no significant difference in identification scores for 'known' words (M = 67.5%, SD 21.8) and 'unknown' words (M= 64.6%, SD = 14.56) although there was a trend for lower scores for unknown words. Interestingly, some of the unknown words yielded high scores, e.g. the mean identification score for *hobble* (meaning unknown to 100% of participants) was 82%.

3.5. Production scores

Production scores were obtained by calculating the percentage of /b/-/v/ productions for each participant that were correctly identified by the ten English listeners. Intermediate-Advanced speakers of English (M = 69.6%,) were significantly more accurate in their production than the Beginner group (M = 59.1%) ($F(1,18) = 8.27$, $p = .010$), although there was a great degree of individual variability in both proficiency groups (See Fig. 3). The effect of position was not statistically significant ($F(1,19) = 3.72$, $p = .07$, n.s.).

Figure 3: Box plot of mean production scores per word position and level of proficiency

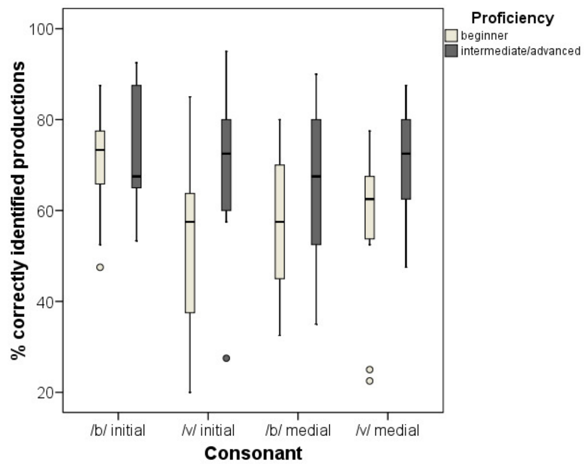
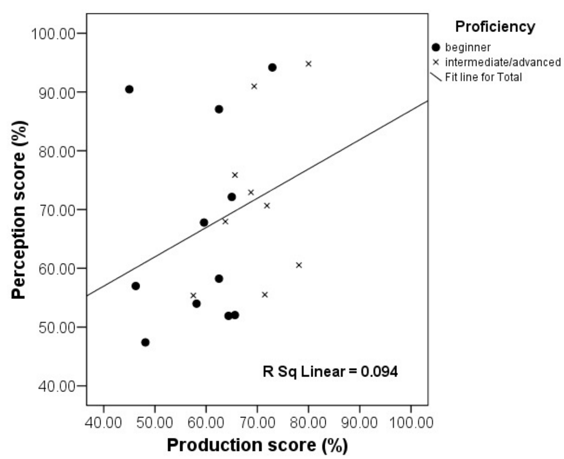


Figure 4 shows that production scores (i.e., how accurately the Colombians' productions of /b/-/v/ were perceived by native listeners) and their perception scores for the same 16 words were not significantly correlated.

Figure 4: Correlation between perception and production scores



3.6. Production and word knowledge results

Correlations between participants' word knowledge (mean percentage of correctly translated words) and their production accuracy was examined for word-initial and word-medial stimuli. There was a significant correlation between these two measures for /b/-/v/ in word-initial position ($r=0.479$; $p<0.05$): participants with better lexical knowledge of the words in the word-initial category tended to produce word-initial /b/-/v/ with greater accuracy.

4. DISCUSSION

This study found that visual cues were of great importance to the Colombians in their identification of the non-native /b/-/v/ contrast. Whereas previous studies with Peninsular Spanish speakers using the same contrast had found that there was no statistical difference between A and AV conditions, showing a greater weighting given to auditory cues (Hazan et al, 2006), our results with Colombian speakers showed that they did not show any difference between V and AV conditions, and that scores for the V and AV conditions were superior to scores in the A condition. As both Peninsular and Colombian speakers share the same L1 phoneme inventory, this result gives some support to the view that the weighting of visual cues in speech perception can be at least partly determined by cultural factors such as the tendency to use eye contact in speech communication. Some support for this view of increased eye contact in Colombian speakers comes from Norris's single case study (Norris, 2007) which states that "the differences lie in the modes of proxemics and gaze, for example, a member of the Colombian discourse system takes up a closer distance and engages in more direct eye contact than a member of the Hispanic discourse system". Cultural factors were suggested in previous studies (Sekiyama and Tokhura, 1993) as an explanation for the reduced weighting of visual cues in speakers of countries like Japan, where there is a culture of 'gaze avoidance'.

With regard to lexical effects, results showed that word knowledge was not a significant predictor in the identification of the non-native contrast but that 'known' lexical items tended to be more accurately produced than unknown words, at least in word-initial position. Therefore the ability to perceive a novel contrast is not always governed by lexical representations. In this study even the participants with low proficiency in English discriminated novel contrasts without referring to the lexicon, albeit inconsistently. Rather, perception seems to rely on a combination of factors namely the relationship between the L1 and L2 sound systems and an interaction between top-down and bottom-up processing mechanisms.

Finally, our study replicated previous findings (e.g., Bradlow et al., 1997) of weak correlations between the ability to perceive and produce a novel phonemic contrast. In both proficiency groups, there was evidence of speakers who could perceive the contrast accurately but could not produce it consistently, and of speakers with accurate productions but perception around chance level.

5. REFERENCES

Bradlow, A.R., Pisoni, D.B., Yamada, R.A. and Tohkura, Y. 1997. Training Japanese listeners to identify English /r/ and /l/ IV: Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, 101, 2299-2310.

Boersma P., Weenink D. 1997. Praat. Retrieved 12th January 2010 from University of Amsterdam. Website: http://www.fon.hum.uva.nl/praat/

Forster K, 2002. DMDX Display Software. Retrieved 18th June 2009 from University of Arizona, Department of Psychology. Website: http://www.u.arizona.edu/~jforster/dmdx.htm

Hayes-Harb R., 2007. Lexical and statistical evidence in the acquisition of second language phonemes. *Second Language Research* 23 (1), 65 – 94.

Hazan V., Sennema A., Faulkner A., Ortega-Llebaria M., Chung H. 2006. The use of visual cues in the perception of non-native consonant contrasts. *Journal of the Acoustical Society of America,* 119 (3), 1740 – 1751.

Hazan V., Sennema A., Iba M., Faulkner A., 2005. Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication,* 47 (3), 360 – 378.

Mora, J. C. 2005. Lexical knowledge effects on the discrimination of non-native phonemic contrasts in words and nonwords by Catalan/Spanish bilingual learners of English. *Proceedings of the ISCA Workshop on Plasticity in Speech Perception*, 43–46. London: University College London

Moreno A., Marino J.B. 1998. *Spanish Dialects: Phonetic Transcription. Proceedings ICSLP 1998*, paper 0598

Norris S. 2007. The micropolitics of personal national and ethnicity identity. *Discourse & Society,* 18 (5). 653 – 674.

Pallier C., Colome A., Sebastian-Galles N. 2001. The influence of native-language phonology on lexical access: exemplar-based vs. abstract lexical entries. *Psychological Science,* 12, 445–49

Sekiyama K. & Burnham D. 2008. Impact of language on development of auditory-visual speech perception. *Developmental Science,* 11 (2), 306 – 320.

Sekiyama K,, Burnham D., Tam, H., Erdener D. 2003. Auditory-visual speech perception development. Japanese and English speakers. *Proceedings AVSP 2003*, 43 – 47.

Sekiyama K, Tokhura, 1993. Inter-language differences in the influence of visual cues in speech perception. *Journal of Phonetics,* 21, 427 – 444.

A longitudinal study of speech rate and pauses in non-native Finnish

Minnaleena Toivola, Mietta Lennes, Jenni Korvala & Eija Aho

University of Helsinki
minna.toivola@helsinki.fi, mietta.lennes@helsinki.fi, jenni.korvala@helsinki.fi, eija.aho@helsinki.fi

ABSTRACT

Very few longitudinal studies exist on the temporal properties of second-language (L2) speech, and most studies have not investigated L2 speech at the earliest stages of learning. The aim of this follow-up study is to look at the potential developments in the temporal properties of L2 Finnish spoken by low-proficiency adult learners. Changes are reported in the articulation rate and the number and duration of pauses for the native speakers of Thai, Chinese, Russian and Vietnamese during the one-year period of observation.

Keywords: articulation rate, pause, second language, Finnish

1. INTRODUCTION

The temporal properties of speech are known to affect both fluency and the perceived strength of foreign accent, at least to some extent. These effects are clearest at the early stages of L2 learning (Derwing et al. 2004). However, relatively few studies deal with phonetic characteristics larger than a segment. In a cross-sectional study, Trofimovich and Baker (2006) examined fluency related temporal factors in immigrant speech. They found no effect of L2 experience on speech rate, and the only factor related to the learners' amount of L2 experience was stress timing. However, pause duration and speech rate contributed to foreign accent more than pause frequency, stress timing or peak alignment.

There are extremely few longitudinal studies of L2 acquisition (Ortega and Iberri-Shea 2005), although they provide a useful method for investigating the development of L2 phonetic features. A two-year study was conducted by Derwing et al. (2008) on the development of oral fluency and comprehensibility of two groups of adult immigrants (native speakers of Mandarin vs. Russian or Ukrainian). According to the foreign accent ratings made by native speakers of English, the fluency and comprehensibility of the speakers of Slavic languages improved somewhat, whereas the Mandarin speakers showed no progress. In addition, Derwing et al. (2009) used the informants of the previous study to measure the temporal factors of speech over the two-year period. Moreover, it has been observed that the perceived quality of L2 pronunciation may be affected by the task given to the speakers (Cucchiarini et al. 2002; Derwing et al. 2004; Gut 2007). For instance, read-aloud sentences and task-oriented conversation make somewhat different demands on the speaker and give away different properties in the L2 speech. There is a considerable lack of research on the speech of low-proficiency L2 speakers, since most studies have been conducted with immigrants in English-speaking countries. Since English is so widely used and heard, most of the subjects (often university students) may have begun their English studies well before immigration. Thus, they may no longer be considered as true beginners at the time of participating in the experiments.

So far, there have not been any experimental studies of the temporal aspects of non-native Finnish speech apart from the research conducted by Toivola et al. (2009). The present longitudinal study is focused on the articulation rate and the number and duration of pauses in read-aloud and conversational Finnish. In this study, the temporal aspects of low proficiency L2 Finnish speech are compared with those of native Finnish over a time period of 6–12 months.

2. MATERIAL AND METHODS

2.1. Speakers

For the present study, speech samples were recorded from nine adult immigrants with low proficiency of Finnish (native speakers of Thai, Chinese, Russian, and Vietnamese; age 23–38 years, median 29). The individual speakers will be referred to by codes, consisting of three letters for the speaker's native language (e.g., *tha* for a Thai speaker), one capital letter for gender (F = female, M = male), and a running index number. All the speakers were adult learners of Finnish with no prior experience of the language before immigrating to Finland (with the exception of *rusM4*, who had been studying Finnish for two months), although their length of residence in Finland varied slightly. By the time of the present study, two follow-up recordings had been completed at approximately six-month intervals from the first recording.

At the time of the first recording, the non-native speakers had been attending beginner-level Finnish language courses for the median duration of 8 months. They had completed either secondary school (younger speakers) or vocational school (elder speakers). According to an informal assessment by the experimenters, all speakers had distinct foreign accents. In addition, 18 native speakers of Finnish (11 female; age 20–47 years, $M = 25$) were recorded as a control group. All speakers were living in the capital city area of Finland. The non-native speakers had been living in Finland for the median duration of 12 months. Although most of the non-natives had immigrated fairly recently, two speakers (*thaF1* and *vieF1*) had already stayed in Finland for almost three years and *chiF2* for two years. However, the speakers *thaF1* and *chiF2* had been using English as their first language of communication before attending any Finnish lessons, as many Finns know English well and are eager to use it. Moreover, the non-natives (including *vieF1*) had quite often used their own mother tongue at the beginning of their stay.

In the first recording session, 10 participants were native speakers of Russian, 2 native Mandarin Chinese, 5 Thai and 3 Vietnamese. For the first follow-up recording (session 2), eight speakers reattended (4 Russian, 1 Chinese, 2 Thai and 1 Vietnamese). Only five speakers (2 Russian, 1 Chinese and 2 Thai, one of whom did not show up in session 2) attended the second follow-up recording (session 3), a year after the first session.

2.2. Recordings

The recordings were made in a sound-treated studio using high-quality audio equipment. Each speaker's voice was recorded with a head-mounted microphone at a sample rate of 44.1 kHz and sample size of 16 bits. The speakers were asked to read aloud a Finnish text of 125 words (13 sentences) extracted from a library brochure. They were instructed to read the text at their normal speaking rate. The same text was used in all recording sessions. The median duration of the read-aloud text was 1.12 minutes for the native speakers and 2.13–2.15 minutes for the non-natives. However, the recording of the read-aloud text did not succeed or take place in all three sessions for the speakers *thaF1*, *thaF3* and *rusF6*.

In order to be able to compare the speech produced for different tasks, picture-elicited dialogues were recorded from all speakers. The participants in each dialogue were provided with slightly different pictures, and they were instructed to find ten differences between the pictures. The speakers were sitting face to face at a table and they were not able to see each other's pictures. The non-natives were accompanied with either a native or a non-native speaker. The control material consisted of dialogues between two native speakers. The median duration of the dialogues was 2.00 minutes for the native speakers and 3.70–3.95 minutes for the non-natives. Each participant's voice was recorded on a separate channel. Each recording of the complete read-aloud text was saved as one sound file. For each dialogue, the two channels were separated and saved as two sound files of identical duration, each representing the speech of one speaker.

2.3. Analysis

All sound files were annotated with the Praat program (Boersma and Weenink 2010). Each utterance was first delineated and transcribed nearly orthographically, excluding punctuation. An utterance was defined as a stretch of speech during which the speaker was continuously articulating. Thus, utterances may contain hesitation sounds and incomplete words. For read-aloud speech, the corresponding written sentences were marked in a separate annotation tier. Individual words and syllables were then delineated and labeled in their separate annotation tiers, allowing for an accurate syllable count and the exclusion of incomplete words and fillers. Using a Praat script designed for this purpose, the duration of each pause and the articulation rate in each utterance (as syllables per second) were measured for all sound files in one go. The first and last pause in each sound file were excluded from the analysis. The articulation rate was calculated by dividing the

Figure 1: Pause duration within sentences in a read-aloud Finnish text for low-proficiency non-native speakers. The non-natives participated in three recording sessions at intervals of approximately six months. The horizontal line indicates the mean (0.229 s) of the median durations of sentence-internal pauses for 18 native Finnish speakers.

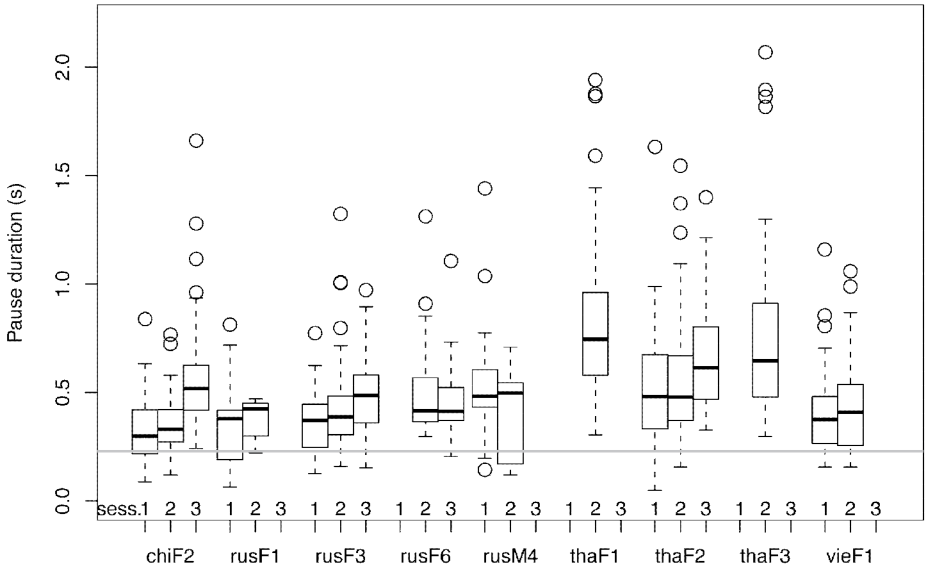
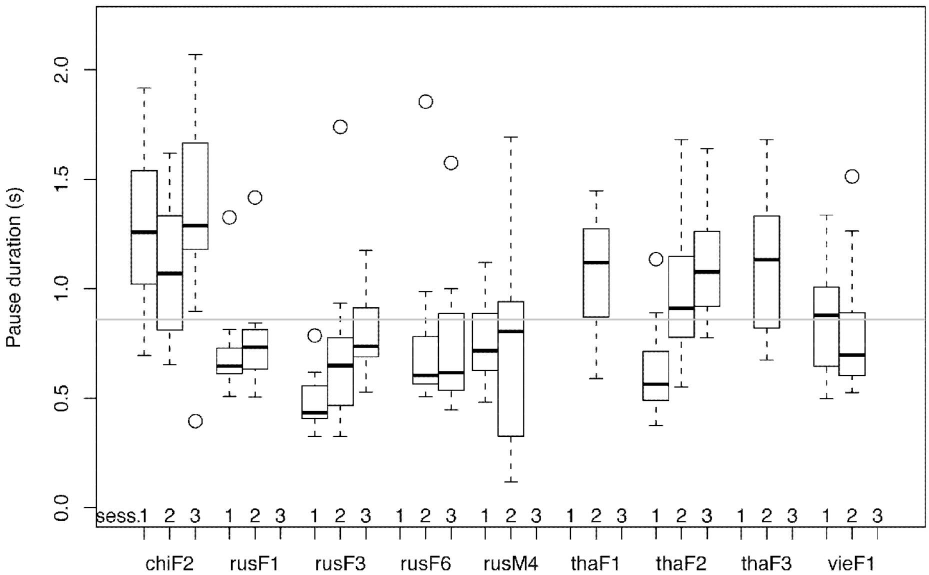


Figure 2: Pause duration between sentences in a read-aloud Finnish text for low-proficiency non-native speakers. The non-natives participated in three recording sessions at intervals of approximately six months. The horizontal line indicates the mean (0.860 s) of the median durations of the pauses between sentences for 18 native Finnish speakers.



number of completely produced syllables by either the total duration of the utterance (the basic articulation rate) or with the sum of the durations of the complete syllables, which will be referred to as the net articulation rate.

3. RESULTS AND DISCUSSION

3.1. Pauses

For the present study, pause durations were analysed for read-aloud speech only, since the pauses in a conversation are dependent on all the participants. The frequency of pauses was found to decrease during the period of observation. Within the same read-aloud text, the native Finnish speakers paused 1–21 times (median 4) within sentences, whereas the non-native speakers paused 21–66 times (*M* = 29) in session 1, 15–70 times (*M* = 49) in session 2 and 19–43 times (*M* = 27) in session 3. Disregarding a few exceptions, the number of pauses held between the 13 sentences in the read-aloud text was rather unanimously 12 for both native and non-native speakers in all sessions, excluding the initial and final silence in each sound file. As a general trend, the total number of pauses decreased during follow-up, although some speakers (e.g., *thaF2*) exhibited a temporary increase (cf. figure 3). However, as the reading material was identical for each recording session, it is possible that the speakers were simply learning to read that particular text.

In most cases, a slight increase was observed in the duration of pauses both within and between the sentences from one recording session to another (see figures 1 and 2). However, this overall trend was not significant for individual speakers. The median duration of all pauses between sentences was 0.860 s for the native Finnish speakers and, including all three recording sessions, 0.797 s for the non-native speakers. For sentence-internal pauses, the median duration was 0.229 s for the native Finnish speakers and 0.461 s for the non-native speakers (in the three sessions). Between sentences, native Finns tended to have longer pauses than the native speakers of Russian or Vietnamese, whereas the speakers of Chinese and Thai tended to exceed the pause duration of Finns. The extremely long pauses may be due to differences in the Finnish skills of the speakers. It has also been informally observed that Russian speakers are generally faster to move from one sentence to the next. As for pausing within sentences, the native Finnish speakers usually made much shorter and fewer pauses than the non-natives, and their pauses always occurred at clause boundaries or other natural breaks in the text (e.g., between the items of a list).

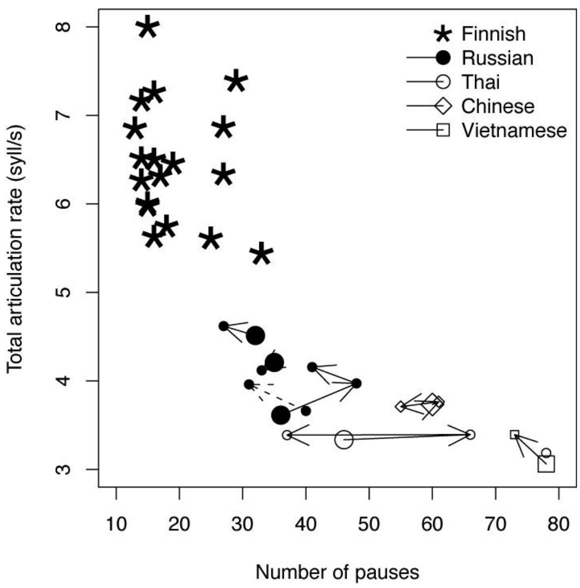
3.2. Articulation rate

With regard to the articulation rate, the native Finnish speakers were systematically faster than the non-native speakers in both read-aloud speech (see figure 3) and the picture-elicited dialogues (see figure 4). This result holds for both the basic articulation rate (including incomplete words and other “fillers”) and the net articulation rate (calculated from completely produced syllables only). The overall net articulation rates for each speaker are shown in table 1. With all speakers and recording sessions pooled together, the median of

Table 1: The total net articulation rate for all utterances in a read-aloud Finnish text and picture-elicited dialogues for the non-native speakers in three recording sessions.

SPEAKER	READ-ALOUD TEXT			PICTURE-ELICITED DIALOGUE		
	AR1	AR2	AR3	AR1	AR2	AR3
chiF2	3,73	3,76	3,71	3,82	3,68	4,10
rusF1	4,51	4,62		3,85	4,23	
rusF3	3,61	3,97	4,16	3,97	4,10	3,91
rusF6		3,66	3,96	3,90	4,13	4,21
rusM4	4,21	4,12		3,89	4,23	
thaF1		3,19		4,40	4,51	
thaF2	3,34	3,39	3,39	3,96	3,73	3,64
thaF3		2,75		3,39		3,15
vieF1	3,06	3,39		3,66	3,68	

Figure 3: Total number of pauses vs. total net articulation rate in a read-aloud Finnish text for native vs. non-native speakers.



the net articulation rate in read-aloud speech was 6.40 for the native Finns and 3.60 for the non-natives, and in picture-elicited dialogues for the natives 5.48 and 4.02 for the non-natives. However, for two speakers (*rusF1* and *rusF3*) the net articulation rate was greater for read-aloud speech than for the dialogues. The standard deviation of the net articulation rate in read-aloud speech was 0.87 s for the native Finns and 0.81 s for the non-native speakers, whereas in picture-elicited dialogues the standard deviation was 1.49 s for native Finns and 1.40 s for non-native speakers.

3.3. Length of runs

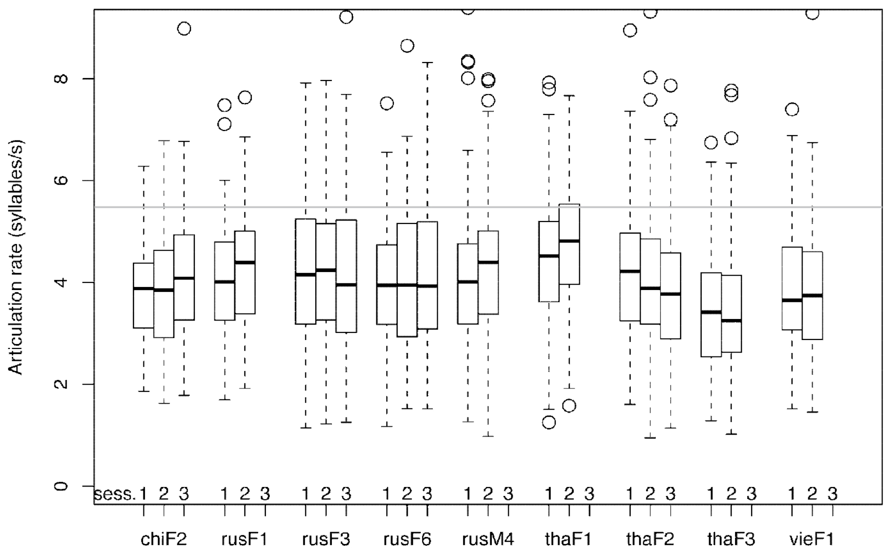
At a maximum, most natives produced 10–28 (median 17) complete words per utterance during their picture-elicited dialogues with other native speakers. The non-natives managed to produce maximum numbers of 4–17 (M = 7) full words per utterance in session 1, 5–10 (M = 7) words in session 2 and 6–12 (M = 8.5) words in session 3, even though filled pauses and hesitation sounds or incomplete words were allowed between the full words in each utterance. The native speakers produced median numbers of 2–5 (median of medians 3) complete words per utterance (in dialogues with other native speakers), whereas the non-native speakers only produced medians of 1–3 (session 1), 1–4 (session 2) and 1–2 (session 3) full words. Thus, very subtle improvements may have occurred in the maximal length of runs for the non-natives over the period of observation.

Within the same read-aloud text, each native speaker produced a mean of 0.3 incomplete words, whereas the non-native speakers produced a mean of 2.63 incomplete words.

4. CONCLUSIONS

Very little is known about the potential influence of the speaker's native language on the temporal features or other features related with foreign accent or perceived fluency in L2. Many researchers have either studied L2 speakers who share the same native language (e.g., Trofimovich and Baker 2006) or reported the pooled results for speakers of various different native languages (e.g., Gut 2007). According to the current results, however, it is very likely that many features in speech may vary greatly with regard to both the L1 and the individual speaker.

Figure 4. Net articulation rate for utterances in picture-elicited dialogues for low-proficiency non-native speakers in three recording sessions. Incomplete words, hesitation sounds and other articulated filled pauses are excluded. The horizontal line indicates the mean (5.48 syllables / s) of the medians of the corresponding articulation rates for 18 native Finns.



Several differences were observed between the temporal features of read-aloud and conversational speech. For Finnish, the grapheme-phoneme correspondence in written language is high, which may often help in learning to read Finnish. However, this feature may also be quite misleading unless it is properly explained to the non-native learner. On the other hand, there are differences between the standard written Finnish and the casual spoken language. Therefore, it is important to study conversational L2 speech in more detail. After all, the purpose of learning to speak a language is to be able to communicate and interact with others.

5. ACKNOWLEDGEMENTS

This study is part of the larger research project *ProoF – Pronunciation of Finnish by immigrants in Finland*, http://www.ling.helsinki.fi/ykiel/research/ProoF/, funded by the Academy of Finland (project number 122624).

6. REFERENCES

Boersma, P., Weenink, D. 2010. *Praat: doing phonetics by computer (Version 5.1.26)* [Computer program]. <http://www.praat.org/>. Last access 25 Feb 2010.

Cucchiarini, C., Strik, H., Boves, L. 2002. Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech. *Journal of the Acoustical Society of America.* 111(6), 2862–2873.

Derwing, T., Rossiter, M., Munro, M., Thomson, R. 2004. L2 fluency: Judgments on different tasks. *Language Learning* 54. 655–679.

Derwing, T., Munro, M., Thomson, R. 2008. A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistic* 29. 359–380.

Derwing, T., Munro, M., Thomson, R. ,Rossiter, M. 2009. The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition* 31(04). 533–557.

Gut, U. 2007. Foreign accent. In: C. Müller (ed.), *Speaker classification*. Berlin: Springer. 75–87.

Ortega, L., Iberri-Shea, G. 2005. Longitudinal research in second language acquisition: Recent trends and future directions. *Annual Review of Applied Linguistics* 25. 26–45.

Toivola, M., Lennes, M., Aho, E. 2009. Speech rate and pauses in non-native Finnish. *Proceedings of Interspeech 2009, Brighton, UK.* 1707-1710.

Trofimovich, P., Baker, W. 2006. Learning second language suprasegmentals: Effect of L2 Experience on Prosody and Fluency Characteristics of L2 speech. *Studies in Second Language Acquisition* 28. 1–30.

An acoustic analysis of L2 English speech rhythm among Greek children and adolescents

Eleni Tsiartsioni

Aristotle University of Thessaloniki, Greece
eltsiartsioni@hotmail.com

ABSTRACT

This study examines the production of features of the L2 English speech rhythm by Greek children and adolescents of three ages. Recordings were made twice, once before and once after a pronunciation teaching intervention, which focused, among others, on features of L2 English rhythm. The data were analysed acoustically with the use of the PVI measure, which expresses the degree of variability in vocalic and consonantal intervals. Durational measurements of vowel length and aspiration of stops were also conducted and the percentage of deleted schwa tokens was also estimated. The results indicate that the groups which received pronunciation instruction showed a change in the PVI values towards the target-like direction. The analysis of the duration of selected features revealed that the sounds that changed significantly were the sounds that pronunciation instruction focused on, i.e. the duration of the unstressed vowels and the duration of aspiration. A greater percentage of schwa deletion was observed after the teaching intervention than before it. The reduction in the duration of schwa appears to be related to the increase in vocPVI, whereas the increase in consPVI could be related to the increase of the duration of aspiration and the percentage of schwa deletion.

Keywords: FL phonological acquisition, speech rhythm, PVI, pronunciation teaching.

1. INTRODUCTION

This paper presents a preliminary interpretation of the results of part of the author's ongoing PhD research on the acquisition of features of English speech rhythm and stop voicing system by Greek learners of English. In particular, the present study examines the acquisition of features of the English speech rhythm by Greek learners of 10, 13 and 16 years old who learn English in a foreign language / formal setting context.

For many years pronunciation teaching has been largely ignored by FL teaching methodologies and research (Goodwin et al 1994, Pardo 2004). It is, therefore, not surprising that a review of studies on teachers' attitude to pronunciation teaching indicated that many teachers are unsure of the effectiveness of pronunciation for intelligibility and communication (Pardo 2004). A central question is therefore whether pronunciation can be effectively taught. Pardo's (2004) review of 25 studies on the effect of pronunciation instruction revealed that there is a positive effect of well-planned, quality pronunciation training and that use should be made of specific teaching techniques, since pronunciation is not simply 'picked-up'.

Whereas the vast majority of the studies reviewed by Pardo (2004) examined phonological acquisition that occurs in a naturalistic L2 setting, empirical research on pronunciation instruction which occurs in a formal EFL (English as a Foreign Langauge) setting also seems to confirm the positive effect of teaching pronunciation (Ekstrand 1982, Olson and Samuels 1982, Thogmartin 1982). Regarding the role of learners' age, contrary to what appears to happen in naturalistic L2 settings, where younger learners have an advantage over older learners in terms of pronunciation ability, research on phonological acquisition that takes place in a foreign language environment tentatively suggests that older students are better at acquiring target-language pronunciation than younger students (Ekstrand 1982, Fullana 2006, Thogmartin 1982). However, it has been shown that the differences between younger and older learners appear to minimise once the younger learners reach the same state of cognitive development as older learners (Muñoz 2003, 2006). This suggests that older learners can have an initial advantage in terms of rate of acquisition over younger ones but in the long run this difference is neutralized.

A number of studies have stressed the importance of suprasegmentals for communication and comprehensibility (for a review see Gong 2002: 26, 29). However, it has been equally stated that suprasegmentals and in particular FL rhythm may be among the most challenging pronunciation features to be learnt by speakers of a different language. Chela-Flores (1994: 235-236) claims that “the failure to make sufficient difference in length between the vowels in stressed and unstressed syllables seems to be the basic cause of difficulty among NNS of English”. It follows from this remark that when learning the English rhythm, native speakers of a ‘syllable-timed’ language have to learn (a) how to give sufficient stress to the main words and (b) how to reduce the unstressed syllables effectively. These suggestions have played an important role in setting the targets of the pronunciation teaching course and also in deciding on the content of instruction in the present study.

The effectiveness of pronunciation teaching focusing on rhythm has been tested by Chela-Flores (2001). The results of ten lessons showed an improvement in students’ perception and controlled production of the rhythmic patterns and also self-perceived improvement in detecting their own mistakes. L2 rhythm production has been investigated in a number of studies (see Grenon and White 2008 for a review). These studies showed that speakers tend to produce L2 rhythm with values intermediate between L1 and L2.

Comparing the characteristics of English and Greek speech rhythm, English is considered to be a prototypical example of ‘stress-timed’ languages (Abercombie 1967), with distinctive vowel length, unstressed syllables reduced both in length and quality, complex syllable structure and ‘heavy’ versus ‘light’ syllables, the former attracting stress while the latter being unstressed. On the other hand, although traditionally classified as a ‘syllable-timed language’ (Mackridge 1985), Greek has been found to display characteristics both of ‘syllable-timed’ languages, such as no alternation between strong and weak syllables and lack of distinctive vowel length, and also characteristics of ‘stress-timed’ ones, such as lexical stress and a clearly discernible ‘beat’ (Dauer 1983).

The present study approaches rhythm using the *Pairwise Variability Index* (PVI, Grabe and Low 2002). The PVI is an acoustic measurement which expresses the variability in successive vocalic and consonantal intervals and is based on the premise that the perception of rhythm classes results from differences in the variability of vocalic and consonantal intervals. According to this approach, it is expected that: a) in syllable-timed languages the duration of successive measurements is relatively similar, therefore there should be low variability indices (low vocalic PVI and low intervocalic PVI) and b) in stress-timed languages high variability indices should be computed reflecting complex syllable structure and reduced vowels.

Grabe and Low’s (2002) PVI scores for English confirm its classification as a stress-timed language. Regarding the PVI scores for Greek rhythm, Grabe and Low (2002) and Baltazani (2007) concluded that Greek rhythm is located somewhere between the rhythm of stress- and syllable-timed languages.

On the other hand, several studies have posed a methodological question on the interpretation of rhythm measures, such as the PVI (Ferjan et al 2008, Grenon and White 2008, Mok and Dellwo 2008); these studies showed that L2 speakers of ‘syllable-timed’ languages may produce L2 English rhythm with values similar to those of native speakers of English. According to the authors, however, this similarity with native values does not necessarily reflect native-like mastery of rhythm but instead may be attributed either to great speaker variability and to features unrelated to rhythm (Ferjan et al 2008), or to speaking rate and selective lengthening (Mok and Dellwo 2008). According to the critics, the rhythm measures do not give a full account of the factors that influence the durational variability in speech. For this reasons, apart from the PVI measure, the present study also examines the duration of selected segments and the percentage of schwa deletion (see the ‘Methodology’ section) which are believed to influence the PVI scores.

2. QUESTIONS OF THE STUDY

The present study aims at exploring the following questions:

1. Is pronunciation teaching effective in a foreign language setting?

2. Does age play a role in the acquisition of pronunciation in a classroom environment? For example, are younger students more favourably predisposed to acquiring FL pronunciation, as happens in naturalistic L2 settings?

3. METHODOLOGY

3.1. Subjects

In the present study two groups of subjects were recorded, an experimental group (n=36), which received pronunciation instruction, and a control group (n=36) which followed the regular English classes at school. Each group was subdivided into three subgroups comprising students of three different ages: 12 ten-year old students/ age group A, 12 thirteen-year old students/ age group B, 12 sixteen-year old students/ age group C. Data for L1 Greek and L1 English were also obtained.

3.2. Speech materials / recording procedure / data analysis

Recordings of students' speech samples were made twice, once before and once after the teaching intervention (Time 1 and Time 2 respectively). The speakers were asked to read an English text for the calculation of English rhythm and a Greek text for the calculation of Greek rhythm. The 10-year-old speakers read a text adapted from their English schoolbook, whereas the older students read 'The North Wind and the Sun' which is a standard text used for phonetic analyses of rhythm.

The recordings were analysed acoustically with the use of waveforms and digital spectrograms generated by the speech analysis software *PRAAT* (Boersma and Weenink 2007). For each recording we measured the duration of 125 vocalic and 125 intervocalic intervals, which gives us a total of 47,250 items measured for the present study. Subsequently, a PVI index was computed for the vocalic and the consonantal intervals for each speaker (vocPVI and consPVI) applying the formula proposed by Grabe and Low (2002). A normalized version of the PVI was employed both for the vocalic and the consonantal intervals.

In order to examine in depth the data of the experimental group, the duration of the unstressed vowel /ә/ and the stressed vowels /æ/, /ɐ/, /ɒ/, /ɑ:/ and /o/, as well as the duration of aspiration (VOT) for /p, t, k/ were measured before and after pronunciation teaching (aspiration was one of the features taught as part of the pronunciation teaching intervention and it was expected that it might be related to consonantal PVI scores). Also, the percentage of the deleted schwa tokens was estimated for each speaker, separately for the function and the content words.

3.3. Framework for pronunciation teaching

The framework of pronunciation teaching which is used in the present study is adopted from Celce-Murcia et al (1996), who propose five teaching stages for pronunciation teaching which move away from controlled to free activities (p. 36). These stages involve the following: (a) description and analysis of features, (b) listening discrimination activities, (c) controlled practice, (d) guided practice and (e) communicative practice.

The students of the experimental group received 50 lessons of pronunciation instruction on English stops and speech rhythm. Nineteen of these lessons were devoted to the teaching of aspects of the English rhythm. Each lesson lasted 10-15 minutes and all lessons were embedded in the regular English classes at school. The pronunciation lessons were taught by the researcher, who was also the main English teacher of the experimental classes.The lessons on rhythm focused on the teaching of the following aspects: word stress, sentence stress (identification of stressed / unstressed syllables and words, visual representation of stress patterns), content versus function words, reduced speech: reduced vowels and the schwa. The techniques used for pronunciation teaching involved the use of rubber bands, graphic representation, rhymes, poetry, jazz chants, all of which have been reported to be quite useful in the teaching of foreign language rhythm (see Gong 2002 for a review).

4. RESULTS

This section presents the results of the PVI analysis for L1 and L2 speech rhythm. Figure 1 illustrates the vocalic and consonantal PVI scores (vocPVI and cosnPVI respectively) for L1 Greek and L1 English for each age group. The asterisks above the bars show the cases where a statistically significant difference between the two languages is observed (according to independent sample t-tests). The results show that, as a general trend, English exhibits greater vocalic and consonantal variability than Greek, although not to a statistically significant extent for all age groups. The lack of statistically significant difference between L1 Greek and English for the vocPVI of ages B and C could perhaps be attributed to the great inter-speaker variability. The overall PVI profile of the two languages is in accordance with previous descriptions of English being more 'stress-timed' than Greek, although the lack of statistical difference in some cases may suggest that languages are not so neatly separated as previously believed.

Figure 1: PVI values for L1 Greek and L1 English.

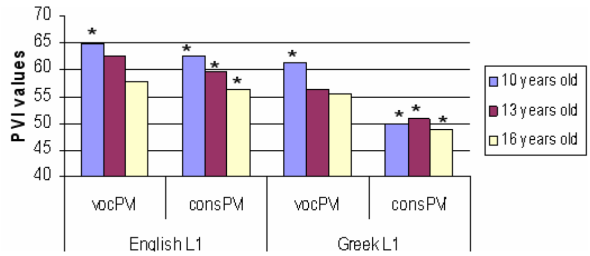
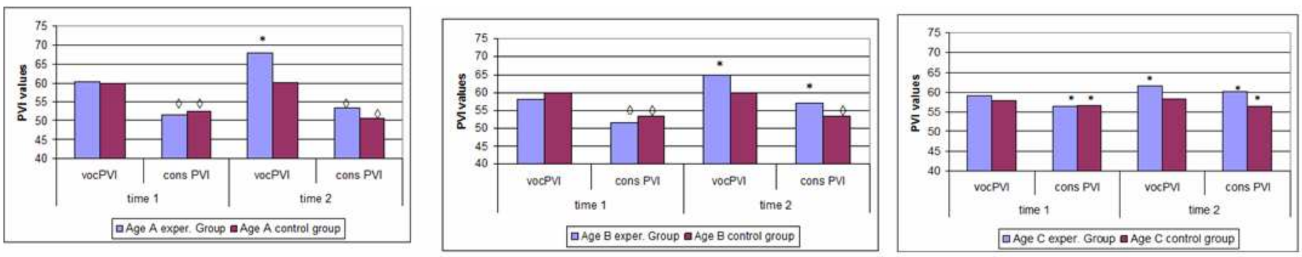


Figure 2 presents the PVI scores for L2 English produced by the experimental groups, which received pronunciation instruction, and the control groups of the three ages, which did not receive special pronunciation instruction. The asterisks above the bars show the cases where there is a statistically significant difference with L1 Greek, whereas the diamond-shaped marks show the cases where a statistically significant difference with L1 English is observed. The results indicate that before the teaching intervention there was no statistically significant difference between the control and the experimental groups of all ages (independent sample t-tests, p>0.05). No statistically significant difference is found in the PVI scores for the control groups between Time 1 and 2, which suggests that their performance did not improve over time. An increase in the PVI scores of the experimental groups was observed, although the difference was not always statistically significant. In particular a statistically significant difference between Time 1 and 2 was found for the vocPVI of ages A and B and the consPVI of age B.

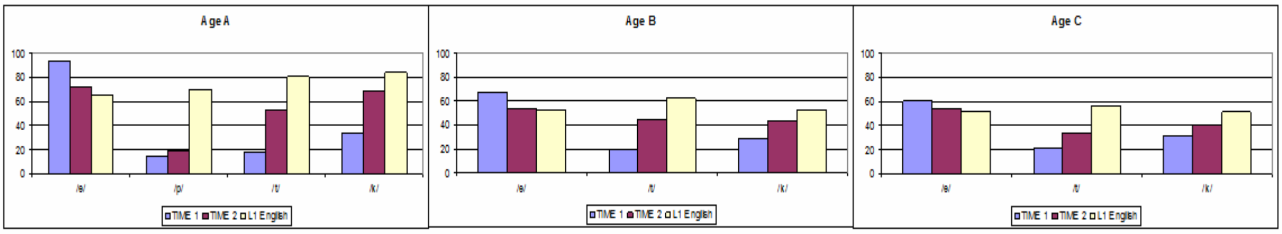
Figure 2: PVI values for L2 English, Ages A, B and C respectively.



Regarding the comparison of L2 data with L1 Greek and English, the vocPVI for all groups was not statistically different from L1 Greek and L1 English at Time 1. This lack of statistical difference could perhaps be related to the fact that the difference in the PVI scores of L1 Greek and L1 English was very small and for ages B and C the difference was not statistically significant (figure 1). The L2 consPVI for ages A and B was statistically different from English and no different from Greek at Time 1. For age C, however, the speakers of the control and the experimental group showed statistically different consPVI from L1 Greek and no different from L1 English at Time 1. This finding suggests that the speakers of this group were already mastering the English consonantal variability even before the pronunciation teaching intervention, perhaps due to their greater experience with the language than the other two age groups (all groups started English at the age of 9). After the teaching intervention, the vocPVI scores of all three experimental groups improved, as they became statistically different from L1 Greek and no different from L1 English. This suggests that students may have started implementing vowel reduction and elision as a result of the teaching intervention. The students' performance regarding the consPVI was not uniform across all experimental groups. In particular, at Time 2 for age A the consPVI was statistically different from English and no different from Greek; for age B an improvement was found, as the consPVI scores were statistically different from L1 Greek and no different from English at Time 2; for age C the consPVI was statistically different from Greek and no different from English, as at Time 1.

In order to further investigate the factors that have lead to a change in the PVI scores of the experimental group, we measured the duration of the unstressed vowel /ɘ/ and the stressed vowels /æ/, /ɐ/, /ɒ/, /ɑː/ and /o/, as well as the duration of aspiration for /p, t, k/ in the text read by the experimental group. The results revealed that in all age groups the same trend appeared: a statistically significant difference between Time 1 and Time 2 was found only for the sounds that pronunciation instruction focused on, namely for the duration of the unstressed vowel /ɘ/ and the duration of aspiration (paired-sample t-tests, $p<0.05$). The duration of /ɘ/ decreased and the duration of aspiration increased for the experimental groups of all ages. Figure 3 presents the duration of the features where a statistically significant difference was observed between Time 1 and 2 (the text read by age groups B and C contained no instances of word initial /p/ therefore no VOT data for this sound are provided).

Figure 3: The duration of /ɘ/ and of stops' aspiration (in ms) at Time1 and 2. L1 English values are also provided.



The comparison of the /ɘ/ duration between L1 and L2 English revealed that at Time 1 the L2 learners produced longer duration for /ɘ/ than the L1 English speakers to a statistically significant extent ($p<0.05$), however at Time 2 the difference with English was not statistically significant. This suggests that the learners managed to suppress the duration of the /ɘ/ as a result of the teaching intervention. Regarding the duration of stops' aspiration, the statistical analyses showed that there was a statistically significant difference between L1 and L2 English both at Time 1 and at Time 2. This finding indicates that L2 learners produced statistically longer VOT values after the teaching intervention than before it, but did not reach the L1 English target.

Table 1. The percentage of schwa deletion in the experimental groups. L1 English data are also given.

	Age A		**Age B**		**Age C**	
Type of words	Content	Function	Content	Function	Content	Function
Time 1	0%	0%	0%	1,1%	7,1%	1,9%
Time 2	0%	3,2%	10,9%	7,3%	16,7%	9,7%
L1 English	5,7%	3,1%	29,2%	4,5%	30,8%	8,4%

Table 1 shows the percentage of the deleted schwa tokens compared to the total number of schwa tokens in content and function words before and after pronunciation instruction. The results indicate that at Time 2 the percentage of schwa deletion increases compared to Time 1 for all groups. Group C produced a greater percentage of schwa deletion both at Time 1 and 2 than the younger age groups.

5. DISCUSSION

The present study examined the production of L2 English rhythm by Greek learners of different ages before and after pronunciation instruction. The first research question asks if pronunciation teaching is effective in a foreign language context. The analysis of the results shows that well-organized and planned pronunciation teaching can improve students production, even if it occurs in a foreign language environment with all the limitations that characterize it, for example lack of native English teacher, limited exposure and practice outside classroom, limited interaction in the target language in everyday life and limited amount of time devoted to English lessons at schools. Despite these limitations, an increase both in the vocPVI and consPVI of the experimental groups is observed, even though this increase was not always statistically significant. On the other hand, no statistically significant difference in the PVI scores is observed in the production of the control groups, which received no pronunciation instruction. The analysis of the vowel duration and the VOT for aspirated stops revealed that the features that changed significantly were the features that pronunciation instruction focused on, namely the duration of /ɘ/ and the increase in stops' aspiration. Since /ɘ/ was the only vowel whose duration changed significantly after the teaching intervention, whereas the duration of the other vowels examined did not change significantly, it can be assumed that the reduction of the duration of /ɘ/ could be related to the increase in the vocalic variability at Time 2. The increase in the duration of aspiration, as well as the increase in the percentage of schwa deletion leading to the creation of complex consonant clusters could have contributed to the increase in consonantal duration variability.

The second research question asks if age plays a role in pronunciation learning that occurs in a formal foreign language setting. The results of the present study show that the 13-year-old group exhibited greater improvement than the younger 10-year-old group. In particular, the 13-year-old group showed a statistically significant increase both in vocPVI and consPVI at Time 2, whereas the 10-year-old group showed a statistically significant increase in the vocPVI only. Also after pronunciation instruction the vocPVI and consPVI values of age group B were statistically different from L1 Greek and no different from L1 English, whereas for age group A the consPVI remained statistically different from L1 English and no different from Greek. The greater improvement in the 13-year-old speakers compared to that in the 10-year-old ones could be attributed to their greater maturity and cognitive development, which may be facilitative factors when learning a language in a formal foreign language context.

The comparison of the results of age groups A and B with age group C appears to be more difficult, since the 16-year-old speakers seem to have had a different starting point at Time 1 than the 10- and 13-year-old groups. In particular, the age group C appeared to be close to English at Time 1 as far as the consPVI is concerned, since the consPVI values before the teaching intervention were statistically different from L1 Greek and no different from L1 English. The percentage of /ɘ/ deletion was greater for this group, too, both at Time 1 and 2. It appears that the 16-year-old speakers were already mastering the English consonantal duration variability before the pronunciation teaching intervention, perhaps due to the greater experience they already had with the language than the younger age groups (all groups started at 9, so at the time of the experiment group C had been learning English for more years than the other two groups).

Concluding, this research has shown that instruction focusing on the rhythmic patterns of L2 English can have a positive effect on aspects of students' production of the FL rhythm. The study has shown that the13-year-old group showed a bigger improvement than the younger 10-year-old group, perhaps due to their cognitive development and maturation. However, it is difficult to compare these age groups with the older 16-year-old students, since the latter seem to have started from a different point, as they were close to the L1 English consPVI from Time 1.

6. REFERENCES

Abercombie D. 1967. *Elements of general phonetics.* Edinburgh: Edinburgh University Press.

Baltazani M. 2007. Prosodic rhythm and the status of vowel reduction in Greek. In: Selected Papers on Theoretical and Applied Linguistics from the 17th International Symposium of Theoretical and Applied Linguistics. Thessaloniki. 31-43.

Boersma P., Weenink, D. 2007. Praat. Doing phonetics by computer (version 5.0). <http://www.praat.org/>.

Celce-Murcia, M., Brinton, D., Goodwin, J. 1996. *Teaching pronunciation: A reference for teachers of English to speakers of other languages*. Cambridge: Cambridge University Press.

Chela-Flores, B. 1994. On the acquisition of English rhythm: theoretical and practical issues. *International Review of Applied Linguistics* 32. 232-242.

Chela-Flores, B. 2001. Optimising the teaching of English suprasegmentals. *Lengues Modernas* 28-29. 255-273.

Dauer, R.M. 1983. Stress–timing and syllable–timing reanalyzed. *Journal of Phonetics* 11. 51-62.

Ekstrand L.H. 1982. English without a book revisited: The effect of age on second language acquisition in a formal setting. In: Krashen, S., Scarcella, R., Long, M. H. (eds.), *Child adult differences in second language acquisition*. Roley, M.A.: Newbury House. 136-158.

Ferjan N., Ross, T., Arvaniti, A. 2008. L2 speech and rhythm metrics. Paper presented at ASA 2008, Paris.

Fullana N. 2006. The development of English (FL) perception and production skills: Starting age and exposure effects. In: Muñoz, C. (ed.), *Age and the rate of foreign language learning*. Clevedon: Multilingual Matters. 41-64.

Gong J. 2002. Introducing English rhythm in Chinese EFL classrooms: a literature review. *Post-Script* 3(1). 26-42.

Goodwin, J., Brinton, D., Celce-Murcia, M. 1994. Pronunciation assessment in the ESL/EFL curriculum. In: Morley, J. (ed.), *Pronunciation, pedagogy and theory: new views, new directions*. TESOL. 3-16.

Grabe E., Low, E. 2002. Duration variability in speech and the rhythm class hypothesis. In: *Papers in laboratory phonology* 7, Berlin: Mouton de Gruyter. 515-546.

Grenon, I., White, L. 2008. Acquiring rhythm: a comparison of L1 and L2 speakers of Canadian English and Japanese. *Proceedings of the 32nd Boston University conference on language development (BUCLD)*. 155-166.

Mackridge, P. 1985. *The modern Greek language.* Oxford, Oxford University Press.

Mok, P., Dellwo, V. 2008. Comparing native and non-native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English. *Proceedings of the 4th Speech Prosody 2008 Campinas Brazil*. 423-426.

Muñoz, C. 2003. Variation in oral skills development and age of onset. In: Garcia-Mayo, M. P, Garcia-Leccumberri, M. L. (eds.), *Age and the acquisition of English as a foreign language*. Clevedon: Multilingual Matters.

Muñoz C. 2006. The effects of age on foreign language learning: the BAF project. In C. Muñoz (ed.), *Age and the rate of foreign language learning.* Clevedon: Multilingual Matters. 1-39.

Olson, L.L., Samuels, S. J. 1982. The relationship between age and accuracy of foreign language pronunciation. In: Krashen, S., Scarcella, R., Long, M. H. (eds.), *Child adult differences in second language acquisition*. Roley, M.A.: Newbury House. 84-92.

Pardo, D.B. 2004. Can pronunciation be taught? A review of research and implications for teaching. *Revisita Alicantina de Estudios Inglesses* 17. 6-38.

Ross, T., Ferjan, N, Arvaniti, A. 2008. On quantifying speech rhythm. Poster presented at the Workshop on Empirical Approaches to Rhythm. March 2008, UCL, London.

Thogmartin, C. 1982. Age, individual differences in musical and verbal aptitude, and pronunciation achievement by elementary school children learning a foreign language. *International Review of Applied Linguistics* 20. 66-72.

Casual speech processes: L1 knowledge and L2 speech perception

Annelie Tuinman[1] *and Anne Cutler*[1,2]

[1]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[2]MARCS Auditory Laboratories, University of Western Sydney, Australia
Annelie.Tuinman@mpi.nl, Anne.Cutler@mpi.nl

ABSTRACT

Every language manifests casual speech processes, and hence every second language too. This study examined how listeners deal with second-language casual speech processes, as a function of the processes in their native language. We compared a match case, where a second-language process (/t/-reduction) is also operative in native speech, with a mismatch case, where a second-language process (/r/-insertion) is absent from native speech. In each case native and non-native listeners judged stimuli in which a given phoneme (in sentence context) varied along a continuum from absent to present. Second-language listeners in general mimicked native performance in the match case, but deviated significantly from native performance in the mismatch case. Together these results make it clear that the mapping from first to second language is as important in the interpretation of casual speech processes as in other dimensions of speech perception. Unfamiliar casual speech processes are difficult to adapt to in a second language. Casual speech processes that are already familiar from native speech, however, are easy to adapt to; indeed, our results even suggest that it is possible for subtle difference in their occurrence patterns across the two languages to be detected, and to be accommodated to in second-language listening.

Keywords: L1, L2, casual speech, reduction, insertion.

1. INTRODUCTION

The perception of speech in a second language (L2) cannot be understood except in the light of the perceiver's first-language (L1) knowledge; so much has been known to speech and language researchers for at least the better part of a century (see, e.g., Polivanov, 1931: "The phonological representations of our native language are so tightly coupled to our perception that even when we hear words or sentences from a language with quite different phonology, we tend to analyse these words in terms of the phonemic representations of the native language"). Experimental reports now provide abundant documentation of L1 effects on speech perception in L2, and models of the L1-to-L2 influence (e.g., Best & Tyler, 2007; Flege, 1995) offer detailed accounts of its genesis.

Like models of speech perception in general, however, models of L2 speech perception are understandably based on a somewhat idealised situation. The mapping of a phoneme or sequence of phonemes to stored representations can be predicted very well by the perceptual models, but the modelled situation will only arise if the input actually presents an acoustic form corresponding to each proposed segment. As listeners and speech researchers know only too well, however, real speech abounds with casual speech processes such as assimilation, reduction, deletion and intrusion, all of which lead to phonetic forms which deviate drastically from the canonical pronunciation of the words intended by the speaker.

In recent years, psycholinguistics has turned increasingly to investigation of real speech, and how listeners deal with the non-canonical forms it presents. A grossly over-simplified summary of the accrued results to date is that listeners are extremely good at exploiting the fine phonetic detail of utterances and identifying intended words even when casual speech processes have altered them from their canonical form, but that the alterations can often (temporarily) mislead listeners, and can often result in word recognition being harder than it would have been for the canonically pronounced versions. The fine differences between intended phonemes and phonemes resulting from a casual speech process have been shown to be exploited by listeners, for example in the case of place of articulation assimilation (e.g., to distinguish the /p/ of English *ripe* in *ripe berries* from the assimilated final phoneme of *right berries*; Gow, 2002), in neutralisation (e.g., to distinguish the final /p/ of Dutch *slip* from the devoiced final sound of *slib*; Warner, Jongman, Sereno & Kemps, 2004), and in liaison (e.g., to distinguish the word-initial /p/ in French *trop partisan* from the liaison realisation of a word-final /p/ in *trop artisan*; Spinelli, McQueen & Cutler, 2003). Listeners are successful at identifying word forms despite assimilation of place (Gaskell & Marslen-Wilson, 1996; Gow, 2001) or of voice (Snoeren, Segui, & Hallé, 2008) and despite reduction (Ernestus, Baayen, & Schreuder, 2002) or other non-canonical realisation (e.g., Alphen & McQueen, 2006; Sumner & Samuel, 2005). Despite all this success at dealing with real-speech forms, however, listeners are also often misled. Thus in a phoneme detection task they respond to phonemes which are not actually in the input at all, because they have been deleted in a casual pronunciation (Kemps, Ernestus, Schreuder, & Baayen, 2004), and they respond to phonemes which are accidentally there, such as a medial /p/ in a casual pronunciation of *something* (Warner & Weber, 2001). Their word recognition response times are slowed by many different types of casual-speech forms (Andruski, Blumstein, & Burton, 1994; LoCasto & Connine, 2002; Racine & Grosjean, 2000), and they can be seriously misled, at least temporarily, into assuming that a quite different word is being heard (Brouwer, Mitterer, & Ernestus, 2008).

All of this research has, of course, been carried out with L1 listeners. But if even these experienced listeners are misled, what is going to happen when L2 listeners hear the same sort of input? Hear it they will, because all languages manifest casual speech processes, and L2 listeners cannot permanently confine themselves to speech situations in which the input is as close to canonical perfection as it is in the classroom or on language tapes. In the current investigation we make a start on charting the perceptual effects of casual speech processes in L2 listening.

Interestingly for our purposes, there are some casual speech processes which are found in many languages, and some which are found in only few. Thus assimilation is widespread across languages, with assimilation of place being extremely common (indeed obligatory in some languages, such as Japanese), and assimilation of voice also quite frequent; but patterns such as the Hungarian assimilation of /lr/ sequences to /rr/ are much rarer (Mitterer, Csépe, & Blomert, 2006). Listeners are better at dealing with the processes that are more common in their L1; e.g., for English speakers, assimilation of place is easier to deal with in word recognition tasks than assimilation of voice, while for French speakers the reverse is the case (Darcy, Peperkamp, & Dupoux, 2007). It is therefore interesting to ask whether this advantage can be transferred to L2 input, i.e., whether listeners will also find it easy to deal with an L2 process if they already have experience with the same process in their L1.

One process that is found in many languages is /t/-reduction (Guy, 1980). In the Germanic languages English, German and Dutch, the process patterns very similarly. For instance, /t/ is highly likely to be deleted after /s/ or before a bilabial, so that most utterances of English *postman*, German *Postbeamter* 'postal worker' or Dutch *postbode* 'postman' are equally unlikely to contain much of a detectable trace of /t/. We compare this common process with a far less common process, namely the insertion of /r/ between words beginning and ending with vowels, in British English sequences such as *idea of*. This process is unknown in many other languages, for instance in Dutch. For the /t/-reduction case, we used Dutch speech presented to L1 speakers and to speakers of German with proficient L2 Dutch; since both languages have this reduction process, we refer to this as a case in which the L1 and L2 match on this feature. For the /r/-insertion case, we used British English speech presented to L1 listeners and to speakers of Dutch with proficient L2 English; since English has this process but Dutch does not, we refer to this case as one in which L1 and L2 mismatch.

The experiments which we carried out were broadly similar, in that in each case we constructed a phonetic continuum of stimuli, in which the phoneme under investigation (/t/, /r/) varied from effectively absent through partially realised to indubitably present. In each case, the listeners' task amounted to judging whether the phoneme was present or absent. In each case we compared the L2 listeners' judgements with the judgements given by L1 listeners presented with the same continuum of stimuli. In the match case, native speakers of German with proficient L2 Dutch and native speakers of Dutch decided in two perception experiments whether or not target words ended in /t/; the target words were verbs in the first experiment and nouns and adjectives in the second. In the mismatch case, native speakers of Dutch with proficient L2 English and native speakers of English decided whether target words were occurrences of *ice* or of *rice*.

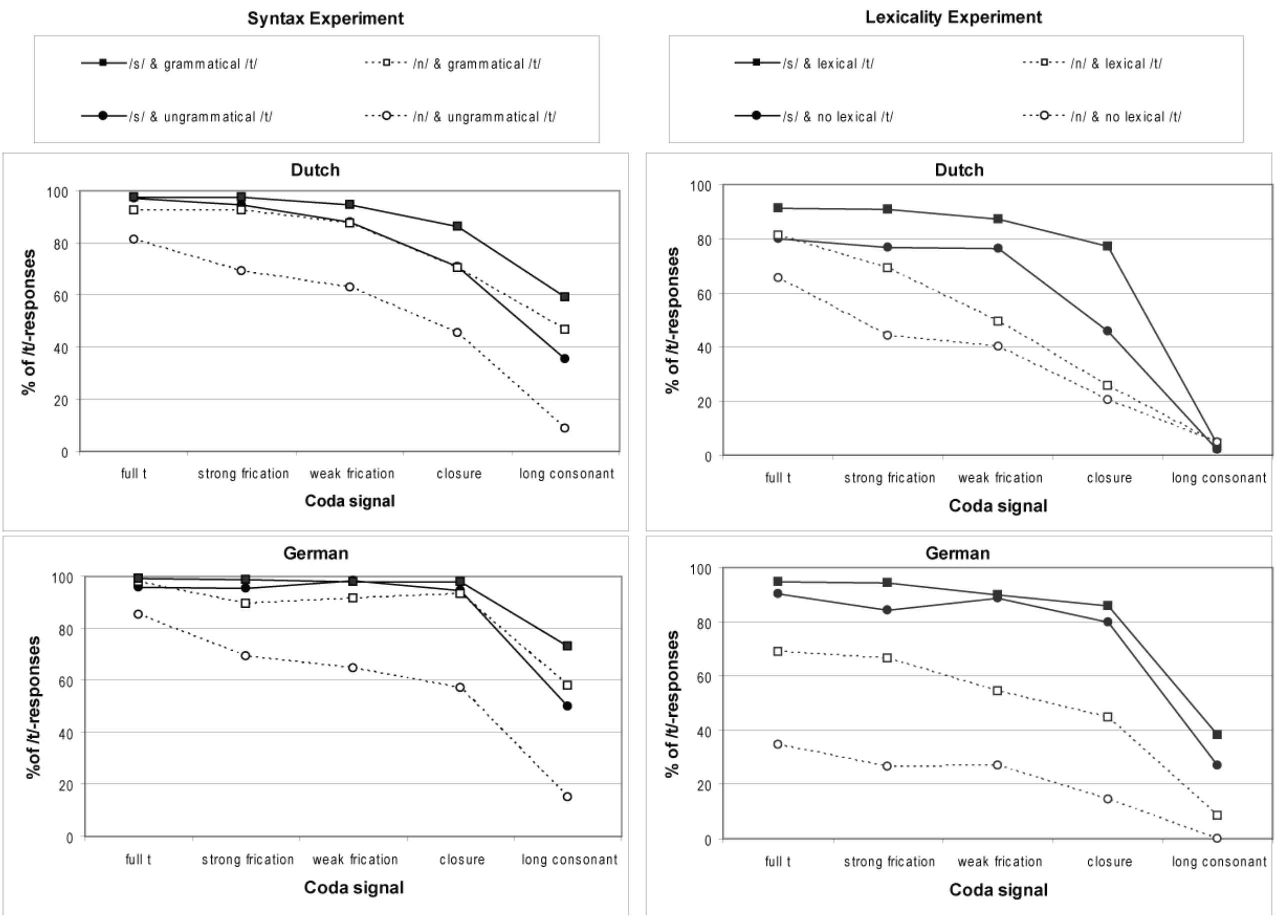
2. PERCEPTION OF REDUCED /t/

2.1. Method

Twenty-one native speakers of Dutch and 16 native speakers of German participated in the first experiment and 16 native speakers of Dutch and 16 native speakers of German in the second experiment. The German participants had a high level of proficiency in Dutch as L2.

Five realizations of /t/, from full production to complete deletion, were presented in two acoustic contexts, after /n/ (where /t/-reduction is unlikely) and after /s/ (where /t/-reduction occurs frequently). The selection of the /s/ and /n/ context and the different forms of reductions were based on patterns observed in a corpus study (Mitterer & Ernestus, 2006). In each sentence, listeners judged whether the target word ended in /t/ or not. In the Syntax Experiment, target words were verbs (e.g., *ren* ‘run’, *kus* ‘kiss’). This made it possible to use grammar (preceding *ik* ‘I’, *zij* ‘she’) to predict whether or not the ending should be /t/; the Dutch present tense third person singular inflection is /t/ while the first person inflection is null. In the Lexicality Experiment, target words were nouns and adjectives and lexical information produced the same result: interpreting /t/ made the target word a correct word (*charmant* 'charming') or not (*kanon*[t] 'gun').

Figure 1: Percentages of /t/-responses for Dutch listeners and German listeners



2.2. Results

The results for the Syntax Experiment with verbs and the Lexicality Experiment with nouns and adjectives were analyzed separately, in each case with a linear-mixed effects model. For the Syntax Experiment, participant was entered as a random factor, and Native Language, Coda Signal (from full production to complete deletion), Preceding Context (/n/ vs. /s/) and Grammar (/t/ or no /t/ predicted) as fixed factors. For the Lexicality Experiment, participant was entered as a random factor, and Native Language, Coda Signal, Preceding Context and Lexicality (/t/ or no /t/ predicted) as fixed factors. For the Native Language variable, the German speaker group was mapped on the intercept. Figure 1 shows the mean /t/-response percentages.

Analysis started with a full model and in stepwise fashion insignificant interactions were pruned. For the Syntax Experiment, there was an overall significant tendency for more /t/-responses by German participants. This tendency, however, was moderated by various interactions with the other experimental variables. To understand the nature of the interactions, we examined the effects of Native Language, Preceding Context and Grammar on all five levels of Coda Signal. This analysis showed that for the full /t/ and strong frication coda signal there were no overall difference between Dutch and German listeners, but for the full /t/ Germans had a larger effect of Grammar and a smaller effect of Preceding Context than Dutch listeners. For the weak frication coda signal the Germans gave more /t/-responses than the Dutch and the effect was enlarged if the Grammar predicts the presence of a /t/. Again, for the closure coda signal there was an overall effect of Native Language which was moderated by Grammar and Preceding Context, so that Germans gave more /t/-responses in all cases except when the Preceding Context was /n/ and the Grammar predicted no /t/. For the long consonant coda signal—when there is no hint of /t/, but actually an extra long /n/ or /s/—German overall gave more /t/-responses than the Dutch.

The Lexicality Experiment showed main effects of Preceding Context and Lexicality; more /t/-responses were given after /s/ than after /n/, and more /t/ responses were given if an existing word resulted. Overall, the effect of Preceding Context was larger for the L2 listeners. However, the reported effects were moderated by various interactions. Therefore, we examined the effects of Native Language, Preceding Context and Lexicality on all five levels of Coda Signal. The strong Coda Signals showed a consistent pattern with main effects for Preceding Context and Lexicality and a significant interaction between Native Language and Preceding Context—L2 listeners had a larger context effect than L1. For the weak Coda Signals the effect of Lexicality was larger for L2 listeners, but depended on the Preceding Context.

In summary, the German L2 listeners' responses are, overall, sensitive to the same factors as the Dutch L1 responses; however, in some cases the L2 responses include significantly more /t/ reports than the L1 baseline. A comparison with the production facts suggests an explanation for this pattern. In German, reduction of morphologically functional /t/ (such as a verb inflection) is inhibited, whereas in Dutch, reduction is equally likely for morphological and for non-morphological /t/. Separate production experiments with non-Dutch-speaking Germans and Dutch native speakers confirmed this pattern (Tuinman, 2006). In the light of this comparison, the tendency of the L2 listeners to produce more /t/ responses, especially when Grammar predicted a /t/, suggests that they were sensitive to the L1-L2 difference, and tried to compensate for it.

3. PERCEPTION OF INTRUSIVE /r/

3.1. Method

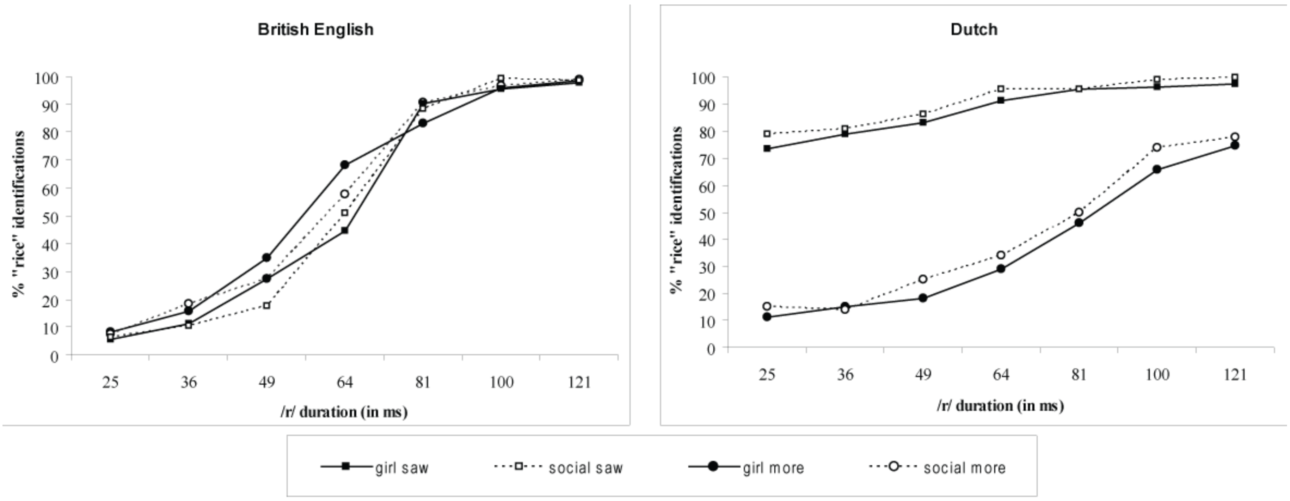
Eighteen native speakers of Dutch and 18 native speakers of British English took part in the experiment. The Dutch participants had a high level of proficiency in English as a second language. On average, they had received 7 years of English instruction in primary and secondary education.

Seven realizations of /r/, from short (25 ms) to long (121 ms), were presented in four sentence contexts with a contextual bias and an orthographic bias for the perception of /r/. In each sentence, listeners judged the critical word *(r)ice* and had to decide whether they heard *ice* or *rice*. Sentences with the context *the social worker* and *given to the poor* were intended to have a semantic bias towards *rice* rather than *ice*, while sentences with *the little girl* and *given to her brother* were assumed to be less biased towards *rice* and more biased towards *ice*. The orthographic bias was established by the words *saw* and *more* preceding the target word *(r)ice*. As the phrase *more ice* includes an /r/ in the spelling, the perception of an /r/ in the speech signal can be attributed to *more*, while in the case of *saw ic*e, perception of an /r/ sound cannot be mapped to the spelling; the bias should therefore manifest itself in terms of more reports of *rice* after *saw* than after *more*.

3.2. Results

Figure 2 shows the percentage of rice judgments as a function of the three independent variables, separately for the L2 and L1 listeners. The principal result is clear at a glance: the response patterns are different from each other. British English L1 listeners based their responses mainly on the durational variation rather than on the other variables, while the L2 listeners barely used the durational information, but show a clear separation between the responses in sentences with saw (squares) versus more (circles).

Figure 2: Percentages of "rice" identifications for British English and Dutch listeners



The results were analyzed with a linear-mixed effects model, with Participant as a random factor and Native Language, Orthography, Context, and Duration as fixed factors. Duration was entered as a numerical variable, centered around zero, and Orthography and Context were contrast-coded (-/r/ bias = -0.5, +/r/ bias = 0.5). For Native Language, the L1 group was mapped on the intercept. Analysis started with a full model and in stepwise fashion, insignificant interactions were pruned. The L1 group had a significant effect of /r/ Duration, no effect of Context, and a small negative effect of Orthography. That is, L1 listeners in fact gave somewhat more *rice* responses when the preceding word was *more* than when the preceding word was *saw*. The L2 listeners showed a significant effect of Orthography, and gave more *rice* responses when the preceding word was *saw* than when the preceding word was *more*. The L2 listeners were also less categorical in their responses than the L1 listeners. Additionally, the L2 group was influenced by Context, and made less use of /r/ Duration when there was an Orthographic Bias.

4. CONCLUSION

The combined results of our two sets of studies motivate the inescapable conclusion that casual speech processes are subject to the same tight coupling with L1 listening experience as every other aspect of L2 speech perception. The L2 listeners' performance in the match case (/t/-reduction experiment) related quite differently to the native performance baseline than the performance of the L2 listeners in the mismatch case (/r/-intrusion experiment). In the match case, the German listeners were broadly sensitive to the same range of factors in the Dutch input as the Dutch L1 listeners: they were in general more likely to report the presence of a /t/ when it followed /s/ rather than /n/, when it formed a grammatical string, and when it made a real word. In the mismatch case, however, the Dutch listeners showed quite a different pattern of sensitivity than the English L1 listeners: while the L1 listeners based their responses overall on the acoustic characteristics of the stimuli, were quite insensitive to the sentence meaning and certainly did not incline to report /r/ when there was none in the orthographic representation, the L2 listeners mimicked none of these patterns. They made relatively little use of the acoustic information, they were significantly influenced by the sentence meaning, and they were far more likely to report /r/ when the orthography contained none.

We conclude, therefore, that the casual speech patterns of the L1 influence interpretation of casual speech processes in an L2. A process familiar from the L1 can be processed easily; an unfamiliar L2 process with no corresponding L1 experience is very hard to adapt to.

However, this is not the end of the story, because in some respects the L2 listeners did deviate from L1 performance in the match case as well. Especially they were more likely than the L1 listeners to report a /t/ when the sentence grammar predicted it, i.e. when it was a morphological inflection. This is precisely the kind of /t/ that is less likely to be reduced in their L1. This pattern therefore raises the intriguing possibility, certainly worthy of future investigation, that listeners not only can deal easily with a casual speech process of the L2 that matches an L1 process, but they are even sensitive to subtle differences in its distribution of occurrence across the languages, and can accommodate their perceptual responses accordingly.

5. ACKNOWLEDGEMENTS

This research was supported by a doctoral grant from the Max Planck Society and by a Fulbright travelling fellowship, both to the first author. We are grateful to Holger Mitterer for advice on all aspects of the project.

6. REFERENCES

Alphen, P. M. v., McQueen, J. M. 2006. The effect of voice onset time differences on lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 32, 187-196.

Andruski, J. E., Blumstein, S. E., Burton, M. 1994. The effect of subphonetic differences on lexical access. *Cognition*, 52, 163-187.

Best, C. T., Tyler, M. D. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In: Munro, M., Bohn, O.-S. (eds.), *Second language speech learning Amsterdam*: John Benjamins.

Brouwer, S., Mitterer, H., Ernestus, M. 2008. How acoustically reduced forms activate the lexicon: Evidence from eye-tracking. Abstracts of *Laboratory Phonology 11*. Wellington, New Zealand.

Darcy, I., Peperkamp, S., Dupoux, E. 2007. Bilinguals play by the rules: Perceptual compensation for assimilation in late L2-learners. In: Cole, J. , Hualde, J. (eds.), *Papers in Laboratory Phonology 9*. Berlin: Mouton de Gruyter. 411-442.

Ernestus, M., Baayen, H., Schreuder, R. 2002. The recognition of reduced word forms. *Brain and Language*, 81, 162-173.

Flege, J. E. 1995. Second language speech learning: Theory, findings, and problems. In: Strange, W. (ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Baltimore: York Press. 233-277.

Gaskell, M. G., Marslen-Wilson, W. D. 1996. Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 144-158.

Gow, D. W. 2001. Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language*, 45, 133-159.

Gow, D. W. 2002. Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance*, 28, 163-179.

Guy, G. R. 1980. Variation in the group and the individual. In: Labov, W. (ed.), *Locating language in time and space*. New York: Academic Press. 1-36

Kemps, R. J., Ernestus, M., Schreuder, R., Baayen, H. 2004. Processing reduced word forms: The suffix restoration effect. *Brain and Language*, 90, 117-127.

LoCasto, P., Connine, C. M. 2002. Rule-governed missing information in spoken word recognition: Schwa vowel deletion. *Perception and Psychophysics*, 64, 208-219.

Mitterer, H., Csépe, V., Blomert, L. 2006. The role of perceptual integration in the perception of assimilation word forms. *Quarterly Journal of Experimental Psychology*, 59, 1395-1424.

Mitterer, H., Ernestus, M. 2006. Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch. *Journal of Phonetics*, 34, 73-103.

Polivanov, E. 1931. La perception des sons d'une langue étrangère. *Travaux du Cercle Linguistique de Prague*, 4, 79-96.

Racine, I., Grosjean, F. 2000. Influence de l'effacement du schwa sur la reconnaissance des mots en parole continue. *L'Année Psychologique*, 100, 393-417.

Snoeren, N., Segui, J., Hallé, P. 2008. Perceptual processing of partially and fully assimilated words in French. *Journal of Experimental Psychology: Human Perception and Performance*, 34, 193-204.

Spinelli, E., McQueen, J. M., Cutler, A. 2003. Processing resyllabified words in French. *Journal of Memory and Language*, 48, 233-254.

Sumner, M., Samuel, A. G. 2005. Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language*, 52, 322-338.

Tuinman, A. 2006. Overcompensation for /t/-reduction in Dutch by German-Dutch bilinguals, Abstracts of *Laboratory Phonology 10*. Paris, France.

Warner, N., Jongman, A., Sereno, J., Kemps, R. J. 2004. Incomplete neutralization and other sub-phonemic durational differences in production and perception of Dutch. *Journal of Phonetics*, 32, 251-276.

Warner, N., Weber, A. 2001. Perception of epenthetic stops. *Journal of Phonetics*, 29, 53-87.

A neurophysiological investigation of processing phoneme substitutions in L2

Merel van Rees Vellinga[1], *Adriana Hanulíková*[1], *Andrea Weber*[1] *and Pienie Zwitserlood*[2]

[1]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[2]Westfälische Wilhelms-Universität Münster, Münster, Germany

{merel.vanreesvellinga; adriana.hanulikova; andrea.weber}@mpi.nl, zwitser@uni-muenster.de

ABSTRACT

The present electrophysiological study examined whether linguistic experience with pronunciation variants in non-native speech (i.e., th-substitutions in English) influences processing at the prelexical level. In a mismatch negativity (MMN) study, the English pseudoword /θond/ was presented along with the deviants /tond/ and /sond/, while ERP-data were collected from Dutch listeners. If experience influences the MMN, smaller amplitude deviances should be observed for /tond/, the deviant typical for Dutch learners of English, than for the less typical deviant /sond/. If, on the other hand, the MMN mainly reflects perceptual similarity, smaller amplitude deviances should be found for the perceptually similar /sond/ deviant. The results of Dutch listeners revealed a significant deviance interaction, with /sond/ eliciting a smaller MMN and longer latencies than /tond/. This suggests that perceptual similarity rather than linguistic experience influences prelexical processing of th-substitutions for Dutch L2 listeners.

Keywords: MMN, th-substitutions, Dutch, L2.

1. INTRODUCTION AND BACKGROUND

An experience-based account of processing pronunciation variants assumes that a frequent variant will be processed differently from an infrequent one. The present study takes variant forms that vary in production frequency in foreign-accented speech as the starting-point for investigating the prelexical processing of such L2 mispronunciations. An answer is sought to the question of whether preferred variants of Dutch learners of English lead to different neurophysiological responses in Dutch listeners than disfavoured variants. The case is made with the English interdental fricative [θ], a sound that poses great difficulties for many learners of English. In particular, Dutch speakers often substitute [θ] with [t] (e.g., they pronounce *theft* as *teft*), and to a lesser extent also with [s] (Wester et al. 2007; Hanulíková & Weber 2010).

Recently, Hanulíková and Weber (in prep.) showed in an eye-tracking study that spoken-word recognition by L2 listeners is influenced by how frequent particular th-substitutions occur in the listeners' own L2 speech. In this study, Dutch and German listeners heard L2 English words in which word-initial [θ] was substituted (e.g., *theft* pronounced as *teft* and *seft*). Dutch and German L2 speakers vary with respect to the frequency with which they produce these substitutions. While hearing the auditory probe, participants looked at a computer screen displaying four printed words: the English th-word (e.g., *theft*), a phonological competitor (e.g., *left*), and two unrelated distracters (e.g., *kiss*, *mask*). The amount of looks to printed words in this task is assumed to reflect the strength of lexical activation during spoken-word recognition (e.g. Allopenna, Magnuson & Tanenhaus 1998). It was found that L2 listeners looked most often at the th-words when the auditory prime represented the accent-specific predominant substitute. That is, Dutch listeners looked more at *theft* when hearing *teft* than when hearing *seft*, but German listeners for whom the predominant substitute is [s] (Hancin-Bhatt 1994; Hanulíková & Weber 2010) looked more often at *theft* when hearing *seft* than when hearing *teft*. This is evidence that language-specific experience with mispronunciations influences lexical processing. Listeners recognise L2 words more easily when they are pronounced in the way that they are familiar with from their own accent, which is in line with frequency-based accounts of phonological variation (Connine 2004; Mitterer & Ernestus 2006).

The present study seeks to find neurophysiological evidence for the question of whether experience with mispronunciations, in the form of segmental substitutions in L2 speech, already influences the prelexical level of processing. A well-established method to investigate experience-based auditory memory traces is the

Mismatch Negativity (MMN) paradigm. With respect to language-specific features, the MMN, an Event-Related Potential (ERP) component discovered by Näätänen et al. (1978), so far has been used to study differences in neural representations for within- and across-category stimuli in L1 speech. The MMN is an early, automatic brain response that becomes visible as a negative component of an ERP upon the detection of a deviant feature in a stimulus. Relating sounds to meaning involves mapping sounds to phonological categories, and this process is strongly influenced by experience-based phonetic representations. Language-specific experience affects the processing of speech sounds, which has been shown in EEG research before (e.g. Näätänen 1997; Jacobsen 2003; Kirmse et al. 2008; Dehaene-Lambertz 1997; Brunellière et al. 2009). The specific aim of this study was to test whether experience with accent-specific th-mispronunciations in English influences the activation of memory traces in an MMN design. More specifically, we wanted to examine whether the presentation of the predominant th-substitution [t] results in smaller MMNs or longer MMN latencies for Dutch listeners than the presentation of the less frequent substitution [s]. Dutch participants were presented with an English pseudoword with initial [θ] ([thond]) as the standard condition, and with two pseudowords starting with either [t] or [s] ([tond] and [sond]) as the deviant conditions. If familiarity with th-substitutions influences prelexical processing, the less frequent mispronunciation [sond] should result in a larger MMN amplitude and/or shorter MMN latencies, while the more frequent variant [tond] should result in a smaller MMN amplitude and/or longer MMN latencies. If familiarity does not influence prelexical processing, then we predict that the deviant with [s], which is perceptually more similar to [θ], should result in a smaller MMN than the deviant with [t], which is perceptually more distinct from [θ] (Cutler et al. 2004).

2. RESEARCH DESIGN AND MATERIALS

Sixteen native listeners of Dutch participated in this study (mean age of 23, SD 3.3, nine females), all right-handers, with good hearing, with no speech or language problems, and with a good command of English.

The standard stimulus was the English monosyllabic pseudoword ‘thond’ [θond], and the two deviants were ‘tond’ [tond] and ‘sond’ [sond]. The stimuli were recorded by a native speaker of American English. Using Praat (Boersma & Weenink 2002), the stimuli were cross- and identity-spliced to avoid the elicitation of MMNs owing to small differences in features other than the beginning phoneme in the recorded materials. Variation in the stimuli was added by the creation of pitch changes of +6, +12, -6 and -12 Hz. These pitch variations made it possible to abstract away from the specific acoustic properties of one token (see Bien et al. 2009). Application of the pitch variations resulted in five tokens per pseudoword, thereby creating fifteen different tokens. The length of the stimuli was 593 ms for [θond], 499 ms for [tond] and 609 ms for [sond].

The EEG recordings were made with the actiCAP system (Brain Products GmbH, Munich, Germany). Impedances were kept below 20 kΩ throughout the recordings, and a total of 38 electrodes were placed on standard electrode sites. The electrodes were referenced to the right mastoid. In addition, one electrode was placed underneath the left eye to monitor the vertical oculogram. The horizontal oculograms were monitored by electrodes F9 and F10. EEG signals were amplified with BrainAmp DC amplifiers.

2.1. Procedure

Participants were equipped with a 38-electrode cap on a common connector, and were placed in front of a computer screen in a soundproof, electrically shielded EEG-booth. The auditory stimuli were presented over speakers at approximately 60 dB. Participants were instructed to watch a self-selected silent film while the stimuli were being presented, and were instructed to ignore the auditory stimuli. Stimuli were presented in two different blocks. The first block was a multideviant block, wherein the standard [θond] was presented for 80% and the deviants [tond] and [sond] each for 10% of the time. The stimuli were presented in a random order, with the criteria that a block started with at least eight standards, and that deviants were always presented with at least two standards in between. Stimuli were presented at 1000 ms inter-stimulus intervals (ISIs). The multideviant block consisted of 1000 stimuli and lasted for 26 minutes.

After a short break, the second block with an equiprobable design started. This block, in which [θond], [tond], and [sond] were each presented for 33.3 % of the time, functioned as a baseline for comparison of the standard and deviants from the first block. It furthermore served as a control for the alignment of [t] with [θ] and [s]. When producing the plosive [t], the airflow is first stopped before it is released with a burst. Since it is impossible to measure the closure phase for isolated words, the MMN latencies for [t] in the equiprobable block would allow a control for MMN latencies in the multideviant block. The equiprobable block consisted of 600 stimuli, with each stimulus randomly presented 200 times, and lasted for 19 minutes.

After the EEG recordings, participants completed a discrimination ABX-task, that was set up to confirm that participants could perceptually discriminate the three stimuli [θond], [tond] and [sond] in an offline task. On twelve trials, the three stimuli were presented in random order at the A and B positions, with the X matching either the A or B stimulus on each trial. The participants' task was to match the X position to one of the two preceding stimuli. On average, participants responded on 84% of the trials correctly.

3. RESULTS

3.1. EEG-Recordings

The analyses were carried out with Advanced Source Analyses (ASA) software (Advanced Neuro Technology, Enschede, The Netherlands). The raw data were first re-referenced to the right mastoid (M2). To monitor the vertical and horizontal oculograms (VEOG and HEOG), bipolar channels were then added by subtracting Fp1 from V1 for VEOG monitoring and by subtracting F10 from F9 for HEOG monitoring. The data were then 35 Hz low pass filtered with a slope of 12 dB/oct. Artefacts were excluded for the -75 to 75 μV range. For all epochs, a baseline correction of -250 ms was employed. Segmentation took place for all epochs of interest, for a time window of -250 to 550 ms. The averages of the two blocks and three conditions were first computed for each participant, and the signal of each participant was visually inspected to detect possible bad electrodes or other abnormalities. The grand averages (GAs) of all conditions were computed and, for the analysis of the first block, the GA elicited by the standard [θond] was subtracted from those of the deviants [tond] and [sond]. For the second block, the procedure was the same, but since [θond], [tond] and [sond] were presented equally often, they were all considered as standard here.

Figure 1a shows for the multideviant block the time course of the topography of the difference in brain responses from 31 until 415 ms after onset, with each frame corresponding to a time window of 32 ms. The top row shows the topography of the deviant [sond] minus the standard [θond], the bottom row depicts the deviant [tond] minus the standard [θond]. The accordant topography for the equiprobable block can be found in Figure 1b. As can be seen in Figures 1a and b, both deviants [tond] and [sond] showed an early negativity, followed by a strong positivity, which again changes into a stronger negativity. This resembles the transition from N1 to P2 and to MMN, as also becomes visible in Figure 2. Furthermore, a difference in latency can be seen in Figures 1a and b: the responses for [tond] always occurred about one or two frames earlier than those for [sond], indicating a difference of at least 32 ms. The maps in Figure 1c show the difference waves for identical stimuli, of [sond] in the multideviant block minus the same stimulus [sond] in the equiporable block in the top row, and the difference of [tond] in the multideviant block minus [tond] in the equiprobable block in the bottom row. Time windows are chosen from 102 to 486 ms, with an interval of 32 ms. Again, in Figure 1c a latency difference can be seen as main effect. The latency effect, that is always earlier for [tond] than for [sond], suggests that it is not experience that is being measured here. Rather, it seems that acoustic similarity between phonemes determines processing.

For all conditions, peak scores were extracted according to the relevant MMN time windows. These were determined by visual inspection, and by calculation of the point of maximum amplitude for the separate conditions. For [tond], the relevant time window was set at 200 to 300 ms after stimulus onset, and for [sond] at 250 to 350 after stimulus onset. Electrodes that were included in the repeated measures analysis of variance (ANOVA) were visually selected on the basis of their relevance to MMN analysis. Of the 38 electrodes used for recording, nineteen were selected: Fp1, Fp2, F3, F4, C3, C4, F7, F8, T7, Fz, Cz, FC1, FC2, CP1, CP2, FC5, FC6, CP5 and CP6.

The repeated measures ANOVA for the multideviant block showed significant effects for the MMNs of both [tond] and [sond]. Also, the MMN of [tond] minus [θond] vs. the MMN of [sond] minus [θond] was significant for maximum amplitude ($F(1, 15) = 7.9$, $p<0.02$), maximum latency ($F(1,15) = 63.4$, $p<0.01$), and area measured in μV per ms ($F(1,15) = 6.5$, $p<0.05$). Figure 2 shows the difference waves of [tond] and

[sond] minus [θond] for the single electrode Fz, with N1, P2 and MMN marked. For the equiprobable block the difference waves were also significant different regarding maximum amplitude ($F(1,15) = 11.2$, $p<.005$), maximum latency ($F(1,15) = 53.4$, $p<.001$), and area ($F(1,15) = 12.5$, $p<.004$).

Figure 1a: Deviant minus standard block 1 (2 µV). Top row: /s/ - /th/. Bottom row: /t/ - /th/.

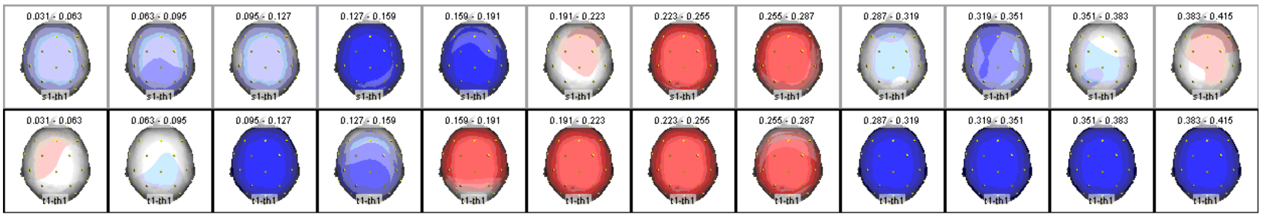


Figure 1b: Deviant minus standard block 2 (2 µV). Top row: /s/ - /th/. Bottom row: /t/ - /th/.

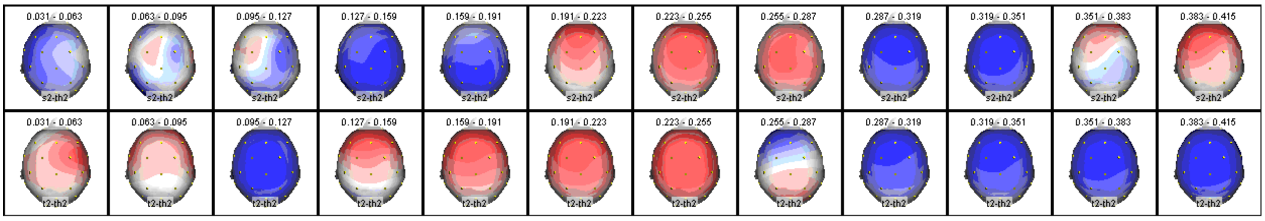


Figure 1c: Difference map block 1 / block 2 (2 µV). Top row: /s1/ - /s2/. Bottom row: /t1/ - /t2/.

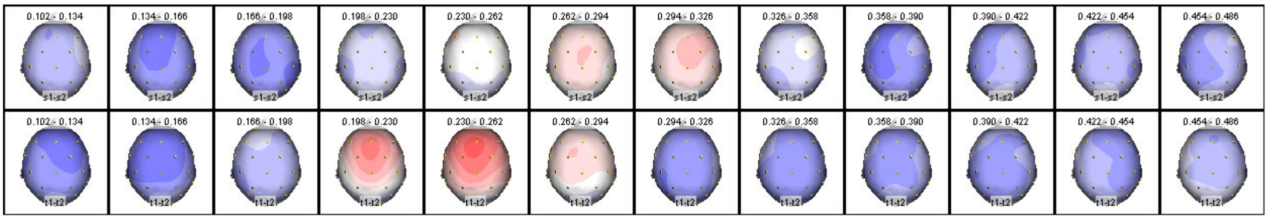
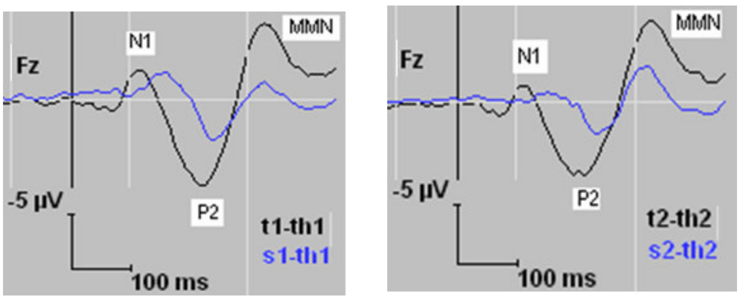


Figure 2: Difference waves for Fz in the multideviant block (left) and in the equiprobable block (right).



Analysis of the difference waves of [tond] and [sond] as deviants in block 1 versus their difference as standards in block 2 revealed no significant interactions for maximum amplitude ($F(1, 15) = .4$, $p>.5$) or area ($F(1,15) = .120$, $p>.7$). There was, however, a significant interaction for maximum latency ($F(1,15) = 44.4$, $p<.001$). As [tond] evokes an earlier MMN than [sond], this indicates that it is not experience but acoustical similarity that is observed here. Since the observed P2 seemed rather large, and P2 effects have been found in studies on language-specific effects (Dehaene-Lambertz 1997; Brunellière et al. 2009) it was analysed here as well. The time window of the P2 for [tond] in block 1 minus [tond] in block 2 was set at 100 to 200 ms, and it was compared with the P2 of [sond] in block 1 minus [sond] in block 2, set at a time window of 110 to 210 ms. No significant interactions were found for maximum amplitude, maximum latency, and area.

4. DISCUSSION AND CONCLUSION

The prelexical processing of the highly preferred th-substitution [t] and the less preferred th-substitutions [s] in English was investigated in an MMN study with Dutch listeners. The elicitation of MMNs was clearly shown when [θond] was the standard and [tond] and [sond] functioned as deviants, with [tond] eliciting a larger MMN and a shorter latency than [sond]. This result pattern contradicts predictions based on language experience, and is rather in line with an explanation based on perceptual similarities. In terms of perceptual similarity, [s] is closer to [θ] than [t] for Dutch listeners (Cutler et al. 2004). Since [tond] elicited a larger MMN than [sond], the results suggest that perceptual similarity was the driving force for the present findings. This is consistent with other findings showing that the MMN is correlated with acoustic and perceptual distance (e.g. Shtyrov et al. 2007) and that MMN latencies decrease as a function of increased auditory discrimination performance (e.g. Näätänen et al. 1993). It is not consistent, however, with some further MMN studies that did find effects of experience (e.g. Näätänen 1997; Jacobsen 2003; Kirmse et al 2008; Dehaene-Lambertz 1997; Brunellière et al. 2009). However, in contrast to the present study, these experience-based effects were always observed for native listeners. In the earlier studies, phoneme contrasts were also usually chosen based on a within-category versus an across-category difference. Conversely, in the present study the contrast was a th-substitution that is often produced by Dutch learners of English, namely [t], versus the less frequent substitution, [s]. Such an effect of preferences in foreign-accented speech has not been investigated before. It is possible that the experience-based memory representations for the preferred phoneme substitutions are not as well established as native language-specific memory representations.

Nevertheless, prior experience with th-substitutions has been shown to interact with the lexical level of processing (Hanulíková & Weber, in prep.). Before concluding now that, at least for L2 listeners, effects of experience are not influencing prelexical processing, we want to point out some factors that may have affected the observed pattern. One difference to the Hanulíková & Weber study is that the stimuli in the eye-tracking study were spoken by Dutch and German learners of English, but in the present study stimuli were spoken by a native speaker of American English. A native speaker rather than a Dutch speaker was chosen in order to avoid accent-specific advantages for a future comparison with German participants listening to the same stimuli. However, our choice of speaker may have created an English-language context for the listeners, and it remains possible that speaker identity interfered with the application of experience gathered from listening to Dutch learners of English. Moreover, Dutch and English /t/ also differ.

The durational differences of the stimuli should also be considered for their influence on the MMN-amplitudes. As noted, the durations of [θond], [tond], and [sond] were 593, 499 and 609 ms, respectively. Research on durational differences in an MMN paradigm, however, usually focuses on *vowel* length (e.g. Kirmse et al. 2008), whereas here, the durational differences originated from *consonants*. In addition, in studies on vowel length, stimuli are spectrally kept constant, whereas we employed different consonants. Therefore, it is not likely that durational differences alone could have accounted for the amplitude and latency differences found here. Effects of stimulus length are furthermore unexpected because of the relatively small durational differences; effects are generally found when stimuli are at least 40% different in duration (e.g. Jacobsen et al. 2003; Kirmse et al. 2008). In comparison, [tond] was 15% shorter than [θond] in the present study, making a duration-based account of MMN elicitation unlikely.

Surprisingly, the equiprobable block also showed significant interactions between the difference waves of [tond]-[θond] versus [sond]-[θond]. This was unexpected, because this design is well-attested as a control condition. Presenting the three stimuli equally often was expected to result in three ‘standard’ waveforms - hence, no MMNs were expected. One explanation for the elicitation of MMNs in the equiprobable block could be that the order of the blocks (with the equiprobable block being second) was never reversed. It could be that the exposure to 800 [th] standard stimuli in the first block created a very solid neural representation of the standard. The second block, in which a further 200 occurrences of the standard were presented, may have further strengthened this standard representation. The deviants, which were each presented only 100 times in the first block, may not have had enough instances to function as a standard as well. Another possible explanation for the results of the second block in this study may be that the stimuli did not have the exact same frequency as the deviant in the first block. This procedure has been adapted by some studies (see e.g. Maess et al. 2007), although others do report the procedure adopted here (e.g. Horváth et al. 2008).

In sum, we investigated the prelexical processing of English th-substitutions in an MMN-paradigm with Dutch L2 listeners. Our main goal was to test whether the deviant stimulus [tond], which is a highly typical mispronunciation for Dutch learners of English, is processed differently from the less typical deviant [sond]. The results clearly show that word initial [t] and [s] lead to different brain responses in terms of latency when they are compared with word-initial [θ], with [t] producing later MMNs effects than [s]. Thus, although Dutch listeners are highly familiar with [t]-substitutes from their L2 speech, this experience was not reflected in the MMN effects (in this case, the opposite pattern of results should have been observed). Rather perceptual similarity between the substitutes and the English interdental fricative [θ] can explain the findings. It remains possible, however, that experience still modulates the MMN effects even though it is not the main force for the effects. A comparison of Dutch listeners in the present study with German listeners, for whom the opposite experience pattern holds (i.e., German learners substitute [θ] mostly with [s] and less often with [t]), will help to clarify this possibility in the future.

5. REFERENCES

Advanced Source Analysis Software, ANT Software, Enschede, The Netherlands. [Computer Programme]

Allopenna, P.D., Magnuson, J.S., Tanenhaus, M.K. 1998. Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *Journal of Memory and Language* 38. 419-439.

Bien, H., Lagemann, L., Dobel, C., Zwitserlood, P. 2009. Early automatic categorization of speech sounds is not misled by changes in the surface form - a dissociation of behavioural and neurophysiological data.

Boersma, P. and Weenink, D. 2002. Praat 5.0: a system for doing phonetics with the computer (version 5.0) [Computer Programme]

Brunellière, A., Dufour, S., Nguyen, N. and Frauenfelder, U. H. 2009. Behavioral and electrophysiological evidence for the impact of regional variation on phoneme perception. *Cognition* 111. 390-396.

Coenen, E., Zwitserlood, P., Bölte, J. 2001. Variation and assimilation in German: consequences for lexical access and representation. *Language and Cognitive Processes* 16. 535-564.

Connine, C.M. 2004. It's not what you hear but how often you hear it: on the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin and Review* 11. 1084–1089.

Cutler, A., Weber, A., Smits, R., Cooper, N. (2004). Patterns of English phoneme confusions by native and non native listeners. *Journal of the Acoustical Society of America* 116. 3668-3678.

Dehaene-Lambertz, G. 1997. Electrophysiological correlates of categorical phoneme perception in adults. *NeuroReport, 8,* 919-924.

Hancin-Bhatt, B.J. 1994. Phonological transfer in second language perception and production. Doctoral Dissertation, University of Illinois.

Hanulíková, A., Weber, A. In prep. Sink positive: Linguistic experience with th-substitutions influences non-native word recognition.

Hanulíková, A., Weber, A. Production of English interdental fricatives by Dutch, German, and English speakers. Proceedings of the Sixth International Symposium on the Acquisition of Second Language Speech, New Sounds 2010, Poznan.

Hautus, M.J., Meng, X. 2002. Decision strategies in the ABX psychophysical task. *Perception & Psychophysics* 64. 89-106.

Horváth, J., Czigler, I., Jacobsen, T., Maess, B., Schröger, E., Winkler, I. 2008. MMN or no MMN: no magnitude of deviance effect on the MMN amplitude. *Psychophysiology* 45. 60-69.

Jacobsen, T., Horváth, J., Schröger, E., Lattner, S., Widmann, A., Winkler, I. 2004. Pre-attentive auditory processing of lexicality. *Brain and Language* 88. 54-67.

Jacobsen, T., Schröger, E. 2003. Measuring duration mismatch negativity. *Clinical Neurophysiolog* 114. 1133-1143.

Kirmse, U., Ylinen, S., Tervaniemi, M., Vainio, M., Schröger, E., Jacobsen, T. 2008. Modulation of the mismatch negativity MMN) to vowel duration changes in native speakers of Finnish and German as a result of language experience. *International Journal of Psychophysiolog* 67. 131-143.

Maess, B., Jacobsen, T., Schröger, E., Friederici, A.D. 2007. Localizing pre-attentive auditory memory-based comparison: magnetic mismatch negativity to pitch change. *NeuroImage* 37. 561-571.

Mitterer, H., Ernestus, M. 2006. Listeners Recover /t/s that speakers reduce: evidence from /t/-lenition in Dutch. *Journal of Phonetics* 34. 37-103

Näätänen, R., Gaillard, A.W., Mantysalo, S. 1978. Early Selective-attention effect on evoked potential reinterpreted. *Acta Psychologica* 42. 313-329.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R.J., Luuk, A., Allik, J., Sinkkonen, J., Alho, K. 1997. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Letters to Nature* 385. 432-434.

Näätänen, R., Schröger, E., Karakas, S., Tervaniemi, M., Paavilainen, P. 1993. Development of a memory trace for a complex sound in the human brain. *NeuroReport* 4. 503-506

Shtyrov, Y., Pulvermüller, F. 2007. Language in the mismatch negativity design. motivations, benefits, and prospects. *Journal of Psychophysiology* 21. 176-187.

Wester, F, Gilbers, D., Lowie, W. 2007. Substitution of dental fricatives in English by Dutch L2 speakers. *Language Sciences* 29. 477-491.

On the Acquisition of English Stress by Spanish Native Speakers

Verónica del Camen Villafaña Rojas

University of Essex
vdcvil@essex.ac.uk

ABSTRACT

This paper presents the preliminary results of ongoing research on the acquisition of English stress by Spanish native speakers learning English as a foreign language in Mexico. The data were gathered from three tasks a production, a perception and a repetition task. Both real and nonce words were used; these were verbs ending in the suffixes: –ate and –ise; adjectives ending in the suffix –atory and nouns ending in the suffix –ator. The analysis presented here compares the data obtained from 9 Spanish native speakers who are all English instructors in Mexico with the data from a control group; 12 English native speakers. The results seem to indicate that not all words are equally difficult to acquire; although the participants performed in an English-like way in the repetition task, their performance in the perception and production tasks on items containing the suffixes –atory and –ator were less satisfactory. This paper will suggest that the L2 learners have not learned the rules of stress placement in English but instead they seem to have stored the items in their lexicon.

Keywords: English stress, Second Language Phonology, Perception, Production, and Language Acquisition.

1. INTRODUCTION

Previous studies on the acquisition of English stress patterns by second language learners have shown that perceiving stress accurately does not always imply the ability to produce stress correctly and that incorrect perception of stress does not imply incorrect production (Lee, 2006; Altmann, 2006). Archibald (1993) using the metrical parameter approach, reported that speakers of languages of predicable stress (Polish and Hungarian) perform more poorly in the production of English stress than in the perception of it; conversely Youssef and Mazurkewich, (1998) who adapted their study from Archibald's (1993), concluded the opposite in their study with Egyptian Arabic learners of English. With the aim of shedding more light on the acquisition of English stress by speakers of a language with phonological but unpredictable stress on the word level, Spanish, from the perspective of both perception and production, the present study included a production, a perception and a repetition task. It seeks to find out whether L2 learners' pronunciation problems are connected with their perception deficiency of the target sounds. In order to pursue this objective, both real and nonce words were used. The target words are verbs with the suffixes: –ate and –ise; adjectives with the suffix –atory and nouns ending in –ator.

It is predicted that, if the L2 learners have acquired the stress pattern of English, they will produce both real and unreal words in an English-like way. If they have, on the other hand, stored the lexical items as they would for Spanish real words with antepenultimate stress[1], their performance will be better in real items and they may assign antepenultimate stress to nonce forms by means of analogy. In the perception task, L2 learners are predicted to perform better than English native speakers (NS hereafter) due to the fact that they are taught from primary school to identify the stressed syllable in Spanish and also they are used to marking lexical stress while the former are not. The repetition task aims to find evidence to suggest that although the NS and L2 learners have difficulty in perception tasks where they are required to identify the syllable that bears main stress (as reported by Youssef and Mazurkewich 1998), if they are able to repeat the word with the same stress pattern this could indicate that they are able to perceive stress correctly in the word.

2. METHODOLOGY

In order to collect the data, the participants were assigned three experimental tasks: a production, a perception and a repetition task.

2.1. Stimuli

Previous studies on the acquisition of the English stress placement have used either nonce words or real words as part of their research. The stimuli of the present study consisted of eighty four words; forty-two real words and forty-two nonce forms. There were three real and three unreal words for each of the following seven word types: HL–ate, LL–ate, HL–ise, LL–ise, LLL–ise , LLL–atory, and LLL–ator, where H stands for a syllable with the structure: consonant, vowel, consonant (CVC) or consonant and a long vowel (CVV). L stands for a syllable with the structure: consonant, vowel (CV). The nonce forms were created following the guidelines for creating nonce words, as suggested by Hochberg (1988). Twenty-one real words and twenty-one unreal words were used in the production and repetition tasks and the other forty-two words were used in the perception task.

2.2. Subjects

Nine Mexican Spanish native speakers were recruited for this study. They work as English intructors in Toluca in the Faculty of Languages of the Autonomouns University of the State of Mexico in Mexico. They are advance learners of English, certified with ALTE 5 proficiency level, according to the ALTE framework[2]. They were chosen so as to get a better understanding of the input the students of the Faculty of Languages learning English receive. A control group consisting of 10 undergraduate British English native speakers and 2 American English native speakers, who work as English instructiors in the same Faculty as the Spanish native speakers, was also used in this study. They range in age from 20 to 45.

2.3. Procedure

The participants were assigned three experimental tasks: a production, a perception, and a repetition task, in that sequential order. In each task the words used were randomised. The participants were tested individually in a sound-attenuated room. Prior to each actual experiment, all subjects read a brief text explaining the procedure of the task and received practiced items. The production task was an online task where the focus word was presented in bold in contextualised sentences. Only one sentence appeared on the screen at a time. The participants were asked to read each word out loud and their production was recorded using a digital recorder. The second task was the perception task, for this task, the participants listened to a native speaker of British English as they uttered each of the target items twice using headphones. The participants received a piece of paper with the words separated in syllables, and were asked to mark the most prominent syllable with a stress mark (´), resembling the way one marks lexical stress in Spanish. They were allowed intervals of about five seconds between each item. Lastly, the participants heard each of the words used in the production task over headphones; the words were uttered by a native speaker of British English. The participants were asked to repeat each word after they heard it and their production was recorded using a digital recorder. There was an interval of about five seconds between each word.

2.4. Scoring

Following Louriz (2004), the collected data from the production and repetition task was analysed using the software "WASP" so as to identify the syllable with main stress. This software displays spectrograms and fundamental frequency and it clearly points the syllable that bears primary stress. The results from the perception task were printed so there was no need to transcribe them. All data were entered into SPSS and analysed with one-way repeated measures ANOVA to judge significance.

3. RESULTS

Below are the results obtained in the three tasks: production of real items (real prod.), production of unreal items (unreal prod.), perception of real items (real per.), perception of unreal items (unreal per.), repetition of real items (real rep.) and repetition of unreal items (unreal rep.). The sections are divided by the suffix used. Each table shows both the correct and incorrect responses by the L2 learners (ALTE 5) and the control group (NS). The column labelled *incorrect* indicates the percentage of responses when these syllables received main stress; the syllables are counted from the right edge of the word. The *correct* column on the right shows the percentage of such responses. To avoid percentages misleading the interpretation of the results, the final row in each figure indicates the number of tokens a word received main stress in that syllable.

3.1. –ate words

Table 1 shows the responses obtained in words with the syllable combination HL–ate. Although most Spanish speakers stressed the antepenultimate syllable for both real and nonce words in the production task, the difference between their scores and the one from NS is significant (p = .021), which implies that they could not be said to be performing in an English-like way. As predicted, it seems the L2 learners are storing stress in the lexicon and because the difference in their performance in real and nonce forms is not significant, they could be claimed to be assigning stress by means of analogy. It is also important to notice where errors occur in the L2 learners' production of words, it is always the final syllable that is stressed, which may suggest some L1 transfer due to the fact that in Spanish verb forms nearly all the time receive final stress. As predicted, in the perception task the L2 learners perform better than the NS, having difficulty only with nonce forms; the NS perform equally in real and nonce words. There difference in this task is not statistically significant. Neither group shows problems in the repetition task.

Table 1: Stress patterns in words with the syllable combination HL-ate (amputate, complicate, conjugate, fangitate, asmigate, domflicate, palpitate, calculate, fambitate, osbigate)

	% incorrect				**% correct**	
	penult		final		**antepenult**	
	NS	ALTE 5	NS	ALTE 5	**NS**	**ALTE 5**
real prod.	0	0	0	22.2	**100**	**77.7**
unreal prod.	0	0	0	25.9	**100**	**74.1**
real per.	16.6	0	0	0	**83.3**	**100**
unreal per.	8.3	7.4	8.3	18.5	**83.3**	**74.1**
real rep.	0	0	0	0	**100**	**100**
unreal rep.	0	0	0	0	**100**	**100**
# tokens	9	2	3	18	**204**	**142**

Table 2 shows the responses obtained in words with the syllable combination LL–ate. Similar to the responses shown in Figure 1, most Spanish speakers stressed the antepenultimate syllable for both real and nonce words in the production task, the percentage, however, is much lower. The difference between their scores and the one from the NS, who again obtained 100% accuracy, is significant (p = .004), which could be interpreted as further evidence that the L2 learners have not acquired the English stress patterns of this type of words, but rather have stored the items in the lexicon. Because the difference between the scores obtained for real and nonce forms is not significant, they could be said to assign stress by means of analogy to the nonce forms. Where errors occur in the L2 learners' production of words, they imply again stressing the final syllable instead of the antepenult. As suggested before, this could be the result of L1 transfer. This time, all participants seem to have overall less difficulty in the perception task. They perform, however, significantly better in identifying main stress in real words than in nonce forms (p = .007). Neither group shows problems in the repetition task.

Table 2: Stress patterns in words with the syllable combination LL-ate (fabricate, saturate, moderate, comonate, fipolate, baturate, celebrate, operate, collocate, gabelate, inustate, fatorate)

	% incorrect				**% correct**	
	penult		final		**antepenult**	
	NS	ALTE 5	NS	ALTE 5	**NS**	**ALTE 5**
real prod.	0	0	0	40.7	**100**	**59.2**
unreal prod.	0	0	0	37.1	**100**	**62.9**
real per.	0	0	2.7	3.7	**97.2**	**96.2**
unreal per.	8.3	11.1	5.5	18.5	**86.1**	**70.3**
real rep.	0	0	0	0	**100**	**100**

unreal rep.	0	0	0	0	**100**	**100**
# tokens	3	3	3	27	**210**	**132**

3.2. –ise words

Table 3 shows the responses obtained in words with the syllable combination HL–ise. Similar to the responses shown in Figure 1, and Figure 2 most Spanish speakers stressed the antepenultimate syllable for both real and nonce words in the production task, again errors occur when the final syllable is stressed. The difference between the scores of the two groups is significant (p = .008), offering further evidence for the claims stated above. In the perception task the NS perform significantly better than the L2 learners (p = .035). This time in the repetition task the L2 learners seem to have some difficulty with repeating nonce forms, the difference between their scores with real and nonce words is not significant.

Table 3: Stress patterns in words with the syllable combination HL-ise (agonise, terririse, digitise, figitise, degorise, agimise, memorise, criticize, televise, feborise, ledevide, magolise)

	% incorrect				**% correct**	
	penult		final		**antepenult**	
	NS	ALTE 5	NS	ALTE 5	**NS**	**ALTE 5**
real prod.	0	0	0	29.6	**100**	**77.7**
unreal prod.	0	0	0	29.6	**100**	**77.7**
real per.	8.3	7.4	2.7	29.6	**88.8**	**62.9**
unreal per.	2.7	7.4	2.7	25.9	**94.4**	**66.6**
real rep.	0	0	0	0	**100**	**100**
unreal rep.	0	0	0	11	**100**	**88**
# tokens	4	4	2	32	**210**	**126**

Table 4 shows the responses obtained in words with the syllable combination LL–ise. These results offer evidence for the claims made above with respect to the production tasks since they mirror responses shown in Figure 1 and Figure 2. The difference between the scores of the two groups is significant (p = .036). In the perception task the NS perform equally in nonce and real words but the difference between their scores and those from the L2 learners is not significant. In the same task the L2 learners seem to have more difficulty with perceiving stress in nonce forms, the difference, however, between their scores with real and nonce words is not significant. Neither group shows problems in the repetition task.

Table 4: Stress patterns in words with the syllable combination HL-ise (authorise, socialise, privatise, forbanise, bruforise, permirise, formalise, verbalise, organise, audorise, frutalise, dorbalise)

	% incorrect				**% correct**	
	penult		final		**antepenult**	
	NS	ALTE 5	NS	ALTE 5	**NS**	**ALTE 5**
real prod.	0	0	0	22.2	**100**	**77.7**
unreal prod.	0	0	0	29.6	**100**	**70.3**
real per.	8.3	3.7	0	14.8	**91.6**	**81.4**
unreal per.	5.5	11.1	2.7	18.5	**91.6**	**70.3**
real rep.	0	0	0	0	**100**	**100**
unreal rep.	0	0	0	0	**100**	**100**
# tokens	5	4	1	23	**210**	**135**

Table 5 shows the responses from words with the syllable combination LLL–ise. It seems the L2 learners have more difficulty assigning antepenultimate stress in this type of words. The difference between the scores of NS and the L2 learners in real words is significant (p = .032) and so is the difference for both groups between their scores in real and nonce forms (p = .006). This is not surprising if we consider that the real words maintain stress in the same syllable after suffixation (i.e. monopoly+ise, category+ise). In the perception task the NS perform better in nonce than in real words and the difference between their scores and the L2 learners' is significant (p = .006). Although in the repetition task the L2 learners perform better with nonce forms than the NS, the opposite happens for real words; what is more, the difference between the groups is significant (*p* = .043) only in the scores for real words.

Table 5: Stress patterns in words with the syllable combination LLL-ise (monopolise, economise, categorise, fimocratise, telephonise, modecorise, apologise, democratise, capitalise, lafitalise, pateborise, edogorise)

	% incorrect								**% correct**	
	preante		antepenult		penult		final		**antepenult**	
	NS	ALTE 5	NS	ALTE 5	NS	ALTE 5	NS	ALTE 5	**NS**	**ALTE 5**
real prod.	2.7	7.4	0	22.2[3]*	0	0	0	29.6	**97.2**	**40.7**
unreal prod.	19.4	11.1	0	N/A	0	7.4	0	29.6	**80.5**	**51.8**
real per.	5.5	0	0	0	0	7.4	5.5	22.2	**88.8**	**70.3**
unreal per.	2.7	0	0	7.4*	0	7.4	0	7.4	**97.2**	**77.7**
real rep.	0	0	0	11.1*	0	0	0	3.7	**100**	**85.1**
unreal rep.	0	0	11.1*	0	0	0	0	0	**88.8**	**100**
# tokens	11	5	4	11	0	6	2	25	**199**	**115**

3.3. –atory words

Table 6 shows the responses from words with the syllable combination LLL–atory. Both groups appear to have difficulty with the production of this type of words, with the nonce forms being the most difficult. The difference between their scores in real and nonce words is significant (*p* = .020). Just like in the four-syllable words ending in –ise, one should bear in mind that the real words maintain stress in the same syllable after suffixation (i.e. prepare+atory) and because the nonce forms do not come from real items, they represent difficulty to the speakers. In the perception task, the NS perform slightly better than the L2 learners, but the difference is not is significant. In the repetition task both groups perform better in real words than with nonce forms and the difference between these scores is significant (*p* = .028).

Table 6: Stress patterns in words with the syllable combination LLL-atory (respiratory, preparatory, declamatory, diplomatory, tisipatory, grasilatory, regulatory, obligatory, derogatory, firomatory, apasatory, cablifatory)

	% incorrect								**% correct**	
	preante		antepenult		penult		final		**antepenult**	
	NS	ALTE 5	NS	ALTE 5	NS	ALTE 5	NS	ALTE 5	**NS**	**ALTE 5**
real prod.	13.8	14.8	N/A	N/A	8.3	7.4	0	29.6	**77.7**	**48.1**
unreal prod.	25	14.8	N/A	N/A	25	14.8	0	33.3	**50**	**37.1**
real per.	0	3.7	16.6[4]	7.4*	0	7.4	0	3.7	**83.3**	**77.7**
unreal per.	0	0	8.3*	0	5.5	14.8	0	0	**86.1**	**85.1**
real rep.	0	0	N/A	N/A	0	0	0	3.7	**100**	**96.2**
unreal rep.	2.7	0	2.7*	7.4*	0	0	0	0	**94.4**	**92.5**
# tokens	15	9	10	4	14	12	0	19	**177**	**118**

3.4. –ator words

Table 7 shows the responses from words with the syllable combination LLL–ator. In is in this type of words that the L2 learners show the poorest performance in the production task. The difference between their scores and those of the NS is highly significant (*p* < .001). This may be a consequence of main stress being in the forth syllable from the right edge. Spanish does not have preantepenultimate stress expect on words plus clitics. The majority of the L2 learners' responses assign stress to the penultimate syllable, as they would if they applied the stress pattern of Spanish due to the diphthong /eɪ/, therefore it implies transfer from the L1. In the perception task, the NS perform better than the L2 learners, and the difference between the scores of the two groups is significant (*p* = .003). In the repetition task both groups do not have any incorrect responses.

Table 7: Stress patterns in words with the syllable combination LLL-ator (animator, navigator, separator, evidator, bogerator, naficator, moderator, illustrator, generator, padigator, afinator, loberator)

	% incorrect						**% correct**	
	antepenult		penult		final		**preantepenult**	
	NS	ALTE 5	NS	ALTE 5	NS	ALTE 5	**NS**	**ALTE 5**
real prod.	0	0	0	51.8	0	0	**100**	**48.1**
unreal prod.	2.7	0	2.7	66.6	0	0	**94.4**	**33.3**
real per.	5.5	0	8.3	59.2	0	0	**86.1**	**40.7**
unreal per.	2.7	3.7	2.7	37	0	0	**94.4**	**59.2**
real rep.	0	0	0	0	0	0	**100**	**100**
unreal rep.	0	0	0	0	0	0	**100**	**100**
# tokens	4	1	5	58	0	0	**207**	**103**

4. CONCLUSIONS

The results presented above indicate that because the L2 learners do not behave in and English-like way in the production tasks, they do not have acquired the stress pattern of English, but instead, they have stored the lexical items as they would for Spanish real words with antepenultimate stress and assign stress to the nonce words by means of analogy. With respect to the perception task, the data did not support the prediction that L2 learners would perform better than the NS. The fact that both groups perform almost perfectly in all word types in the repetition task, may indicate that the participants are able to perceive stress, but fail to identify in spelling the syllable that bears it. Further research is suggested combining the methodology of the perception and the repetition task to better support the claims about the participants ability to perceive stress correctly.

5. REFERENCES

Archibald, J. 1993. *Language Learnability and L2 Phonology.* Dordrecht: Kluwer .

Altmann, H. 2006. The Perception and Production of Second Language Stress: A crosslinguistic experimental study. PhD Dissertation. University of Delaware.

Hochberg, J. 1988. First steps in the acquisition of Spanish stress. *Journal of Child Language* 15, 273-292

Lee, S. 2006. The relationship between perception and production in L2 English stress acquisition. *Studies in Phonetics, Phonology and Morphology* 12.3, 661-685.

Louriz, N. 2004. Acquisition of L3 Phonology: A case study. Unpublished PhD Thesis, University of Essex. UK.

Youssef, A. & I. Mazurkewich (1998). The acquisition of English metrical parameters and syllable structure by adult native speakers of Egyptian Arabic (Cairene dialect). In: S. Flynn, G. Martohardjono, & W. O'Neil (eds.) *The generative study of second language acquisition.* Mahwah: Lawrence Erlbaum Associates, 303-332 .

Villafaña Rojas V. 2006. *Stress Placement in Spanish*. Unpublished MA Dissertation, University of Essex, UK.

NOTES

[1] Villafaña (2006) showed that Spanish native speakers tend to assign antepenultimate stress to nonce forms by means of analogy. It suggests that antepenultimate stress is stored in the lexicon, rather than being the result of extrametricality being present in some stems.

[2] ALTE – The Association of Language Testers in Europe. http://www.alte.org/about/index.php

[3] *These percentages correspond to the words: *categorise* in the production and repetition tasks, and *capitalise* and *pateborise* in the perception task which receive stress in the forth syllable from the right. Therefore assigning antepenultimate stress to these words is regarded as incorrect. The tokens, where these words received stress in the forth syllable from the right, are included in the %correct column.

[4] *These percentages correspond to the words: *tisipatory* in the repetition tasks, and *regulatory*, and *cablifatory* in the perception task which were recorded with stress in the forth syllable from the right.

The role of metrical stress in L2 word recognition

Silke Weber and John Archibald

University of Calgary
ksuweber@ucalgary.ca, john.archibald@ucalgary.ca

ABSTRACT

Foreign accents sometimes but not always reduce the intelligibility of non-native speech (e.g. Derwing and Munro 1997). The present study investigates the question why some types of pronunciation errors are more problematic at others. Specifically, we examine the effect of two phonologically different types of stress errors on word recognition. Eight native speakers of German participated in a lexical decision task in which they listened to correctly stressed and incorrectly stressed German nouns as well as non-words and had to decide whether each item was a word in German or not. Based on response latencies, we argue that stress errors only impede word recognition if they result in a violation of metrical foot structure.

Keywords: Metrical feet, stress errors, intelligibility, non-native accent.

1. INTRODUCTION

A range of studies conducted by Munro and Derwing indicate that foreign accent, comprehensibility and intelligibility are partly independent measures (e.g. Derwing and Munro 1997; Munro and Derwing 1995, 2000; Munro et al. 2006). Even non-native speakers with strong foreign accents may be perfectly intelligible and comprehensible. Intelligibility can be defined as the conformity of speaker intention and listener interpretation. Comprehensibility refers to the ease with which an utterance can be understood.

The evidence for crucial differences between the three measures amounts to showing that accent ratings are usually harsher than comprehensibility ratings, which in turn are harsher than actual intelligibility scores. In Munro and Derwing (1995), for instance, participants were asked to listen to utterances produced by non-native speakers of English and to provide accentedness and comprehensibility ratings on nine-point scales. It was found that the accentedness scores of utterances that were easily comprehensible (score of 1 or 2) ranged from one to nine. In other words, even non-native speakers with strong foreign accents may be perfectly intelligible and comprehensible.

This finding raises two important questions: First, which types of pronunciation errors reduce the intelligibility and comprehensibility of non-native speakers? Second, why are some types of errors more problematic than others? The present study addresses these questions by investigating the effect of two phonologically different types of stress errors on word recognition. Based on our findings, we are going to argue that stress errors only impede word recognition if they result in a violation of metrical foot structure.

This paper is structured as follows: In §2, we discuss how stress properties of the listener's L1 might affect word recognition; in §3, we discuss relevant properties of German word stress. We turn to our own study in §4. In §5, we discuss our findings and indicate directions for further research.

2. PREVIOUS RESERACH ON THE ACQUISITION OF STRESS

2.1. The production of L2 stress

It is well established that non-native speakers sometimes misplace word stress. In one of their experiments, Wayland et al. (2006), for instance, tested native Thai L2 speakers of English on their knowledge of the position of stress in real English words. The participants produced 36 sentences of the type "I said X this time", where X was either a bi- or a trisyllabic English noun or verb. Half of the items were consistent, the other half was inconsistent with English stress regularities (e.g. Guion et al. 2003). The error rate was 10% for words following the regular pattern, 28% for words with an irregular stress pattern. Similarly, Archibald (1992, 1993) showed that native speakers of Polish and Spanish, respectively, sometimes made stress errors when producing English words. The Spanish participants (Archibald 1993) misplaced stress on bi- and trisyllabic English items between 8% and 26% of the time, indicating again, that stress errors are fairly common in L2 speech, even in a laboratory setting. In spite of some regularities, the types of errors the participants in these studies made were relatively unpredictable.

While the numbers presented above cannot be considered representative of non-native speakers, they support our assumptions that a) stress errors are common in non-native speech and b) that non-native speakers make different types of stress errors.

2.2. Stress in L1 and L2 word recognition

Whether or not mismatching word stress impedes lexical activation has been shown to depend on two factors: first, the stage of lexical activation the experiment taps into and second, the L1 of the listener. In the following, we will briefly review some evidence for these claims. Moreover, we will propose a third factor, which to the best of our knowledge has not been investigated before: the type of metrical structure violation resulting from the stress error.

A series of fragment priming experiments conducted with native speakers of Spanish (Soto-Faraco et al. 2001), English (Cooper et al. 2002), and Dutch (van Donselaar et al. 2005) demonstrated that stress-matching primes facilitate while stress-mismatching primes inhibit lexical activation. The participants listened to bisyllabic primes consisting of the first two syllables of a longer word. At the offset of the auditory prime, they saw a longer word on screen. The first two syllables of the target items matched the prime segmentally and either matched or mismatched with respect to stress. The participants' task was to decide as quickly and accurately as possible whether the item they saw on screen was a word in the respective language or not. Response times were significantly faster after stress-matching primes indicating that speakers of all three languages exploit stress cues in early lexical activation.

However, Dupoux et al. (2008) demonstrate that French L2 learners of Spanish cannot reliably distinguish Spanish words from non-words that differ only in stress. Similarly, Tremblay (2008) found that French L2 learners of English did not make use of stress information in word recognition indicating that whether or not stress is used in word recognition is language dependent. We will discuss this finding in more detail below.

Our assumption that different types of stress errors might affect intelligibility to a different degree is based on a study by Domahs et al. (2008). Domahs et al. (2008) conducted an ERP study investigating brain responses to correctly and incorrectly stressed trisyllabic German nouns. A third of the stimuli had correct stress on the first syllable, a third on the second and a third on the final syllable. In the experiment, each item occurred three times: Once with correct stress, twice with incorrect stress. The noun was first presented visually on screen and then it was presented aurally, either with correct or incorrect stress. The task was to decide whether the stimulus was stressed correctly or not. Interestingly, it was found that this task was easy to do in some cases and more difficult in others. In some cases the participants frequently indicated that the item was stressed correctly when as a matter of fact it was not. Domahs et al. (2008) conclude that stress errors were easy to detect when they involved a violation of foot structure and more difficult when the other foot was stressed. This finding suggests that different types of stress errors may also impede word recognition to a different degree.

2.3. The perception of L2 stress

Arguably, stress errors can affect word recognition only if the listener is able to accurately perceive stress. Previous research has shown that some listeners reliably perceive word stress while others do not. It is widely assumed that L1 stress properties allow to predict a listener's performance in (L2) stress perception. The currently dominant approach relates the ability to perceive stress to the surface predictability of stress in the listener's L1 (e.g. Peperkamp and Dupoux 2002, Altmann 2006).

The Stress Deafness Model (Peperkamp and Dupoux 2002), for instance, argues that only speakers of languages that lexically store stress are able to reliably perceive it. According to this model, stress is lexically stored if one of the following conditions is met: Either stress is phonologically non-predictable in the listener's native language; or it is predictable but infants fail to notice its predictability before the end of a critical period after which, supposedly, the stress parameter is set. Listeners are expected to have difficulties perceiving stress, if stress in their L1 is predictable and if the stress regularity is accessible to infants early on in the acquisition process.

Evidence for this model comes from studies comparing the performance of Spanish and French listeners in stress perception tasks (Dupoux et al. 1997, Dupoux 2001). Spanish word stress is partly unpredictable while French stress is always phrase- or, according to some analyses, word-final. The experiments showed that speakers of Spanish could reliably store and retrieve stress patterns while speakers of French could not. A study by Altmann (2006) supports these results based on a wider range of participant groups. In this study on the perception of English stress, speakers of partially unpredictable stress languages (English and Spanish) reliably perceived stress, while speakers of predictable stress languages (Turkish, Arabic, French) did not. Note, though, that French speakers performed above chance level in all experiments.

Based on these findings, we assume that our hypothesis can only be reliably tested on languages with partially unpredictable stress, such as German. We are aware, though, that this can only be a tentative assumption as more recent research suggests that additional factors must be considered as an explanation for performance in stress perception experiments (e.g. Kijak 2009).

3. PROPERTIES OF GERMAN WORD STRESS

In German, primary stress has to fall either on the ultimate, the penultimate or the antepenultimate syllable. In words with a closed penult or a final schwa, the penultimate syllable will almost certainly be stressed. A somewhat less-reliable generalization is that superheavy final syllables attract stress. Based on this assumption, Féry (1996), for instance, claims that German is quantity sensitive. However, Jessen (1999) points out that there are at least 100 exceptions to this rule. And, indeed, others have proposed that German is not quantity-sensitive and that penultimate stress is the default (e.g. Wiese 1996). It is even more difficult to make generalisations regarding monomorphemes ending with a final VC or VV syllable. According to Jessen (1999), syllables of this type are almost as likely to be stressed as unstressed.

These examples are sufficient to conclude that word stress in German is at least partly lexical. To illustrate, monomorphemic nouns with three open syllables can be stressed on the initial, the penulitmate or the final syllable. Examples are given in (1).

(1)	(Rí.si)(ko)	'risk'
	Bi (kí.ni)	'bikini'
	(Pa.ro)(díe)	'parody'

Risiko has initial, *Bikini* penultimate, and *Parodie* final stress. Based on the well-established assumption that German is a trochaic language (e.g. Domahs et al. 2008; Féry 1996; Giegerich 1985; Wiese 1996), we assume the following lexically governed metrical structure of these items: *Bikini* has a single trochee at the right edge of the word, while *Risiko* and *Parodie* are made up of a bisyllabic word-initial and a monosyllabic word-final trochee.

In our study, we investigate the effect of violations of this metrical structure.

4. THE STUDY

4.1. Methodology

4.1.1. Participants

One female and seven male native speakers of Standard German participated in our study, which was conducted at the University of Calgary, Canada. At the point of the study they had resided in Canada for an average of 6.9 months (range: 1 day – 16 months) but spoke German on a daily basis. Their age ranged from 22 to 30 years, with an average of 26.8 years. All reported to have normal hearing. The participants were paid $ 10 for their participation.

4.1.2. Stimuli

90 target and 60 filler items were produced by a phonetically trained native speaker of English. This choice of speaker provided the participants with a credible explanation for mispronunciations, i.e. misplaced stress, while at the same time controlling for accurate pronunciation at the segmental level.

The target items were 90 trisyllabic German nouns. All of them contained only stressable syllables, i.e. syllables with full vowels. 30 of them had main stress on the initial syllable, 30 on the penultimate and 30 on the final syllable. In addition to the number of syllables, the syllable structure was controlled for. Items with final stress contained a final CVCC or CVVC syllable; items with penultimate stress contained a penultimate CV(C) syllable and a final CV syllable; items with initial stress contained a penultimate CV and a final CV(C) syllable. The filler items were 60 German non-words, which matched the target items in syllable structure. They were all tested independently for "goodness" as a German word and found to be acceptable.

Each filler and each target item was produced three times: Once with primary stress on the first, once on the second and once on the third syllable. In addition to correctly stressed items, this procedure yielded stimuli with two types of stress errors: Stress errors involving putting primary stress on the other foot and stress errors involving restructured feet.

Table 1: Foot structure of correctly and incorrectly stressed stimuli.

Correct stress/actual stress	Initial	Penultimate	Final
Initial	(**Ri.**si)(ko)	Ri(**si.**ko)	(Ri.si)(**ko**)
Penultimate	(**Bi.**ki)(ni)	Bi(**ki.**ni)	(Bi.ki)(**ni**)
Final	(**Mo.**nu)(ment)	Mo(**nu.**ment)	(Mo.nu)(**ment**)

Table 1 shows that words with initial stress are stressed on the other foot if stressed on the final syllable and vice versa. Both, words with stress on the initial and with stress on the final syllable involve a violation of foot structure if stressed on the penultimate syllable. If stressed correctly, these items are made up of a bisyllabic and a monosyllabic trochee. If the penultimate syllable is stressed, however, they have a bisyllabic trochee at the right edge of the word. Words with stress on the penultimate syllable have restructured feet both when stressed on the initial and when stressed on the final syllable. This yielded 90 correctly stressed items, 60 items stressed on the other foot and 120 items involving a violation of foot structure.

The stimuli were digitally recorded with 48.2 kHz and 16bit (mono) in a sound attenuated room and transposed to 44.1 kHz using Praat.

4.1.3. Task

The task employed was a lexical decision task. The participants were told that they would be listening to nouns some of which were German words and some of which did not mean anything in German. Furthermore, they were told that the items might sometimes sound funny because they were produced by a non-native speaker of German. Their task was to decide as quickly and as accurately as possible whether each word they heard was a German word or not.

The stimuli were presented in three blocks, which were counterbalanced for number of correctly and incorrectly stressed stimuli.

4.2. Results

Both accuracy scores and reaction times were established. Accuracy scores were taken as a measure of intelligibility. Reaction times indicated the degree of comprehensibility, assuming that an item that is easy to understand takes less processing time than an item that is difficult to understand. While Munro and Derwing (1997) operationalize comprehensibility slightly differently, they note that comprehensibility strongly correlates with processing time.

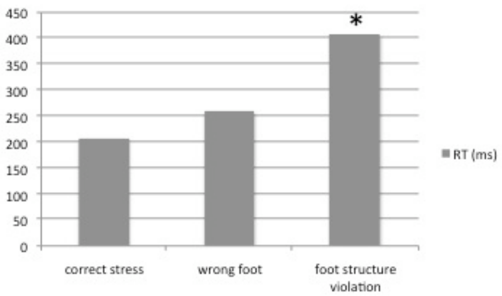
4.2.1.Accuracy scores

All participants correctly identified 94% to 96% of the target items as German words. The fewest errors were made on items stressed correctly, more on items stressed on the wrong foot. Most errors were made on items involving a violation of foot structure. However, the generally very high accuracy scores suggest that a lexical decision task might not be sufficiently sensitive to potential intelligibility problems. Hence, we did not run any statistical analyses to test if the observed differences were significant.

4.2.2. Reaction times

The reaction times support our hypothesis that stress errors involving a violation of foot structure impede comprehensibility more than stress errors where wrong foot is stressed. The mean reaction times of each participant were submitted to an analysis of variance (ANOVA) with one within-participant factor: word stress. The three levels were a) correct stress, b) incorrect stress with primary stress on the wrong foot and c) incorrect stress resulting in a violation foot structure. Repeated measures were used to compare the differences between the three types of items. The main effect of stress type was significant, $F(2,5)=18.029$, $p<0.5$. However, a more detailed analysis showed that the differences were only significant between items stressed on the wrong foot and items with restructured feet but not between correctly stressed items and items with stress on the wrong foot.

Figure 1: Reaction times.



As an overall decrease in reaction time suggested that there was a learning effect from the first to the second and from the second to the third block, ANOVAs was run for each block individually. In the second block, the difference between items stressed correctly and items stressed on the wrong foot was again non-significant, but the difference between these items and words with restructured feet was significant $F(2, 5)=6.053$, $p<0.5$. In the third block, the difference between all three levels was significant.

5. DISCUSSION AND CONCLUSION

The results of this study suggest that not all types of stress errors impair the recognition of trisyllabic German nouns. Nouns that were stressed on the wrong foot were recognized as fast and as accurately as nouns that were stressed correctly. If stress errors resulted in a violation of the foot structure, however, the participants took significantly longer to identify the stimulus as a German word and in some cases they completely failed to do so. While it has been known that in some languages word stress has an impact on word recognition (e.g. Soto-Faraco 2001; Cutler 2005; Donselaar et al. 2005), the finding that not just stress but also metrical structure affects word recognition is new. It has important implications for the relation of foreign accents, comprehensibility and intelligibility and for theories of word recognition.

It has previously been observed that strong non-native accents do not necessarily result in reduced comprehensibility and intelligibility. The present study suggests that whether or not a pronunciation error matters is systematic and can be explained with reference to the phonological structure of the listener's native language. If one assumed that word stress was processed in terms of linear sequences of stressed and unstressed syllables, the results of this study were surprising. Assuming metrical structure, they are not.

At present, one might want to argue that potential alternative explanations for our findings cannot entirely be ruled out. One alternative explanation might be the participants' proficiency in English. Some of the stimuli with final primary stress in German have English cognates with initial primary stress. Hence, good performance on these items might not be surprising. More participants need to be tested on more stimuli to rule out this explanation. Another alternative explanation might be a signal driven approach according to which some syllables are more apt to be stressed than others because they carry secondary stress in Standard German. We currently investigate this approach by doing an acoustic analysis of our stimuli.

It is our intention to expand this research in several ways. First, we are going to test more participants on a range of different tasks and to employ more stimuli in order to get more robust results and to be able to rule out potential alternative explanations for the current findings. Second, we are going to use this and future studies to come up with a theory as to how word metrical structure is represented in the lexicon and parsed during word recognition. This will allow us to systematically expand this line of research to other languages and non-native listeners.

6. REFERENCES

Altmann, H. 2006. The Perception and Production of Second Language Stress: A Cross-Linguistic Experimental Study. Unpublished doctoral dissertation, University of Delaware.

Archibald, J. 1992. Transfer of L1 parameter settings: Some empirical evidence from Polish metrics. *Canadian Journal of Linguistics* 37(3). 301-339.

Archibald, J. 1993. The learnability of English metrical parameters by adult Spanish speakers. *International Review of Applied Linguistics* 31. 129-142.

Cooper, N., Cutler A., Wales, R. 2002. Constraints of lexical stress on lexical access in English: Evidence from Native and Non-native Listeners. *Language and Speech* 45(3). 207-228.

Cutler, A. 2005. Lexical stress. In Pisoni, D. B. & R. E. Remez (eds.). *The Handbook of Speech Perception*. Malden, MA: Blackwell Publishing. 264-289.

Domahs, U., Wiese, R., Bornkessel-Schlesewsky, I., Schlesewsky M. 2008. The processing of German word stress: evidence for the prosodic hierarchy. *Phonology* 25. 1-36.

Donselaar W., van, Koster M., Cutler A. 2005. Exploring the role of lexical stress in lexical recognition. *The Quarterly Journal of Experimental Psychology* 58A. 251-273.

Dupoux, E., Pallier, C., Sebastian, N., Mehler, J. 1997. A destressing "Deafness" in French? *Journal of Memory and Language* 36. 406-421.

Dupoux, E. Peperkamp, S., Sebastian, N. 2001. A robust method to study stress 'deafness'. *Journal of the Acoustical Society of America* 110. 1606-1618.

Féry, C. 1996. German Foot and Word Stress in OT. *Nordlyd* 24. 63-96.

Guion, S., Clark J.J., Harada, T., Wayland R.P. 2003. Factors affecting stress placement for English nonwords include syllabic structure, lexical class and stress patterns of phonologically similar words. *Language and Speech* 46(4). 403-427.

Hayes, B. 1995. *Metrical Stress Theory: Principles and Case Studies*. Chicago: University of Chicago Press.

Jessen, M. 1999. German. In Harry van der Hulst (ed.). *Word Prosodic Systems in the Languages of Europe*. Berlin: Mouton de Gruyter. 515-545.

Kager, R. 1995. Review of Hayes, Bruce. 1995: Metrical stress theory: Principles and case studies. *Phonology* 12. 437-464.

Kijak, A. 2009. How stressfull is L2 stress? *A cross-linguistic study of L2 perception and production of metrical systems*. Utrecht: LOT.

Liberman, M. 1975. The Intonational System of English. Doctoral dissertation, MIT, reprinted 1979 by Garland, New York and London.

Munro, M. J., Derwing T.M. 1995. Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech* 38(3). 289-306.

Munro, M. J., Derwing T.M. 2001. Modeling perceptions of the accentedness and comprehensibility of L2 speech. *Studies in Second Language Acquisition* 23. 451-468.

Peperkamp, S., Dupoux E. 2002. A typological study of stress 'deafness'. In C. Gussenhoven and N.Warner (eds.) *Papers in Laboratory Phonology* 7. Berlin: Mouton de Gruyter. 203-24.

Soto-Faraco, S., Sebastián-Gallés, N., Cutler A. 2001. Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language* 45. 412-32.

Tremblay, A. 2008. Is L2 lexical access prosodically constrained? On the processing of word stress by French-Canadian L2 learners of English. *Applied Psycholinguistics* 29. 553-584.

Wiese, R. 1996. *The Phonology of German*. Oxford: Clarendon Press.

Identifying foreign speakers with an unfamiliar accent or in an unfamiliar language

Andrea Weber[1] *and Katja Pöllmann*[1,2]

[1]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[2]Universität Bamberg, Bamberg, Germany

andrea.weber@mpi.nl, katja.poellmann@mpi.nl

ABSTRACT

For an accent rating study, we recorded 12 English and Dutch short sentences with 6 native speakers of English and with 6 native speakers of Dutch; each speaker read all sentences, thus performing in both her native language and in her second language. The sentences were presented to 3 listener groups who all knew English but varied in their knowledge of Dutch: Dutch listeners, German listeners who knew Dutch, and German listeners who did not know Dutch. Listeners rated all sentences on a scale from 1 (strong foreign accent) to 9 (definitely native speaker). For the English sentences, we found that even German listeners who were not familiar with Dutch were as good as Dutch listeners themselves in identifying Dutch non-native speakers. Thus, familiarity with the language of the accent did not necessarily influence performance. For the Dutch sentences, we found that familiarity with the target language had a noticeably greater influence on responses. That is, German listeners without any knowledge of Dutch were less good at identifying non-native speakers of Dutch than German listeners with Dutch knowledge were. Interestingly, German listeners without any knowledge of Dutch were still better than chance at identifying non-native English speakers in Dutch sentences. The results suggest that familiarity with the target language may play a bigger role in accent detection than familiarity with the accent language, but in the absence of any knowledge of the target language listeners can still reliably detect foreign speakers.

Keywords: L2 speech, accent detection, L2 listener, L1 listener, familiarity.

1. INTRODUCTION

Native-like pronunciation in a second (L2) language is often seen as the most prominent and persistent difficulty for adult L2 learners. This can be problematic for successful communication because native (L1) listeners often have a negative attitude towards speakers with a foreign accent (e.g., Munro et al. 2006), and are known to be highly sensitive to foreignness in speech. For example, it has been shown that L1 listeners can detect L2 speakers in short stretches of speech (e.g., Flege 1984), in filtered speech (e.g., Munro 1995), or when the traces of accentedness are only minimal (e.g., Munro et al., 2003). An often used task for revealing sensitivity to foreignness in L2 speech is accent rating. In accent-rating studies, words or sentences are recorded with L1 and L2 speakers, and listeners are asked to rate on a scale how native-like the speech sounds; L1 listeners then reliably judge L1 speakers to be more native-like than L2 speakers. A debated issue in accent-rating studies is thereby whether listeners' responses indicate something about the speech itself because they are influenced by its acoustic and phonological properties, or whether they indicate something about the listener and therefore vary with the listeners' language experience, for example. Our approach to this issue was to focus on accent ratings from L2 listeners, a listener population for which there is comparatively little evidence for sensitivity to non-nativeness yet (but see Flege 1988; MacKay et al. 2006; Major 2007), and to vary their familiarity with the language of the accent but also their familiarity with the target language (i.e., the language they are asked to rate). In doing so we want to establish whether familiarity with either the target language or the accent language are a prerequisite for detecting non-nativeness in speech, and whether the degree to which familiarity with the target language and the accent language are beneficial, varies.

For our accent rating study, we used short English and Dutch sentences that had been produced by either native or non-native speakers. The L2 speakers for the English sentence were Dutch L1 speakers, and the L2 speakers for the Dutch sentences were English L1 speakers. While our listeners were all highly proficient in English, they varied in their knowledge of Dutch: they were either native listeners of Dutch, German listeners who knew Dutch, or German listeners who did not know any Dutch. German listeners with Dutch knowledge were students living in the Netherlands, immersed in a Dutch-speaking environment. German listeners without Dutch knowledge were students in the south of Germany (i.e., not close to the Dutch border) who have never learned Dutch or been to the Netherlands. Of particular interest was how well German listeners without any knowledge of Dutch could detect Dutch L2 speakers in English sentences, and whether the same listener group could identify English L2 speakers in Dutch sentences at all.

2. ACCENT RATING

2.1. Methods

2.1.1. Participants

Twenty-four native speakers of Dutch, with an average age of 21, took part in the experiment for a small monetary compensation. All 24 Dutch participants were students at the Radboud University in Nijmegen, the Netherlands, and they had received between 5 and 13 years of formal training in English (average of 8 years). In addition, 21 native speakers of German (average age of 23), also all students at the Radboud University, participated. At the time of testing, they had been living in the Netherlands for 2 years on average (ranging from 1 to 5). After their arrival in the Netherlands, they had received an intensive course in Dutch, and judged their level of proficiency in Dutch to be highly advanced, regularly speaking and listening to Dutch for more than 10 hours per week. They had received formal training in English for 9 years on average. The third group of participants consisted of 20 German listeners without any knowledge of Dutch. These participants were all students at the University of Bamberg in Germany (average age of 24). None of them had studied Dutch or had ever been to the Netherlands. All 20 students from Bamberg declared in a questionnaire to have no knowledge of Dutch, never to speak or listen to Dutch, and 17 of the 20 declared not to be familiar with Dutch accented-English at all (the other 3 stated to be a little bit familiar with it).

2.1.2. Materials and Procedure

Six short English declarative sentences and six short Dutch declarative sentences were chosen as materials (see Table 1). The English sentences were based upon the Bamford-Kowal-Bench sentences (Bench et al. 1979) and the Dutch sentences upon the Plomp and Mimpen (1979) sentences. Care was taken that the English and Dutch sentences were comparable in number of words and syllables. Some of the English sentences contained sounds that were judged to be typical sounds of American English (e.g., /ɚ/ in *stirs*), and others were less obviously marked. Dutch- and German-accented English share some markers of foreignness (e.g., in both accents, English /æ/ is usually produced as /ɛ/; see Swan, & Smith, 2001), and we therefore avoided these known overlapping markers to a large extent in the English sentences. Some of the Dutch sentences contained sounds that are typical of Dutch, and for which it is well known that many learners of Dutch have difficulties producing them correctly (e.g., /œy/ in *duinen*). Furthermore, some of the Dutch sentences had sounds that when produced by American learners of Dutch can be typical markers of an American accent (e.g., the /ɾ/ in *prijzen* incorrectly produced as /ɹ/).

Table 1: English sentences and Dutch sentences (English translation in brackets) used in the accent rating study.

English sentences	Dutch sentences
1. He cut his finger.	1 Alle prijzen waren verhoogd. (All prices were increased.)
2. He found his brother.	2. Het vliegtuig vertrekt over een uur. (The airplane leaves in one hour.)
3. Somebody took the money.	3. De schrijver is goed. (The composer is good.)
4. The mother stirs the tea.	4. Het meisje stond te wachten. (The girl was waiting.)
5. The girl has a picture book.	5. De bus is niet op tijd. (The bus is not on time.)
6. The pond water is dirty.	6. Het was leuk in de duinen. (It was nice in the dunes.)

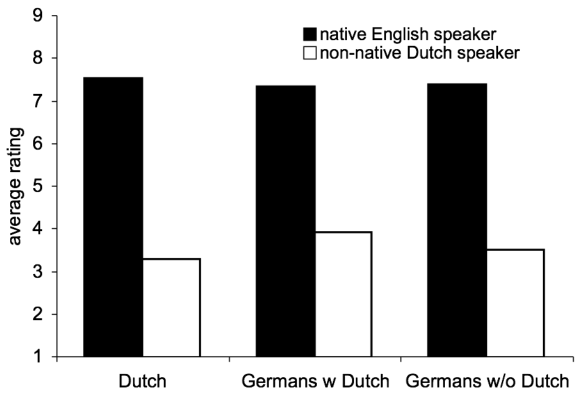
All 12 sentences were recorded by 6 native speakers of American English (2 male and 4 female, average age of 39, ranging from 27 to 68) and by 6 native speakers of Dutch (2 male and 4 female, average age of 46, ranging from to 28 to 71). At the time of the recording, the American speakers had been living in the Netherlands for 3 years on average. While the American speakers judged their proficiency in Dutch to be lower than the Dutch speakers estimated their proficiency in English, all speakers could produce the sentences in their L2 at a normal speech rate and without any noticeable disfluencies. Two independent native listeners furthermore ascertained that none of the speakers was completely accent-free. Sentences were recorded in a quiet room on a digital recorder at 44.1 kHz sampling rate with 16-bit resolution and were later transferred to a computer. Dutch sentences were on average 1.75 s long when produced by the Dutch speakers and 1.77 s when produced by the American speakers; English sentences were on average 1.39 s long when produced by the American speakers and 1.42 s when produced by the Dutch speakers.

The 144 sentences (12 sentences x 12 speakers) were presented in two blocks. One block contained all English sentences spoken by both English and Dutch speakers, and the other block contained all Dutch sentences spoken by both Dutch and English speakers. The order of blocks was counterbalanced across participants, and each participant heard the sentences in a different randomized order with the restriction that no sentence or speaker was heard twice in a row. Listeners were instructed to rate the sentences on a 9-point scale ranging from 1 (strong foreign accent) to 9 (definitely native speaker). Sentences were only played once, and participants were encouraged to use the full range of the scale for their ratings. The experiment started with two practice sentences, one English and one Dutch sentence, and it took approximately 25 minutes to complete the experiment.

2.2. Results

Figure 1 shows the average ratings for the English sentences by the 3 listener groups. As can immediately be seen in Figure 1, all 3 listener groups rated the English sentences significantly higher (i.e., more native-like) when the sentences had been produced by English L1 speakers than when the same sentences had been produced by Dutch L2 speakers.

Figure 1: Average ratings for English sentences by Dutch listeners, German listeners with Dutch knowledge, and German listeners without Dutch knowledge; 1 = strong foreign accent, 9 = definitely native speaker.



Analyses of Variance (ANOVAs) on the English sentences with the factors *listener group* (with the 3 levels 'Dutch', 'German with Dutch knowledge', and 'German without Dutch knowledge') and *speaker* (with the 2 levels 'English' and 'Dutch') showed a weak effect of *listener group* ($F_1[2,59] = 2.62$, $p > .08$; $F_2[2,10] = 5.13$, $p < .03$), a strong effect of *speaker* ($F_1[1,59] = 806.43$, $p < .001$; $F_2[1,5] = 233.57$, $p < .001$), and a significant interaction ($F_1[2,59] = 5.95$, $p < .005$; $F_2[2,10] = 20.06$, $p < .001$). Subsequent pairwise comparisons revealed that the interaction was mainly driven by Dutch listeners being better at differentiating between native English and non-native Dutch speakers than German listeners with Dutch knowledge were. Even though numerically this difference was rather small (Dutch listeners: 7.5 for L1 speaker and 3.2 for L2 speaker; Germans with Dutch knowledge: 7.3 for L1 speaker and 3.9 for L2 speaker), the interaction was significant. While this interaction suggests that sharing L1 language background with the L2 speaker can make accent detection easier, the lack of such a strong interaction between Dutch listeners and German listeners without Dutch knowledge, on the other hand, suggests that the advantage of sharing L1 language background with the L2 speaker does not always help. While it seems somewhat counterintuitive that familiarity with the accent language did not seem to enhance the ability to detect foreignness for L2 listeners, we want to note that previous studies also varied on whether they found a correlation between familiarity with the accent and ratings or not (e.g., Major, 2007; Munro et al., accepted).

Figure 2: Average ratings for Dutch sentences by Dutch listeners, German listeners with Dutch knowledge, and German listeners without Dutch knowledge; 1 = strong foreign accent, 9 = definitely native speaker.

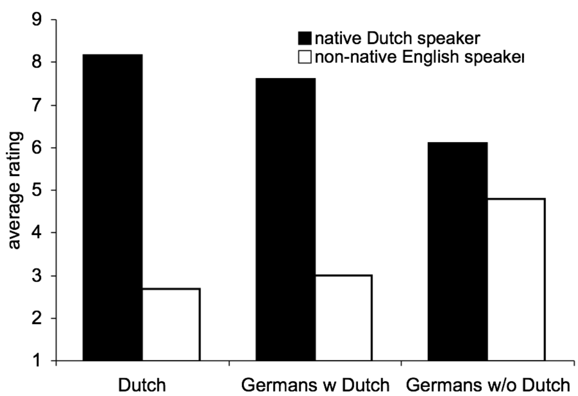


Figure 2 shows the average ratings for the Dutch sentences. As can be seen, Dutch listeners made again the clearest distinction between native Dutch speakers and non-native English speakers: they rated Dutch sentences significantly higher (i.e., more native-like) when the sentences had been produced by native Dutch speakers than when the same sentences had been produced by non-native English speakers. This is not surprising, given that Dutch listeners were rating sentences in their native language. In overall ANOVAs, this time no significant effect of *listener group* ($F_1[2,59] = 1.71$, $p > .1$; $F_2[2,10] = 1.51$, $p > .2$), but a significant effect of *speaker* ($F_1[1,59] = 1324.24$, $p < .001$; $F_2[2,10] = 198.44$, $p < .001$), and a significant interaction ($F_1[2,59] = 136.18$, $p < .001$; $F_2[2,10] = 162.91$, $p < .001$) was found. Further pairwise comparisons showed highly significant interactions of *listener group* and *speaker* between each of the 3 listener groups. Thus, German listeners with Dutch knowledge rated native Dutch speakers lower and non-native English speakers higher than Dutch listeners did. While this difference could simply reflect differences between ratings from L2 and L1 listeners, a comparison with the second group of German L2 listeners showed that knowledge of the target language Dutch further modulated the rating patterns: German listeners without any knowledge of Dutch rated native Dutch speakers still lower and non-native English speakers still higher than German listeners with Dutch knowledge did. This suggests that familiarity with the target language helps L2 listeners to detect foreignness in speech. But as the difference in ratings by German listeners without any knowledge of Dutch further suggests, familiarity with the target language is not a requirement to detect foreignness: even though German listeners from Bamberg did neither speak nor understand Dutch they were not at chance when they had to detect foreignness in the Dutch sentences as a significant *speaker* effect showed ($F_1[1,18] = 54.24$, $p < .001$; $F_2[1,5] = 21.58$, $p < .007$).

3. DISCUSSION

For the English sentences, we found that Dutch listeners, as well as German listeners with and without knowledge of Dutch could distinguish between English and Dutch speakers very well; that is, they all judged sentences from non-native Dutch speakers to be more accented than sentences from native English speakers. Familiarity with the accent language Dutch, however, did not increase German L2 listeners' ability to detect Dutch L2 speakers. In fact, German listeners with Dutch knowledge were slightly less good at detecting L2 speakers than German listeners without Dutch knowledge were. For the Dutch sentences, on the other hand, we found that familiarity with the target language did influence responses strongly. German L2 listeners without any knowledge of Dutch were less good at identifying non-native English speakers than German listeners with Dutch knowledge were (and not surprisingly they were also less good than Dutch L1 listeners were). But importantly, German listeners without any knowledge of Dutch were still better than chance at identifying non-native English speakers in Dutch.

Presumably, native listeners can identify foreign speakers by being able to perceive whether the speech signal deviates from native pronunciation or not. Similarly, proficient L2 listeners may be able to base their judgments on whether or not the speech signal corresponds to the pronunciation norms of the listeners' L2. L2 listeners who are not familiar with the language being spoken, however, cannot base their judgments on recognizing how much the pronunciation corresponds to the target language norms because they do not know what the correct pronunciation of a sound or word in the target language is. How then could German listeners without any Dutch knowledge detect non-native speakers in Dutch?

One possibility is that sentence durations influenced their accent ratings. Non-native speakers often speak more slowly than native speakers, and previous studies have shown that listeners rate more slowly produced speech to be more accented than faster speech (e.g., Munro & Derwing, 1998). Note that sentence durations in the present study were not on average longer when sentences had been produced by non-native speakers than when they had been produced by native speakers. Furthermore, we found no significant correlation between ratings and sentence length (p-values for all correlations larger than .3). But the possibility remains, of course, that single words or syllables within the sentences were produced with a slower speaking rate in the non-native recordings and that German L2 listeners picked up on this cue.

A second possibility is that we underestimated our listeners' knowledge of Dutch. All German listeners from Bamberg stated in a questionnaire that they did not understand or speak any Dutch, but they probably nevertheless had some idea of what Dutch sounds like. If we would ask them whether a spoken Dutch sentence is Dutch or Chinese, we assume that they could reliably decide that the sentence is Dutch. They may also have heard Dutch-accented German before. But then they still should not know whether the sounds and words in the Dutch sentence were produced correctly or bear traces of foreignness.

A third option for how German listeners without Dutch knowledge detected L2 speakers in Dutch sentences has to do with familiarity with the accent language English (either from native English or English-accented German). Our German listeners were highly proficient in English, and they may have based their ratings on recognizing whether the speech signal contained segmental and suprasegmental characteristics of the accent language English or not (still not knowing whether the same characteristics might be Dutch too). We tried to look into this possibility by analysing a subset of our sentences that we judged to be relatively free of segmental accent markers that are typical of an American English accent (e.g., the sentences contained no /r/; Dutch sentences 4-6 in Table 1). Indeed, we found that the difference in average ratings between native and non-native speakers became somewhat less pronounced (while ratings for native Dutch speakers hardly changed, ratings for non-native English speakers rose from 4.8 to 5.3). The remaining difference in ratings was, however, still significant. Thus, perceiving segmental characteristics of the accent language alone probably cannot explain how German listeners without Dutch knowledge detected foreignness in Dutch sentences. This leaves the possibility that listeners responded to something more general in the speech of non-native speakers. Voice quality, for example, has been suggested as a potential marker of nonnativeness, though its role in L2 production has not been thoroughly investigated yet (e.g., Esling 2000). Articulatory effort and carefulness are other potential markers. While with the present study we cannot uncover such possible language-independent, general markers and our results probably reflect a combination of all the above mentioned possibilities, the results of the German listeners make the existence of such general markers more likely. With respect to the question of whether the speech or the listener influence accent ratings, the present findings suggest that it is both. Linguistic experience improves the ability to detect foreignness in speech (speaking for listener effects), but the ability to detect foreignness without experience further suggests a role for the speech signal itself.

4. REFERENCES

Bench J., Kowal A., Bamford J. 1979. The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *British Journal of Audiology* 13. 108-112.

Esling, J. H. 2000. Crosslinguistic aspects of voice quality. In R. D. Kent, M. J. Ball (Eds.). *Voice Quality Measurement* (pp. 25-35), Singular Publishing Group, San Diego.

Flege J. E. 1984. The detection of French accent by American listeners. *Journal of the Acoustical Society of America* 76. 692-707.

Flege J. E. 1988. Factors affecting degree of perceived foreign accent in English sentences. *Journal of the Acoustical Society of America* 84. 70-79.

MacKay I. R. A., Flege J. E., Imai S. 2006. Evaluating the effects of chronological age and sentence duration on degree of perceived foreign accent. *Applied Psycholinguistics* 27. 157-183.

Major R. C. 2007. Identifying a foreign accent in an unfamiliar language. *Studies in Second Language Acquisition* 29, 539–556.

Munro M. J. 1995. Nonsegmental factors in foreign accent: ratings of filtered speech. *Studies in Second Language Acquisition* 17. 17-34.

Munro M. J., Derwing T. M. 1998. The effects of speaking rate on listener evaluations of native and foreign-accented speech. *Language Learning* 48. 159-182.

Munro M. J., Derwing T. M., Burgess C. S. 2003. The detection of foreign accent in backwards speech. In M-J. Sole, D Recasens, J. Romero (Eds.). *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 535-538). Barcelona, Spain: University Autonoma de Barcelona.

Munro M. J., Derwing T. M., Morton S. L. 2006. The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition* 28. 111-131.

Munro M. J., Derwing T. M., Sato K. 2006. Salient accents, covert attitudes: Consciousness-raising for pre-service second language teachers. *Prospect* 21. 67-79.

Munro M. J., Derwing, T. M., Burgess C. S. accepted. Detection of nonnative speaker status from content-masked speech. *Speech Communication*.

Plomp R., Mimpen A. M. 1979. Speech-reception threshold for sentences as a function of age and noise level. *Journal of the Acoustical Society of America* 66. 1333-1342.

Swan M., Smith B. (Eds.). 2001. *Learner English. A teacher's guide to interference and other problems*. Cambridge: Cambridge University Press.

Vowel variation in advanced Polish learners of English

Jarosław Weckwerth

Adam Mickiewicz University, Poznań
wjarek@ifa.amu.edu.pl

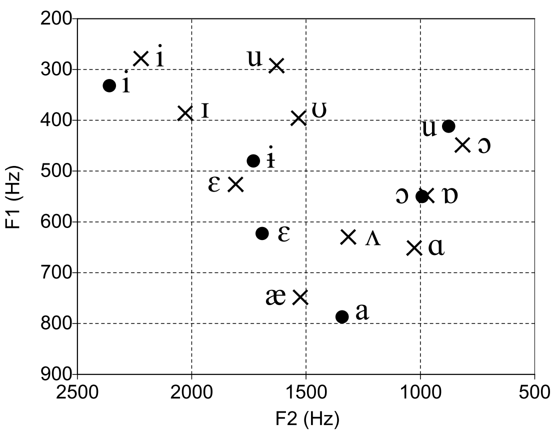
ABSTRACT

The vowels of TRAP, DRESS and STRUT from a wordlist read by 52 female advanced Polish students of English were measured acoustically in terms of F1 and F2. In the entire sample treated as a whole, TRAP showed the most variation, and its distribution overlapped almost completely with those of STRUT and DRESS. Statistical analysis showed that, for the entire sample, there was a significant difference between the three vowels, but a post-hoc test revealed that contrast between TRAP and either DRESS or STRUT was lacking for many subjects. A qualitative analysis of the individual systems showed four patterns: (1) "TREP" systems, where TRAP lacked contrast with DRESS; (2) "TRUP" systems, where it lacked contrast with STRUT; (3) "BIMODAL" systems, where it showed a bimodal distribution, with some instances in DRESS territory, and some in STRUT territory and (4) systems where TRAP formed a more or less separate category. An analysis of the variability within TRAP showed that, for F2, the interaction between preceding and following consonantal context was a stronger effect than following context only, pointing to possible lexical conditioning.

Keywords: Polish L1; English L2; vowels; variation.

1. INTRODUCTION

Figure 1: Standard Southern British English vowels (crosses, based on Ferragne and Pellegrino 2010) overlaid on Polish vowels (based on Bogacka et al. 2006).



Polish has a six-part oral vowel system, as opposed to the English system containing between 9 and 11 monophthongs (depending on the variety). Figure 1 sums up the differences. Virtually all models of L2 phonological acquisition will predict some kind of acquisition/ transfer difficulties in such a situation. This is indeed the case; among the problems faced by Polish learners of English, sources mention e.g. the KIT–FLEECE, FOOT–GOOSE, LOT–THOUGHT and TRAP–DRESS–STRUT[1] contrasts (see e.g. Sobkowiak 2004).

As can be gleaned from Figure 1, English TRAP is located roughly between Polish /a/ and /ɛ/ in the F1–F2 formant space. This makes predictions as to how the vowel will undergo assimilation in the vowel systems of Polish learners somewhat troublesome; descriptive sources are divided on whether it is assimilated to Polish /ɛ/ or /a/. Despite being far less open, STRUT is usually mapped onto Polish /a/. DRESS is the least problematic of the three vowels, mapping straightforwardly onto Polish /ɛ/. Logically, this could lead to a neutralisation of the TRAP–DRESS or TRAP–STRUT contrasts in Polish learners.

The present paper investigates this problem from a variationist standpoint, using acoustic analysis, with a view to inspecting the character of the variation within Polish EFL TRAP.

2. PREVIOUS RESEARCH

In native populations, vowel variation is usually conditioned socially and allophonically. English TRAP is a handbook example: phenomena such as the Northern Cities Shift in the US (e.g. Labov et al. 2006) are well documented. Thus, in the NCS region, all instances of TRAP are raised and fronted, possibly well above

DRESS; in New York (and some other accents of the East Coast), there is a complex pattern of phonetic conditioning, with numerous exceptional lexical items (for a recent study, see Becker and Wong 2009); the vowel has also been subject to investigation in England and Australia, among other places, where the variation is mainly diachronic. The vowel of STRUT is, of course, a rather well-known variable in England, while DRESS has received some attention e.g. in New Zealand.

There has been less research into variation in the vowels of L2 English. Acoustic studies in L2 vowel production usually present the L2 speaker data as means (often group means) and compare those against means from "native speaker" controls. Importantly, (1) intra- and inter-speaker variation in the L2 speakers is seldom inspected in detail; and (2) the identification of the native reference variety is usually rather broad (e.g., "American English" or similar). This is perhaps understandable in the L2 situations that are typically studied, where the speakers acquire their L2 in an immersion setting, in specific local communities, in a country where the L2 is the community language. An excellent example of a study in this vein is Bohn and Flege (1992), where the production of TRAP by German learners in Birmingham, Alabama was inspected, and length of residence was found to improve production.

Some studies relevant to the present one are Jongman and Wade (2007), Leemann (2008), and Cunningham (2008). Usually, the finding is that there is more variation in L2 speakers (even though Morrison 2003 found that English learners of Spanish produce less variability than Spanish native speakers, and interpreted this as a by-product of the density of the L1 system).

3. METHODS

The speech samples used here were taken from an EFL corpus recorded in the School of English, Adam Mickiewicz University, Poznań (Polish *Instytut Filologii Angielskiej*, IFA; hence, IFA Corpus).[2] The aim of the corpus was mainly diagnostic; it was recorded to (1) give the pronunciation teachers at IFA an insight into the pronunciation of students enrolling into the English BA programme at the School; and (2) give both the teachers and students a record of their pronunciation at the beginning of the programme, for possible future reference. As such, it is a convenience sample, not designed with instrumental vowel system analysis in mind. The shortcomings of such an approach will be discussed below.

In all, 127 first-year students were recorded in October 2008, their first month at IFA. The recordings were made as early as possible into the first year to capture their pronunciation in a state largely unaffected by the intensive pronunciation instruction first-years receive at IFA (two 90-minute classes weekly, in addition to other practical English courses, such as Speaking, and a Descriptive Grammar Part 1 course providing a detailed description of the English sound system; all other core courses are taught in English).

For the present study, 52 female speakers were selected. This was done to eliminate any gender-related variation, and facilitate a detailed analysis of the whole group. Detailed demographic data exists only for 30 of the speakers. The mean age was 19.3 years (range 18–22), and the mean onset of English was 9 years (range 6–14). By definition, all of the students had passed their final secondary school exams in advanced English with a minimum score of 80%.

The recordings were made in a sound-proofed room in the studio of the Center for Speech and Language Processing, Poznań. An MXL 770 condenser microphone was used, connected, via an Edirol UA-25 USB audio interface, to a PC computer running Windows XP Professional and the Audacity sound editor. The signal was digitised by the Edirol interface at 44.1 kHz, 16 bits, and saved in uncompressed WAV files.

Each student read a 120-item randomised wordlist (divided into four parts), two ten-item sentence sets, and two short texts; two short samples of spontaneous speech were also obtained from each student. The words in the wordlist were selected with impressionistic analysis of common pronunciation errors in mind. In all, the procedure took about 8–10 minutes per student.

Selected occurrences of the DRESS, TRAP and STRUT target vowels were used. The items were: (1) for DRESS: *bet*, *bed*, *said*, *flesh*, *men*, *head*; (2) for TRAP: *bat*, *bad*, *sad*, *flash*, *man*, *happy*, *trap*, *act*, *practice*, *ran*, *hand*, *band*, *understand*; (3) for STRUT: *but*, *bud*, *flush*, *cut*, *monkey*, *run*. Due to the non-balanced character of the original wordlist, only the *bet–bat–but*, *bed–bad–bud* and *flesh–flash–flush* sets formed complete minimal triple sets. *Men–man* and *said–sad* were minimal pairs for the DRESS–TRAP contrast only. The remaining words were included to increase the number of data points; in particular, *happy*, *trap*, *act* and *practice* served as examples of TRAP in non-alveolar contexts, and the additional pre-nasal TRAP instances were used to explore any possible effects of the pre-nasal context on the allophony of TRAP.

The above is different from the usual approach of including only minimal pairs in a fixed consonantal frame (the most typical being *hVd*, see e.g. Hillenbrand 2003). In this respect, the present approach is similar to that taken by sociphonetics studies using interview techniques, such as those in e.g. Labov et al. (2006) and many other variationist studies, where a perfectly balanced wordlist based around a set of minimal pairs is in fact unusual.

The boundaries of the vowels were annotated in a Praat TextGrid (Boersma and Weenink 2010), and the fundamental frequency and first three formants were measured at the midpoint of each vowel. A small number of anomalous instances was discarded where the speakers evidently did not produce the intended vowels or where the measurements could not be performed satisfyingly. In all, 1274 vowel measurements were analysed, including 864 in non-nasal contexts.

The measurements were subsequently normalised using the NORM normalisation suite (Thomas and Kendall 2007). The "Nearey 1" formant-intrinsic model was used (Nearey 1977).

4. RESULTS

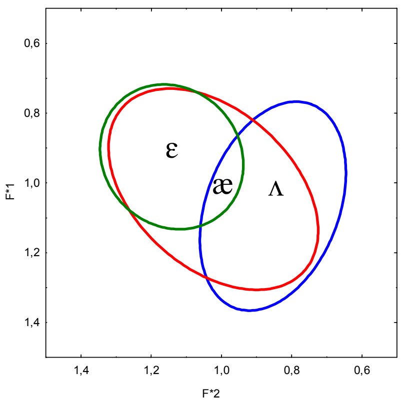


Figure 2. Means (symbols) and .95 confidence ellipses. Log-transformed units as per the Nearey method.

DRESS showed the least variability. The separation between DRESS and STRUT was quite good, with only minimal overlap. TRAP, on the other hand, showed nearly complete overlap with the other two vowels. This was confirmed by a cluster analysis using the K-means method to produce two clusters; DRESS and STRUT were revealed to each cluster within one of the two clusters, with very few miscategorised instances, while TRAP was split roughly in half.

The mean for TRAP was located, in a sense appropriately, between those for DRESS and STRUT, even though STRUT was the most open vowel (albeit by a very small margin). This is in contrast with data from native English – as can be seen from Figure 1, TRAP is normally the most open vowel.

A mixed model analysis of variance components was run on the entire dataset, with F1 and F2 as dependent variables, Speaker as a random predictor, and Vowel as a fixed predictor. Speaker did not prove a significant effect overall, as could be expected after normalisation. There was a significant effect of Vowel, and of the Vowel × Speaker interaction (meaning that the magnitude of the differences between the vowels differed for individual speakers).

The Vowel × Speaker interaction was further investigated using a two-way ANOVA for the entire sample and for each speaker, with a post-hoc Bonferroni test on all pairs of contrasts for both formants. While the ANOVA showed that, for the entire sample, the three vowels were significantly different, there was one speaker for whom a main effect of Vowel was lacking. The Bonferroni test revealed that the situation was much more complex. All speakers had the DRESS–STRUT contrast on at least one formant. Five speakers lacked both the DRESS–TRAP and TRAP–STRUT contrast on both formants. As many as 30 speakers lacked the DRESS–TRAP contrast on both formants; and 12 speakers lacked the TRAP–STRUT contrast on both formants. The remaining five speakers had all the contrasts on at least one formant. Among these, only one speaker displayed a significant difference for all the contrast and both formants. (Of course, due to the low counts of observations for individual speakers, these results must be treated with caution.)

Table 1. Means and SDs for the three vowels.

	F1		F2	
	Mean	SD	Mean	SD
DRESS	0.925305	0.084001	1.141785	0.082704
TRAP	1.017886	0.117288	1.023064	0.121486
STRUT	1.066250	0.121239	0.853546	0.084256

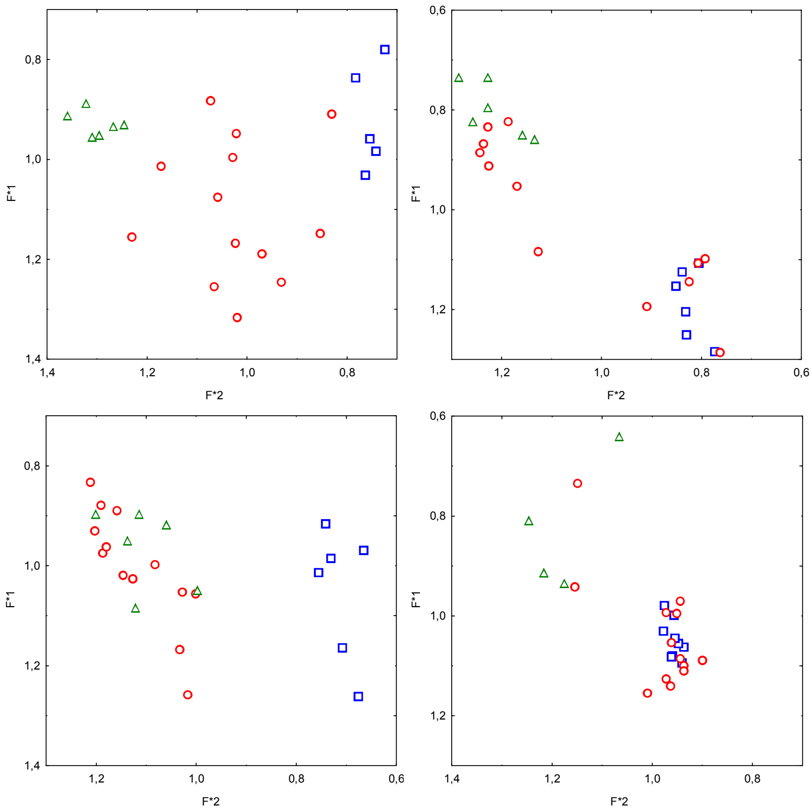


Figure 3. Typical systems: Top left, TRAP; top right, BIMODAL; bottom left, TREP; bottom right, TRUP. Triangles, DRESS; circles, TRAP; squares, STRUT.

This intra-speaker variation was studied qualitatively by visual inspection of individual F1–F2 scatterplots for all speakers. While there were indeed speakers for whom the contrasts were evidently lacking, as suggested by the Bonferroni test, a large proportion of the speakers had a bimodal distribution of TRAP in two clusters, one generally located in or close to the STRUT area, the other – in that of DRESS; conceivably, this could have resulted in either of the two contrasts with TRAP being shown as non-significant by the test above, as a function of the distance between the clusters, and their sizes. There were also systems in which, despite the failure of the Bonferroni test at finding a significant Vowel effect, TRAP instances were visually separate from both STRUT and DRESS.

As a result, a typology of systems was attempted on the basis of the results of the Bonferroni test combined with careful examination of the scatterplots; wherever the results of the test were dubious, the systems were assigned to the TRAP or BIMODAL systems, as detailed below.

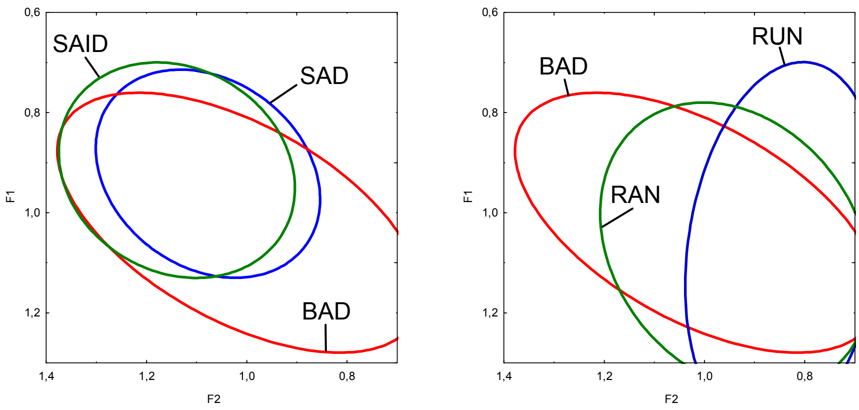
There were 8 systems with general overlap between TRAP and DRESS (informally, “TREP” systems). Four systems displayed overlap with STRUT (“TRUP”). In 17 systems, there was a clear bimodal distribution, with at least four members in each cluster (BIMODAL). The remaining 23 systems did not show any of these patterns (meaning they could be simply termed TRAP systems), even though considerable spread was seen in many speakers. Typical systems from each class are shown in Figure 3.

Since the previous step revealed interesting patterns in the distributions of individual words, a mixed model variance components analysis was performed on the F1 and F2 values for TRAP. Speaker was treated as a random factor, with preceding and following phonetic context (place of articulation) treated as fixed factors. The results are presented in Table 2.

Table 2. Results of a mixed model variance components analysis for TRAP. Df error computed using Satterthwaite’s method.

		Effect (Fixed/Random)	df Effect	MS Effect	df Error	MS Error	F	p
F1	{1} ContextFol	Fixed	3	0.044452	159.4417	0.012298	3.61446	0.014604
	{2} ContextPre	Fixed	5	0.226946	264.6367	0.010946	20.73292	0.000000
	{3} Speaker	Random	51	0.015917	219.6342	0.012541	1.26922	0.124191
	1 × 2	Fixed	3	0.132542	199.0000	0.009175	14.44606	0.000000
F2	{1} ContextFol	Fixed	3	0.027211	161.9790	0.011516	2.36290	0.073208
	{2} ContextPre	Fixed	5	0.204046	267.0266	0.011794	17.30099	0.000000
	{3} Speaker	Random	51	0.012521	191.7677	0.011563	1.08283	0.343801
	1 × 2	Fixed	3	0.156080	199.0000	0.011872	13.14710	0.000000

Figure 4. Distributions of selected individual words. .95 confidence ellipses.



Interestingly, following consonantal context was not returned as a significant effect for F2. However, there was a significant interaction of the preceding and following consonantal context. Since, in the present coding system, this interaction can be argued to be interpretable as "Word", the distributions of individual words were inspected visually. They generally followed the expected context lines, with pre-nasals displaying higher F2 and lower F1, and often clustering within the DRESS area, and instances from other (in particular, pre-bilabial) contexts tending to appear in the lower right of the plots. However, there were notable exceptions. In particular, the word *sad* was frequently found in the DRESS area; in fact, the distribution of *sad* within the whole sample was very similar to that of *said*. The fact that the general distribution of *man* was very similar to that of *men* could conceivably be expected from the nasal context. In contrast, *ran* tended to be located further back and lower on the F1–F2 plane, possibly highlighting the role of preceding context. Figure 4 shows pooled results for all the speakers. Given such distributions, it is not entirely surprising that following phonetic context alone was not found responsible for the variation observed in F2. However, any further investigation could not be attempted with the present limited dataset.

5. DISCUSSION

The present study must be considered a preliminary – perhaps pilot – study of vowel variation in the DRESS, TRAP and STRUT classes in advanced Polish EFL learners of English. The limitations included the following: (1) the wordlist was unbalanced; (2) there was only one reading of each word per speaker; (3) as a result, there were relatively few observations per speaker; (4) other measures, such as vowel length or offgliding, were not taken into consideration; (5) it was not possible to disentangle the effect of phonetic context from any lexical effects. These are all possible avenues for future research.

All the same, the following points are worth highlighting: (1) of the 52 advanced subjects, only five had systems evidently similar to a native distribution; (2) the TRAP=DRESS and TRAP=STRUT "mergers" that could be assumed for Polish L2 English were borne out only partially; (3) there was considerable overlap between TRAP and the remaining two classes; (4) there was considerable variability within TRAP, both inter- and intra-speaker; (5) there was also considerable variation between individual lexical items.

With this small dataset, it is not possible to tap into the possible causes of the variation found in Polish L2 TRAP. However, some speculations may include the following.

The somewhat chaotic variation within TRAP may be simply due to the fact that the English vowel is ambivalent between Polish /ɛ/ and /a/. Importantly, this ambiguity is engrained into the spelling of English borrowings in Polish. Some TRAP borrowings, such as *flesz* 'flash lamp', are mapped onto <e>; others, such as *skan* 'scan', are mapped onto <a>; still others vary either in spelling (*menadżer*/*menedżer* 'manager') or just pronunciation (*Flash* 'Flash graphics technology' is either /flaʃ/ or /flɛʃ/). This latter path of investigation for Polish L2 TRAP was taken by Gonet et al. (2010).

An alternative explanation may be variability in the input (which of course includes the variably adapted borrowings above), as suggested e.g. by Bohn and Bundgaard-Nielsen (2009). Sources of this type may include: (a) interaction with native varieties differing in their treatment of TRAP, including through the media; (b) interaction with non-native varieties differing in their treatment of TRAP; (c) interaction with Polish teachers who may themselves display the variation patterns; (d) attitudinal factors.

Escudero and Boersma (2004), for example, showed that Spanish learners of English differ in their acquisition of the FLEECE–KIT contrast as a function of the variety of English they are exposed to. As mentioned above, TRAP is a good example of variation between native varieties of English. Figure 1 compared the vowels of Polish to those of Standard Southern British English. Of course, such a comparison would be different if e.g. Standard American English was taken as the reference accent, not to mention those accents where TRAP shows more unusual qualities. L2 varieties of English also differ, at least stereotypically, in their treatment of TRAP; in some, e.g. spoken by Spanish learners, the vowel tends to be assimilated to /a/, while in others – e.g. spoken by German learners – with /ɛ/. Again, this is not an easily measurable factor. This type of analysis, however, is more problematic in an EFL population, as meaningfully measuring contact with different varieties of English is difficult, if not impossible, other than through subject self-reports. This leaves the investigation of variability in teacher pronunciation and attitudinal factors as two paths offering more accessible research possibilities.

6. REFERENCES

Becker, K., Wong, A.W. 2009. The short-a system of New York English: An update. *University of Pennsylvania Working Papers in Linguistics* 15(2). (Selected papers from NWAV 37.) < http://repository.upenn.edu/pwpl>

Boersma, P., Weenink, D. 2010. *Praat: Doing phonetics by computer* (Ver. 5.1.05). [Computer program.] <http://www.praat.org/>

Bogacka, A., Schwartz, G., Zydorowicz, P., Połczyńska, M., Orzechowska, P. 2006. The production and perception of schwa in second language acquisition: The case of Polish learners of English. In: Dziubalska-Kołaczyk, K. (ed.), *IFAtuation: A life in IFA. A festschrift for Professor Jacek Fisiak on the occasion of his 70th birthday*. Poznań: Wydawnictwo Naukowe UAM. 71–84.

Bohn, O.-S., Bundgaard-Nielsen, R. L. Second language speech learning with diverse inputs. In: Piske, T., Young-Scholten, M. (eds), *Input matters in SLA*. Bristol: Multilingual Matters. 207–218.

Bohn, O.-S., Flege, J.E. The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition* 14. 131–158.

Cunningham, U. 2008. Acoustic variability in the production of English vowels by native and non-native speakers. In: Waniek-Klimczak, E. (ed.), *Issues in accents of English*. Newcastle upon Tyne: Cambridge Scholars Publishing. 1–15.

Escudero, P., Boersma, P. 2004. Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition* 26. 551–585.

Ferragne, E., Pellegrino, F. 2010. Formant frequencies of vowels in 13 accents of the British Isles. *Journal of the International Phonetic Association* 40. 1–34.

Gonet, W., Szpyra-Kozłowska, J., Święciński, R. 2010. Clashes with ashes. In: Waniek-Klimczak, E. (ed.), *Issues in Accents of English 2*. Newcastle upon Tyne: Cambridge Scholars Publishing.

Hillenbrand, J.M. Illustrations of the IPA: American English: Southern Michigan. *Journal of the International Phonetic Association* 33. 121–126.

Jongman, A., Wade, T. 2007. Acoustic variability and perceptual training: The case of non-native accented speech. In: Bohn, O.-S., Munro, M. (eds), *Language experience in second language speech learning: In honor of James Emil Flege*. Amsterdam: John Benjamins. 135–150.

Labov, W., Ash, S., Boberg, C. 2006. *The Atlas of North American English: Phonetics, phonology and sound change*. Berlin: Mouton de Gruyter.

Leemann, A. 2008. Vowel quality of Swiss EFL speakers. In: Waniek-Klimczak, E. (ed.), *Issues in accents of English*. Newcastle upon Tyne: Cambridge Scholars Publishing. 16–32.

Morrison, G. 2003. Perception and production of Spanish vowels by English speakers. *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona. 1533–1536.

Nearey, T.M. 1977. Phonetic feature systems for vowels. PhD dissertation, University of Alberta.

Sobkowiak, W. 2004. *English phonetics for Poles*. (3rd edition.) Poznań: Wydawnictwo Poznańskie.

Thomas, E.R., Kendall, T. 2007. *NORM: The vowel normalization and plotting suite*. [Online resource.] <http://ncslaap.lib.ncsu.edu/tools/norm/>

Wells, J.C. 1982. *Accents of English*. Cambridge: Cambridge University Press.

NOTES

[1] All English vowel phoneme keywords according to Wells (1982).

[2] The corpus was recorded by Dawid Pietrala and myself. I would like to thank Mr Pietrala for the initial idea and thrust in recording the corpus, and for his invaluable help in preparing and conducting the recordings.

A Computational Model for Accent Identification

Steven H. Weinberger
George Mason University

weinberg@gmu.edu

Stephen A. Kunath
Georgetown University

sak68@georgetown.edu

ABSTRACT

Human listeners have the unique linguistic ability to distinguish the speech of other humans. This ability is assumed to be based upon abstract and implicit linguistic knowledge. This paper aims to identify this knowledge by using a computational procedure that aligns and compares sets of narrowly transcribed speech samples. The procedure results in a finite set of phonological speech patterns. Those samples with similar patterns are grouped into a category that will ultimately be labelled with a native language background. We test the computational procedure against a set of human judgments. This work allows us to determine just which phonological speech patterns are crucial for the proper classification of speech accent. Ultimately, this work will help us to understand what humans are doing when they listen to accented speech.

Keywords: Accent, Computational Phonology, Phonetics, Speaker Identification

1. INTRODUCTION

1.1. Accent Detection

Every normal human is able to listen to the speech of another human and immediately determine whether or not the talker is a member of the listener's speech community (Scovel 1995). For instance, a native speaker of English, born and raised in Pittsburgh, PA, can instantaneously tell that a non-native speaker of English born in Beijing, China, is not a native of Pittsburgh. And because L_2 accents typically reflect the phonological structure of the speaker's native language, we expect that this Pittsburgh listener also should be able to easily distinguish the English of the Chinese talker from the English of a native Arabic talker. Even though ordinary human listeners generally cannot assign the proper native language labels to the talkers, they can nevertheless instantaneously distinguish these different speech samples. Like other linguistic intuitions, most naïve humans cannot adequately explain just how they are able to perform this global accent detection. Indeed, this mystery is real, for as Gut (2007) remarks, "…no exact, comprehensive and universally accepted definition of foreign accent exists" (p. 75). This preliminary study seeks to define global accents and uncover part of the knowledge human listeners have about L2 speech. We will use a computational procedure to define a finite set of distinct phonological characteristics that determine specific accents. That is to say, we want to discover just what it means to have a Chinese accent or an Arabic accent. Various studies deal with how listeners judge and rate non-native speech (Anderson-Hsieh, Johnson, and Koehler 1992; Cunningham-Andersson and Engstrand 1989; Koster and Koet 1993; Magen 1998; Munro, Derwing, and Morton 2006). But these studies have generally dealt with accent rankings rather than with the issue of specific determinations and categorizations of accents. Several computational studies have attempted to automatically sort and classify English speech into accent categories, but these have been limited to just two varieties (Deshpande and Chikkerur 2005). The work we report on here will be describing a computational comparison method that will not rank severity of accents, but rather categorize them into native language classes. We will derive a set of language specific phonological speech patterns for three varieties of non-native English speech.

2. PHONOLOGICAL SPEECH PATTERNS

We assume that humans, when listening to speech, perform some sort of comparison whereby differences are highlighted. These differences form the basis for whatever judgments a listener makes about the talker. In our study, comparisons are made between a native variety of speech and several non-native varieties. The resulting differences we call phonological speech patterns (PSPs), and consist of purely segment and syllable structure generalizations such as vowel shortening, final obstruent devoicing, palatalization, interdental fricative substitutions, vowel epenthesis, and /r/ trilling. They are essentially derived by principled comparisons between phonetically transcribed samples of different English accents. For example, a comparison of Pittsburgh English and the English of a Beijing Chinese talker might result in the set of phonological speech patterns that includes, among others, final obstruent devoicing, vowel epenthesis, vowel shortening and /s/ substitution for interdental fricatives. Our question is whether or not this set of phonological speech patterns is sufficient to define the class of Chinese speakers from Beijing. Certainly some of these phonological speech patterns will be idiosyncratic, and some may be conditioned by the English proficiency of the talker. And yet others may be shared across accent types. But we find that there remains a finite set of patterns that uniquely define Beijing Mandarin Chinese speakers of English.

3. STAT: SPEECH TRANSCRIPTION ANALYSIS TOOL

We have designed a “machine” (STAT) that essentially compares speech samples (Kunath and Weinberger 2009). Conceptually, the machine superimposes, or aligns, one narrow phonetic transcription over another, and the differences are tabulated. In this case, the one transcription is of a native speaker of English, and the other is from a non-native speaker. The alignment is performed using a set of phonetic features that distinguish the difference and relationship between sounds. A key aspect of this is determining a distance measure between two speakers’ pronunciations of the same target utterance. The features used to calculate the distance include the different articulatory descriptions of utterances. Once two utterances have been aligned, a closer comparison begins that identifies patterns of L_2 phonology throughout the utterance. This could be based on a speaker aspirating a stop consonant, which would then trigger the system to look for other instances of aspirated stops. Finally, differences between the transcriptions are returned as PSPs like those listed in section 2.

The STAT system incorporates several distinct components. Users interact with the system primarily via a web interface. All user interfaces are implemented with Ruby on Rails and various JavaScript libraries. Backend processes and algorithms are implemented in Java. We believe that the transcription alignment and speech pattern analysis components of STAT make it a unique tool for linguists studying speech processes. An open source web application bundle including the front-end web interfaces and backend libraries suitable for use in other applications will be made in the future.

4. METHOD

4.1. Accented Speakers

Fifteen non-native English samples were chosen from the *Speech Accent Archive* (Weinberger 2010), a database maintained by George Mason University containing more than 1250 annotated audio samples of speech. Each speaker on the archive reads the same English paragraph. The paragraph is composed of simple words that contain virtually all of the English phonemes. Narrow phonetic transcriptions have been constructed for each speech sample by three trained phoneticians. They are done in Unicode text (see figure 1 for an example of a native English transcription of the elicitation paragraph).

Figure 1: English 1 elicitation transcription.

[pʰl̥iiːz kʰɑlˠ stɛlə æskɚ ɾə bɹĩŋ ðiiːz θĩŋz wɪθɚ fɹʌ̃m ðə stɔɹ sɪks spũunz əv fɹɛʃ snoʊ pʰiiːz faɪːv θɪk sl̥æːbz əv bluː ʧiiːz æn meɪbi ə snæk˺ fɚ hɚ bɹʌðɚ bɑːb wii ɑlˠso niĩːɾə smɑlˠ pʰl̥æstɪk˺ sneɪk æ̃nə bɪːg tʰɔɪ fɹ̥ɑːg fɚ ðə kʰɪːdz ʃii kə̃n skʷuup˺ ðiiːz θɪ̃ŋz ̃ɪntə θɹ̥ii ɹɛːd˺ bæːgz æ̃ːn wii wɪlˠ goʊ miit hɚ wɛ̃nzdeɪ æt˺ ðə tʰɹ̥ẽɪn steɪʃə̃n]

Uniform demographic information for each speaker is also included. We chose five speakers each from the native language backgrounds of Arabic, Russian, and Mandarin Chinese. There were three females and two males from each group. All speakers were independently judged by two native English listeners to show evidence of a non-native accent. These samples were chosen because they represent three major language families, and are therefore regarded as different from each other in structure. They were also chosen because of their relatively abundant representation in the Speech Accent Archive. Table 1 presents the demographic information for the fifteen accented speakers.

Table 1: Accented speaker demographics.

Native Language	Birthplace	Age	Gender	Age of Onset	English Residency
Mandarin 1	Shanxi, China	26	female	13	USA, 2 years
Mandarin 14	Dalian, China	49	male	20	USA, 21 years
Mandarin 2	Nanjing, China	38	female	14	USA, 0.8 years
Mandarin 9	Shanghai, China	38	male	12	USA, 2 years
Mandarin 8	Jingmen, China	29	male	12	USA, 5 years
Arabic 1	Riyadh, Saudi Arabia	38	female	12	USA, 4 years
Arabic 20	Chtaura, Lebanon	36	male	12	USA, 1 year
Arabic 29	Doha, Qatar	19	male	17	USA, 0.3 years
Arabic 9	Jerusalam, Israel	28	female	4	USA, 12 years
Arabic 6	Sanaa, Yemen	21	male	14.5	UK, USA, 2 years
Russian 4	Moscow, Russia	68	female	38	USA, 9 years
Russian 1	Nizhni Novgorod, Russia	37	male	36	USA, 1 year
Russian 14	Pskov, Russia	33	male	18	USA, 6 years
Russian 2	Izmail, Russia	50	female	10	USA, 7 years
Russian 12	St. Petersburg, Russia	62	male	53	USA, 9 years

4.2. STAT Methodology

The fifteen samples of narrowly transcribed speech were computationally aligned with an arbitrary sample of native English from Pittsburgh PA, USA (English 1 in the *Speech Accent Archive*). STAT produced alignment for each sample that included both word and phoneme level information. Each word alignment from STAT indicated which PSPs occurred as well as the placement in the word. Additionally, STAT produced a report that summarized all PSPs that occurred throughout the entire sample. The final report containing the total counts for each PSP was what was used for the purposes of this study.

4.3. Human Judges

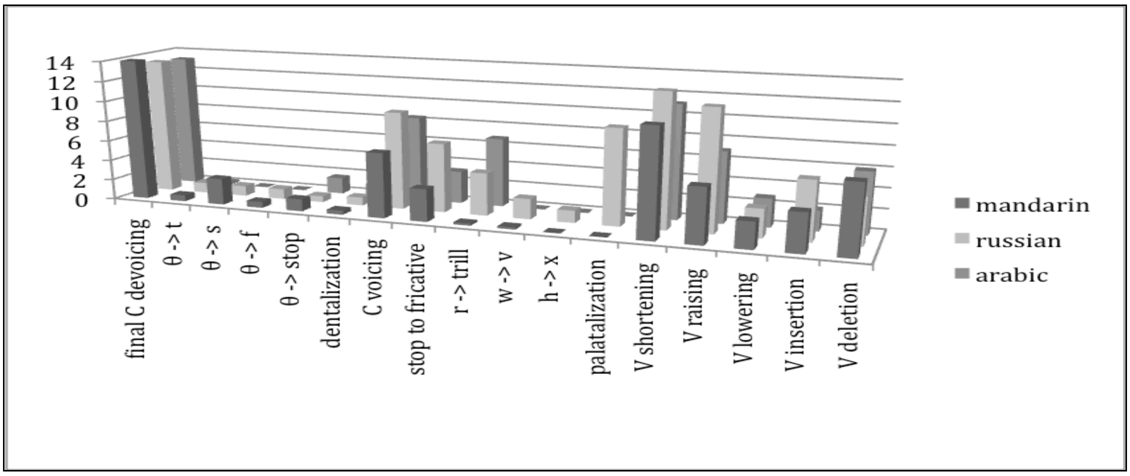
As a diagnostic to the computational approach, we tested American native-English speakers on how they categorized the same fifteen accented samples. This was done via an Amazon.com Mechanical Turk human intelligence task (HIT). The HIT included a baseline question where judges, or workers in this case, had to rate the level of accent of a native English speaker (English 1 in the archive) and also a heavily accented speaker (Spanish 10). Consistent ratings for this baseline judgment task showed that judges had no trouble in identifying accented English speech. Judges for each task listened to each accent sample through a web interface and were instructed to group each sample into one of three unnamed categories. There were 50 judges, 23 females and 27 males. Average age was 31.5, and their place of U.S. birth was evenly distributed across the country. All had higher education degrees. Twelve judges were college students and the remainder held professional positions. The workers completed the task in less than 10 minutes on average. After the task was submitted the results for each worker was analyzed to ensure that groupings were not random. Once workers' submissions were judged to be of high quality they were credited with a reward of $0.75 for their participation.

5. RESULTS

5.1. STAT Results

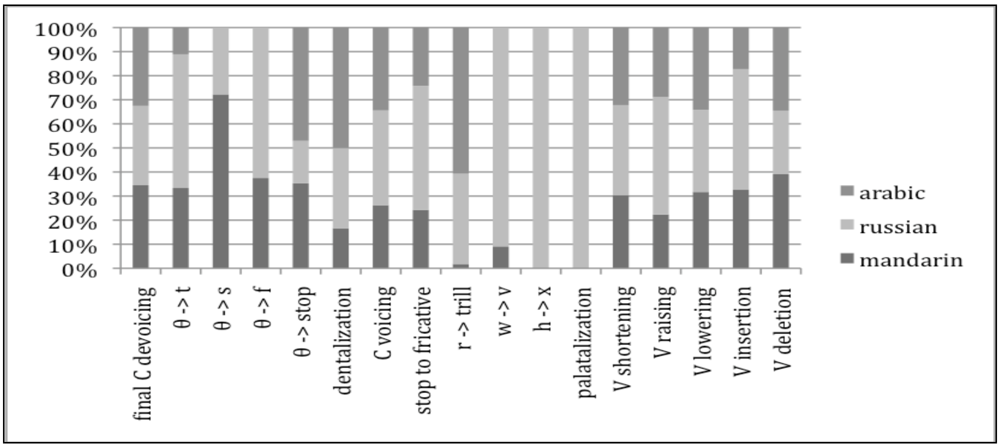
Automated alignment and computational comparison between each accented sample and the native English sample resulted in 17 PSPs. These PSPs reflect the behaviors of each non-native speaker of English. Figure 2 presents the mean results for the three native language categories. It appears that certain PSP behaviors were shared across language categories. For instance, most of the speakers, regardless of native language background, devoiced their final obstruents and shortened their vowels. These are indeed common L_2 phonological behaviors and would most likely contribute to a judgement that the owner of these English behaviors has a foreign accent. But by themselves, these two PSPs do nothing to help distinguish between accents. What we need for that is some indication that one native language group uses a relatively higher proportion of some PSPs while other groups do not. We can see some indications of this situation as we examine palatalization, /r/ trilling, vowel raising, and substitution of alveolar fricatives for interdental fricatives (θ → s).

Figure 2: Mean use of PSPs for each accented language group.



From figure 2, we see that the Arabic group is using more /r/ trilling than the other groups. The Russian group seems to be monopolizing palatalization, and to a lesser extent, these speakers are raising their English vowels. The Mandarin Chinese group is substituting alveolar fricatives for interdentals more than the other groups. When we view these results from a 100% proportional perspective, some of these accent differences become clearer. This is shown in Figure 3:

Figure 3: Proportions of PSP behavior for each accented language goup.



While we remain cautious about some of these generalizations, figure 3 shows that Russians are uniquely distinguished by their high proportion PSPs: h→x, palatalization, w→v, θ→f, θ→t, stop→ fricative, and vowel raising. The Arabic group has high proportions of /r/ trilling, θ→ non-alveolar stop, and dentalization. Mandarin Chinese speakers of English in this study show high proportions of θ→s. The remaining six PSPs: final devoicing, consonant voicing, vowel shortening, vowel lowering, vowel insertion and vowel deletion are relatively uniform across all of these non-native speakers.

5.2. Human Judges Results

Now if humans indeed have the ability to categorize different accents into distinct categories, and further, if this ability is as perfect as simply *noticing* an accent, i.e. determining whether or not the talker is a member of the listener's speech community, then we should expect our 50 native English judges to unerringly group all five Mandarin Chinese speakers together into a unique category, and do the same for the Russian and Arabic speakers of English. But as shown in figures 4 and 5, our judges did not do this. They did well, but not all achieved perfect scores.

Figure 4: Mean category scores for native English judges (5=perfect score).

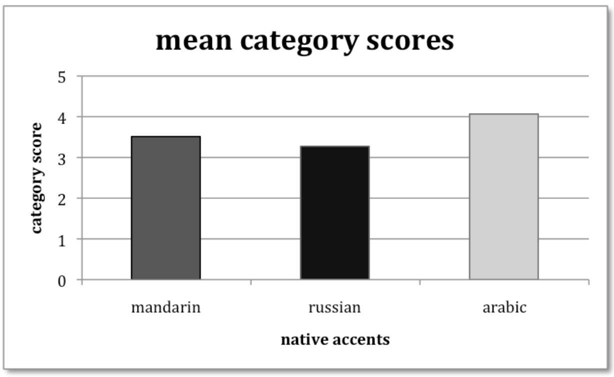
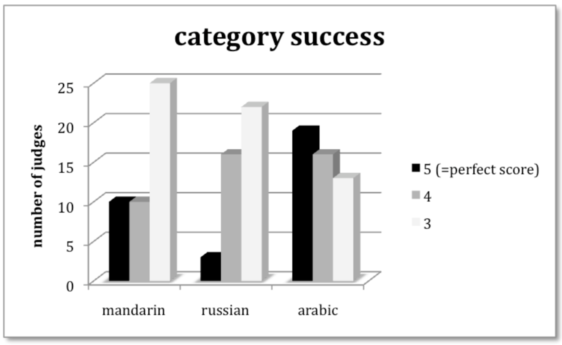


Figure 5: Number of Judges who correctly categorized 3 or more accents



As suggested by the data in figures 4 and 5, a higher proportion of our human judges are most successful when categorizing the Arabic accented English. Fully 35 out of 50 judges grouped four or five of the five Arabic accented samples together. This is compared to 20 of the judges who successfully categorized the Mandarin speech, and 19 who did it for the Russian accents.

6. CONCLUSION

Clearly, the three sets of speech accents in this preliminary study were shown to be categorically different from each other. The results from STAT and from our human judges appear to converge on this. The STAT analysis revealed that the accented English speech from Mandarin Chinese, Arabic, and Russian backgrounds resulted in sets of PSPs that could indeed be used to define particular classes of accent. But there were also overlapping sets. To us this suggests that while some types and proportions of PSPs may be language-specific (like palatalization), others may be more cross-linguistic (such as final obstruent devoicing). While they all serve to advertise a foreign accent, unique constellations of these PSPs seem to function as indicators for specific accents.

The PSPs were segmental in nature. This study made no attempt to analyse intonation, stress patterns, or voice quality. And while the suprasegmental structure no doubt carries important identifying information about an accent, the results we obtained from a purely segmental analysis seemed sufficient to correctly categorize the speech samples. But we should note that there may be a "weighted effect" of some of the PSPs. That is, some PSPs may carry more identifying information than others. This may account for the lack of a one-to-one relationship between total number of language-specific PSPs used by an accent category and the results of the human judges. Recall that while the Russian accented group had a constellation of seven major PSPs associated with it (h→x, palatalization, w→v, θ→f, θ→t, stop→fricative, and vowel raising), the Arabic accented group only had three major PSPs associated (r-trilling, θ→ non-alveolar stop, and dentalization). Yet the human judges demonstrated their highest scoring with the Arabic group, and their lowest scoring with the Russian group. Could it be that r-trilling, θ→ non-alveolar stop, and dentalization are individually, or as a constellation, more salient speech behaviors? A more detailed determination of this weighted effect awaits further work.

Currently, the STAT system is being extended to utilize all transcriptions available through the Speech Accent Archive. The paradigm is also being evaluated to determine the minimum number of speech transcription samples needed to successfully identify a speaker's native language background and degree of accent.

7. REFERENCES

Anderson-Hsieh, Janet, Ruth Johnson, and Kenneth Koehler. 1992. The Relationship Between Native Speaker Judgments of Nonnative Pronunciation and Deviance in Segmentals, Prosody, and Syllable Structure. *Language Learning* 42, no. 4: 529-555.

Cunningham-Andersson, Una, and Olle Engstrand. 1989. Perceived Strength and Identity of Foreign Accent in English. *Phonetica* 46: 138-154.

Deshpande, Shamalee, and Venu Chikkerur. 2005. Accent Classification in Speech. In *Proceedings of the Fourth IEEE Workshop on AUtomatic Identification Advanced Technologies*. IEEE Computer Society.

Gut, Ulrike. 2007. Foreign Accent. In *Speaker Classification*, ed. Christian Muller, I:75-87. Berlin: Springer-Verlag.

Koster, Cor, and Ton Koet. 1993. The Evaluation of Accent in the English of Dutchmen. *Language Learning* 43, no. 1: 69-92.

Kunath, Stephen, and Steven Weinberger. 2009. STAT: Speech Transcription Analysis Tool. In *Proceedings of NAACL HLT 2009: Demonstrations*, 9-12. Boulder, CO: Association for Computational Linguistics.

Magen, Harriet. 1998. The Perception of Foreign-accented Speech. *Journal of Phonetics* 26: 381-400.

Munro, Murray, Tracey Derwing, and Susan Morton. 2006. The Mutual Intelligibility of L2 Speech. *Studies in Second Language Acquisition* 28: 111-131.

Scovel, Thomas. 1995. Differentiation, Recognition, and Identification in the Discrimination of Foreign Accents. In *Phonological Acquisition and Phonological Theory*, ed. Archibald. Hillsdale NJ: Lawrence Erlbaum.

Weinberger, Steven. 2010. *Speech Accent Archive*. George Mason University. http://accent.gmu.edu.

L2 phonological acquisition by young learners: Evidence from production

Henning Wode

English Department, Kiel University, Germany

hb.wode@t-online.de

ABSTRACT

Data from different kinds of immersion (IM) programs in Germany are reviewed to determine the nature of the transfer-based segmental substitutions by young learners. The age range covered is 3;0-10;0. It is argued that (a) what determines the nature of these substitutions is the state of development of the learner's L1; and (b) that, these restrictions aside, the 3-year olds produce the same substitutions as older learners including adults.

Keywords: Production, transfer-based substitutions, young children, immersion

1. INTRODUCTION

This is an interim report on a long-term endeavour to characterize the development of the pronunciation of a non-native language with (very) young children. The key issue is, whether the abilities of human beings to acquire the pronunciation of a non-native language do, or do not, change as a function of age. The age issue has traditionally been discussed in terms of various notions of critical period(s). One weak point in this debate is that because of the lack of appropriate data we do not have an adequate understanding of what the beginnings are with respect to young children. To be sure, the research on infant speech perception in both L1 and L2 acquisition has clarified many aspects in the perceptual domain; but very little is known about the beginnings and/or the development of L2 production with very young children in order to be able to state which properties may change as a function of age. In this paper transfer-based segmental substitutions in production are singled out for closer scrutiny.

2. METHODOLOGY

2.1. IM in Germany

IM has been surprisingly slow to catch on in Germany, in particular, with preschools and primary schools. The first IM programs were established towards the end of the 1960s. They are extended core programs involving only a low dose of late partial IM, and they were meant to serve schools at secondary age ranges. These programs start in grade 5 (age 10) as traditional language-as-subject instruction (LAS). LAS may continue till the end of secondary school (age 18); but in grade 7 an IM component is added in that two of the regular subjects, in general geography and history, are taught entirely in the new language (details Wode 1995, 1998).

Preschool IM and IM in primary schools are just beginning to get established in answer to the 3-language formula as the official language policy of the EU. This formula states that every child is to have the opportunity to learn at least three languages at a functionally appropriate level during his/her time in school (e.g. European Commission 2004, Wode 2009a).

The data for this paper come from English IM programs in Kiel, northern Germany. The children can start at age 3 in preschool via 100% IM, i.e. English is the only language used throughout the entire day. Upon entry into primary school at age 6 IM continues in such a way that all subjects except German are taught in English resulting in appr. 70% of the teaching time for English and only 30% for German. A small number of children without any knowledge of English is also admitted to the primary school IM classes so that these students get their first exposure to English at age 6. They are not given any special treatment; they are expected to catch up on their own – and they certainly do. By the end of grade 4, in most cases even earlier, these children have become indistinguishable from the others on the basis of their English, including their phonology (details Wode 2009a).

As for the beginning of exposure to English, these IM programs, if taken together, allow for three entry points, hence three age groups, namely, age 3 (beginning of preschool), age 6 (beginning of primary school for the children with no prior knowledge of English), and age 12 (the beginning of IM in the late partial IM program after the children have had two years of LAS.

The data for the late partial IM program was elicited by asking triads of students to discuss how they would handle a difficult situation on a class trip to the Scottish Highlands (e.g. Wode 1994, 1998). The primary school data come from picture narratives of the story *Frog, where are you?* (Mayer 1969). This test is administered at the end of each grade level. The preschoolers are tested in a variety of ways including child games, enacting specific roles in puppet shows, identifying objects on picture cards, etc. (details Wode 2009a).

The data are analysed in such a way that the approach can be applied to all age ranges. It is based on determining the frequency of occurrence for each sound/substitution in terms of the percentages relative to the total number of occurrences of the target phoneme. Changes in the frequencies over time/age are indicative of development.

3. RESULTS

3.1. Primary school (6;0-10;0)

Table 1 lists some of those English phonemes that notoriously cause problems for German L2 learners of all age ranges, because these distinctions do not exist in German. Amongst other things, German lacks /θ ð æ dʒ əʊ ɛɪ/; the German /r/ is uvular [R] or [ʁ]; there is no velar [ɫ]; there are no retroflex vowels; and no syllable-final voiced obstruents. Table 1 illustrates two kinds of substitutions, namely, those that result from transfer from the L1 and those that do not. The latter comprise the substitutions for /r/ except [ʁ]; the former all the others including the vowels and [ʁ] as a substitute for /r/.

Recall that it is not only the IM preschoolers that are admitted to the IM program. To be sure, the majority of the children in the IM class have also attended the IM preschool, so that they already have some knowledge of English. Therefore they are referred to as the B-(bilingual) children. However, a smaller group of children are also admitted, although they do not have any knowledge of English at all. They are therefore termed the M-(monolingual) children. As already stated above, the latter do catch up with the B-children by the end of primary school.

Three points seem particularly noteworthy about Table 1: First, as expected, there is a considerable amount of variation. This may be due to the mix of situations that triggered the utterances just as much as to the state of L2 development of the children. Second, in spite of the age differences when first exposed to English there are no striking differences in the structure of the substitutions found across the children of both groups. That is, there is not a single substitution that can be shown to occur only in one group. Age at first exposure to the L2 does not seem to have an impact on the structural properties of the children's production. Moreover, a final point can be added, although it is not displayed in Table 1. In terms of the percentages of target-like vs. non-target-like renditions the M-children trail the B-children by quite a margin in grade 1. This gap, however, is continuously narrowed down as the children progress towards grade 4. In fact, by the end of grade 4 the two groups have become indistinguishable on the basis of their English. This implies that the M-group must have acquired English at a considerably faster rate than the B-group.

Table 1: Percentages of target-like (tl) and non-target-like renditions of some of the sounds of English that are problematic for German learners according to grade level. The substitutions are ordered according to frequency of occurrence. - = deleted. (Wode 2009b adapted from Sieg 2004).

grade **age**	**1** **7;0**			**2** **8;0**			**3** **9;0**			**4** **10;0**		
L2 target	**tl**	**sub**	**%**	**tl.**	**sub**	**%**	**tl**	**sub**	**%**	**tl**	**sub**	**%**
/ð/	29,5	[d] [z] [-]	59,1 7,6 3,8	45,2	[d] [z] [-]	33 19,7 2,1	71,9	[d] [z] [θ] [-] [tθ] [s]	17,7 8 1,5 0,5 0,3 0,1	79	[d] [z] [θ] [dð]	17,3 3,1 0,5 0,1
/θ/	69,2	[s]	30,8	80	[s] [t]	15 5	78,4	[s]	21,6	85,7	[s] [f]	13,3 1
/r/	72,3	[w] [ʁ] [ɹw] [r]	23,2 1,9 1,6 1,1	88,9	[ʁ] [w] [r] [-] [l]	4,6 4,3 1,4 0,5 0,2	73,5	[-] [w] [ʁ] [ɹw] [v]	12,7 8,7 3 1,9 0,2	84,2	[-] [w] [ɹw] [ʁ] [v]	11,4 2,1 1,1 1 0,2
/w/	93,3	[v]	6,7	94,5	[v] [ɹ]	4 1,4	98,9	[v] [ɹ] [ ]	0,8 0,2 0,1	99,1	[v] [ɹ] [ɹw]	0,6 0,2 0,1
[ɫ]	57,1	[l]	42,9	38,9	[l]	61,1	49,6	[l]	50,4	54,7	[l]	45,3
/v-/	50	[w] [f]	30 20	33,3	[w]	66,7	44,2	[w]	55,8	94	[w]	6
/æ/	8,9	[ɛ] [a]	89,8 1,3	26,3	[ɛ] [a]	73,1 0,6	20,2	[ɛ] [a] [ɔ] [ɛɪ]	78,5 0,5 0,5 0,3	17	[ɛ]	83
/ɛɪ/	100	-	-	98,9	[e]	1,1	98,3	[ɛ]	1,7	99,1	[ɛ]	0,9
/ɔʊ/	93,2	[o]	6,8	88,4	[o] [ɔ] [oa]	10,2 1 0,3	88,7	[o] [ɔ] [aʊ]	10,8 0,3 0,2	84,2	[o] [ɔ] [aʊ]	14,9 0,6 0,3

3.2. Preschool (3;0-6;0)

Table 2 is structured in the same way as Table 1 in order to facilitate comparisons between the groups preschoolers and the primary school children.

Table 2: Inter- and intraindividual variation for 10 preschoolers for those L2 English targets that are problematic for German learners. A5, A6, etc.= serial number for preschoolers. ( ) = number of tokens (Wode 2003).

L2 target	**A5**	**A6**	**A8**	**A9**	**A10**	**A11**	**A13**	**A14**	**A15**	**A22**
/θ/	ts (2) s (1)	θ (4) s (3) t (2) tθ (1) ts (1)	θ (8) t (2) ʃ (1)	s (7) θ (1) ts (1)	θ (3) ts (3) d (1)	θ (4) s (2) z (1)	ts (5) θ (2) s (1) ʃ (1)	t (6) θ (1) s (1)	d (12) θ (2) s (2)	f (5) θ (3)
/w/	w (4)	w (9)	w (8)	w (13) v (1)	w (5) v (1)	w (4) v (1)	v (6) w (1)	w (7)	w (8) ɹ (1)	w (6)
/v/	f (3) v (2)	f (4) v (3)	f (4) v (3)	f (6) v (5)	f (4) v (1)	v (4) f (3)	f (3) v (3)	f (4) v (3) w (1)	f (5) v (5)	f (3ll) v (3)
[ɫ]	l (10)	l (11)	l (14)	l (11)	l (18)	l (10) ɫ (2)	l (6) ɫ˃ (1)	l (9) ɫ (1)	l (18) ɫ (1)	l (12) ɫ (3)
/r/	w (10) v (3) ʁ (2)	w (9) v (8)	ɹ (8) w (4)	v (6) ʁ (5) w (2) ɹ (1)	w (5) ɹ (4) v (3) ʁ (2)	w (4) v (2) ɹ (2) r (2) ʁ (1)	v (6) w (2) ɹ (1) l (1)	v (5) w (5)	ɹ (11) w (10) ʀ (1) ʁ (1)	w (9) ɹ (8) ʁ (1)
/æ/	ɛ (4) a (2) ɛːɐ (1)	ɛ (5)	ɛ (5) æ (3)	ɛ (6) æ (1)	a (4) ɛ (1) æ (1)	ɛ (4) æ (1) i (1) ʌ (1) a (1)	ɛ (2) æ (1)	ɛ (2) a (2) ʌ (1)	ɛ (7) æ (1)	ɛ (5)
/ɔʊ/	oʊ (1) ɔʊ (1) o (1) ɔ (1)	ɔʊ (1) o (1) ɔ (1)	o (2)	o (1)	ʊ (2) ɔʊ (1) o (1)	ʊ (2) o (1)	ʊ (2) oʊ (1)	ʊ (2) ɔʊ (1)	ʊ (2) aʊ (1) ɔ (1)	ʊ (3) ɔʊ (3) ɔ (1)
/ɛɪ/	ɛ (3) e (2) i (1) ɛɪ (1)	ɛɪ (3)	ɛɪ (5) ɛ (2) e (1) i (1)	e (1)	ɛɪ (2) e (1)	ɛɪ (2) e (1) i (1)	ɛɪ (2)	ɛɪ (2) e (1)	ɛɪ (4) e (1)	ɛɪ (3)

Given the topic of this paper, two points are particularly interesting in comparing Table 1 and 2. One is that the two tables do not provide any compelling evidence to support the idea that the L2 development of the younger children differs radically from that of the older ones. For one thing, the range of structural variation with the former group is at least as large as with the latter. The other point is that the substitutions, i.e. both the target-like ones and the non-targetlike ones are much the same with both age groups; and this includes the transfer-based segmental substitutions.

The second point relates to the nature of the L2 developmental process. It appears that the beginning with both the younger and the older learners is marked by a larger range of substitutions. As the process moves forward this range of variation is narrowed down in such a way that one variant increasingly dominates and that the frequency of occurrence the other variants is reduced. The kind of situation is particularly in evidence with the B-children in primary school.

4. CONCLUSION

What is still needed is some empirical confirmation that may justify the claim that the state of development of the L1 determines the structure of the transfer-based segmental substitutions even with young children. One such case is reported by Wode (1981) for a child aged 3;11 upon first contact with English. She is one of the four L1 German children whose L2 phonological development has been studied intensively at Kiel University. At the time of first contact with English she did not yet distinguish between /s z (t)ʃ (d)ʒ/ in her L1 production; she replaced them by /θ ð/, as in [θøn] instead of [ʃøn] *schön (beautiful)*, [ʀɑðə] instead of [gaʀɑʒə] *Garage (garage)*, [vaɪθ] instead of [vaɪs] *weiß (white),* [ðɒl] instead of [zɒl] *soll (shall, ought to).* The dental nature of the fricatives was due to the fact that the child had a lisp. She transferred this pattern into her English resulting in, e.g. *Johnny* [ðɒni] (0;4 (zero months, 4 days)), *Ginger* [ðɪnða] (0;9), *yes* [jɛθ] (0;18) *much* [mat^{h}θ] (0;27), *fishing* [fɪθɪŋ] (0;30), guys [gaɪθ] (1;2).

This evidence suggests that the kind of transfer familiar from older learners is already part of the L2 learning process by age 4. The new data from IM preschoolers allows us to suggest that the regularities of transfer from the L1 are already operative around 3;0. Obviously, we need to look at children younger than 3;0, such as toddlers in day care centers to determine the onset of transfer in L2 acquisition.

5. REFERENCES

European Commission 2004. *Promoting language learning and linguistic diversity: An action plan 2004-06*. Luxemburg: Office for Official Publications of the European Communities

Sieg, A. 2004. *Die Entwicklung der Phonologie von der dritten zur vierten Klasse des erste immersiv auf Englisch unterrichteten Jahrgangs der Claus-Rixen-Schule*. MA, Kiel University

Mayer, M. 1969. *Frog, where Are You?* New York: Pied Piper1981

Wode, H. 1981. *Learning a Second Language: An Integrated View of Language Acquisition*. Tübingen: Narr

Wode, H. 1995. *Lernen in der Fremdsprache: Grundzüge von Immersion und bilingualem Unterricht*. Ismaning: Hueber

Wode, H. 1998. A European perspective on immersion teaching: The German scenario. In: J. Arnau & J.M. Artigal (ed.) *Els Programes d'immersió: una Perspectiva Europea – Immersion,* 43-65

Wode, H. 2003. "Young age" in L2 acquisition: The age issue in reverse in phonology. In: Lidia Costamagna & Stefania Giannini (ed.) *La fonologia dell'interlingua: Principi e metodi di analisi*. Pavia: Francoangeli, 71-92

Wode, H. 2009b. Developing non-native pronunciation in immersion settings. In: Piske, Th. & Young-Scholten, M. (ed.) *Input matters in SLA*. Bristol et al.: Multilingual Matters, 238-256

Wode, H. 2009a. *Frühes Fremdsprachenlernen in bilingualen Kindergärten und Grundschulen*. Braunschweig: Westermann

Reinvestigating sound-colour mappings in L1 and L2 vowel perception

Magdalena Wrembel[1] *& Andrzej Grzybowski*[2]

[1]Adam Mickiewicz University, Poznań, [2]Dept. of Ophthalmology, Poznań City Hospital, Poland

magdala@amu.edu.pl, ae.grzybowski@gmail.com

ABSTRACT

The present study is a continuation of previous investigations into the nature of sound-colour associations in a non-synaesthetic population and it aims to provide further evidence for the non-arbitrary nature of cross-modal mappings. The study involved a diagnosis of the participants' colour vision by means of specialised ophthalmological tests in order to exclude any chromatic discrimination disorders. The experiment was run on a specially designed computer program and involved 40 Polish participants proficient in English who were asked to match the randomised auditory stimuli (i.e. English and Polish vowels presented in different testing blocks) with one of 30 coloured rectangles displayed on a computer screen. The analysis of results revealed the statistical significance of sound-colour interactions for all the Polish vowel sounds under investigation and for 9 out of 12 English vowels ($p<.01$). Strong patterns of cross-modal correspondence were also found for specific colour attributes (HSL - hue, saturation and lightness). The findings indicate that vowel-sound mappings in L1 and L2 non-synaesthetic perception appear consistent and non-arbitrary, and follow general tendencies in which bright colours are associated with high front vowel sounds, dark colours are attributed to back vowels, whereas open sounds tend to be perceived as red and central vowels are mapped onto achromatic grey.

Keywords: Vowel colour, sound-colour mappings, cross-modal perception, colour vision.

1. THEORETICAL BACKGROUND

The paper advocates a non-traditional approach to the study of L1 and L2 speech perception that is mediated through colour and relies on cross-modal audio-visual mappings that are reported to be common also in the non-synaesthetic population (cf. Ward et al. 2006). The study is intended as a follow-up of previous investigations into sound-colour associations conducted on English and Polish vowel sound systems (cf. Wrembel 2007, Wrembel and Rataj 2008, Wrembel 2009). It aims to provide further evidence for the non-arbitrary nature of cross-modal mappings and to demonstrate their potential implications for L2 pronunciation pedagogy. The study is based on the assumptions stemming from research on synaesthesia, sound symbolism and non-modularity of human perception.

1.1. Cross-modality and colour perception

The mechanism of colour perception has intrigued scholars from antiquity, however, the earliest concepts were based on philosophical considerations and theoretical speculations rather than anatomical studies (cf. Grzybowski et al. 2008 for a thorough historical outline thereof). A groundbreaking contribution to the understanding of the true nature of colours was presented by Newton in 1688, however, it was only in the 19th c. that two modern theories of colour were proposed including a trichromatic theory by Young and Hemlholtz and an opponent colours theory by Hering. These contradictory theories were eventually reconciled in the 20th c. zone theories of colour vision by Hurvich and Jameson (cf. Grzybowski et al. 2008: 118-120).

The assumptions of the present study pertain to hypotheses put forward by various researchers including Pythagorus and Sir Isaac Newton, about the existence of a physical relationship between the frequencies of light and sound that are responsible for the sub-modality of colour and pitch (cf. Lyons 2001). The concept of non-modularity of human perception has received a lot of support from recent behavioural and brain imaging studies suggesting that cross-modal interactions are common in normal perception and that the cortical pathways previously considered to be sensory-specific are modulated by signals from other modalities (cf. Shimojo and Shams 2001).

1.2. Vowel colour

The colour terminology with respect to vowels was first applied by Jakobson (1962), who identified some regularities in vowel-to-colour associations on the basis of case studies in coloured hearing synaesthesia. Jakobson claimed that there is “a phenomenal affinity between optimal chromaticity (pure red) and vocalic compactness, [between] attenuated chromaticity (yellow-blue) and vocalic diffuseness, optimal achromaticity (black-white) and consonantal diffuseness, attenuated achromaticity (greyed) and consonantal compactness; and, finally, between the value axis of colour (dark-light) and the tonality axis in language" (1962: 488 footnote). Moreover, Jakobson encouraged further research into associations between phonemic features and colour attributes in order to provide insights into the perceptual aspects of speech sounds. According to his proposal, chromaticity corresponds to the vertical axis of the vowel chart, thus the maximally open vowels (i.e. compact in acoustic terms) are regarded as maximally chromatic, i.e. red. On the other hand, the light vs. dark distinction seems to be related to the horizontal position of the tongue and as chromaticity diminishes for close vowels, this distinction appears to be particularly significant.

Jakobson’s hypotheses correspond to a large extent to Marks’ observations (1975) based on case reports of sound-colour synaesthesia. Marks (1975) noted that the black vs. white distinction is related to vowel ‘pitch’, while the red-green distinction correlates with the ratio of F2 to F1. Moreover, Jakobson’s (1962) claim is in line with synaesthetic research results that point to a strong correlation between auditory pitch and visual luminance as well as a general tendency to associate high pitch sounds with light colours and low tones with darker hues (e.g. Simner et al. 2005, Ward et al. 2006).

Another scholar who used the term vowel colour to refer to such features as palatality and labiality was Donegan (1985). She claimed that palatal (i.e. front) vowels tend to be perceived as 'bright' as opposed to 'dark' labial vowels which are characterised by low F2. Vowels which are neither palatal nor labial are called plain or achromatic (Donegan 1985: 66-67).

1.3. Research into sound-colour mappings

Cross-modal mappings between auditory and visual stimuli have been extensively explored in research on synaesthesia (e.g. Day 2004, Simner et al. 2005, Ward et al. 2006), however, relatively few studies to date have investigated this phenomenon in non-synaesthetic perception.

In a perceptual similarity experiment performed on Japanese subjects Miyahara et al. (2006) examined the patterns of correspondences between 4 selected colours and 5 recorded vowels. The findings revealed clear regularities; red was chosen for /a/ significantly more than any other vowel, yellow was chosen for /i/ significantly more than any other vowel, blue was chosen for /o/ significantly more than any other vowel and for /e/ significantly more than for /i/ and /a/. There was also an effect of gender and pitch of the presented stimuli on the colour choice. Furthermore, Miyahara et al. (2006) compared their results to patterns of vowel-colour mappings for the synaesthetic population extracted from Day's (2004) study and identified similar patterns of correspondence, with the ratios of colour-vowel choices being slightly lower for the non-synaesthetic group.

Smith et al. (forthcoming) reported on a study in which English speaking non-synaesthetes were tested on colour associations with exemplars of IPA cardinal vowels. The open vowels demonstrated high consistency responses related to maximally saturated red hues; front-unrounded vowels elicited higher luminance values than back-rounded vowels. Some indication was found for the role of hue with blues and purples selected more often for back rather than front vowels.

The present author’s previous findings revealed a visible tendency for vocalic openness to generate optimal chromaticity (/ɑː/ /æ/ were consistently mapped onto red). Vocalic diffuseness related to tongue raising tended to be associated with attenuated chromaticity, i.e. yellow/green/blue (/iː/ triggered strong associations with yellow and green, /uː/ with blue and brown). Achromaticity was found to be related to neutral tongue positions as /ə/ and /ɜː/ pointed to significant associations with grey (cf. Wrembel 2009).

2. EXPERIMENT

2.1. Method

2.1.1. Participants

The participants included 40 Polish students of the School of English at Adam Mickiewicz University in Poznań (33 females and 7 males, mean age M=20). Group 1 (N=20) consisted of 1st year students with intermediate English proficiency and little previous phonetic training. Group 2 (N=20) involved 3rd year students with advanced proficiency in English who had been subject to intensive phonetic training in the course of their studies. The participants did not report any synaesthetic experiences.

2.1.2. Ophthalmic examinations

The first part of the study involved a diagnosis of the participants' colour vision by means of specialised ophthalmological tests to exclude any chromatic discrimination disorders and to qualify the participants for the main part of the study. To this end, the participants underwent routine ophthalmology examinations, including visual acuity measurement, slit lamp and fundus examination. The colour vision was evaluated by using Ishihara Colour Vision Test and Farnsworth D15 panel test for each eye separately. The Farnsworth D15 panel test was performed in two different manners; as a saturated and desaturated arrangement test. The test was designed to detect congenital and acquired colour vision defects (cf. Lanthony 1978) and it is based on a set of coloured discs which have to be arranged in the correct order. The type of colour vision defect, including protan (red), deutan (green) and tritan (blue), as well its severity can be calculated based on the mistakes in the arrangement of colour and the resulting confusion vector. The exclusion criteria for further study included visual acuity less that 20/20 (1.0), any acute or chronic eye disease (i.e. cornea disease, glaucoma, retinal disease, etc.), refractive error bigger than 5D, more than 2 mistakes in the same eye in two consecutive colour vision tests.

2.1.3. Materials and procedure

The sound stimuli used in the experiment included 12 English pure vowel sounds and 6 Polish vowel sounds recorded in isolation. The stimuli were recorded by native speakers of respective languages in a professional recording studio as 16-bit mono files at a sampling frequency of 16000 Hz using the Audacity software.

The design of the experiment replicated to some extent the one applied in the previous experiments (cf. Wrembel 2007, Wrembel and Rataj 2008, Wrembel 2009), however it included a number of modifications. The present study differed from the previous ones due to the application of a modified visual stimuli presentation mode, i.e. a more fine-grained colour specification was used in order to provide a more psychologically plausible representation of the colour spectrum based on Munsell's colour space. The sound-colour mappings involved the palette of 30 colours, which were selected from over 500 samples available at http://cloford.com/resources/colours/500col.htm. Each colour was represented numerically in RGB (red, green, blue) space and in HSL values (hue, saturation, lightness).

The experiment was run on a specially designed computer program, implemented in Visual Basic, that offered the following functionality:

- it played the sound stimuli (in every testing cycle the order of the sounds was randomised);
- it displayed a selected palette of 30 colours on the screen;
- it registered in a text file the participant’s responses, i.e. the selected colour and response time.

The experiment was carried out as a series of individual sessions. The participants were seated in front of a computer screen in a dark room. They were instructed to listen to individual sounds and choose one colour from the palette of 30 colours, which appeared automatically on the screen after a 'Play' button was clicked. The participant’s choice was made by clicking on one of 30 coloured rectangles presented in 6x5 rows against a light grey background. The auditory stimuli were presented in two sessions; first the experiment was run on 12 English vowel sounds presented in 3 testing blocks, and then the procedure was repeated for the 6 Polish vowel sounds, tested in 4 blocks. Consequently, each participant listened to 60 stimuli in order to check for the consistency of sound-colour associations. In total the experiment generated 2400 tokens of cross-modal mappings that were all subject to further analysis.

2.2. Results

2.2.1. Sound-colour interactions

The results were analysed statistically by means of SPSS. The performed non-parametric chi-square test demonstrated the statistical significance of sound-colour interactions for all the Polish vowel sounds under investigation and for 9 out of 12 English vowels ($p<.01$) for both groups of participants (see Tables 1 and 2). The tables present specific colour associations including the three strongest mappings.

Table 1: Vowel-to-colour mappings for Polish.

	i	ɨ	e	a	o	u
Chi2	291	101	104	189	112	91
df	26	28	26	26	26	28
sig	0.00	0.00	0.00	0.00	0.00	0.00
	Yellow Gold Green	Grey White Navy	Blue Red	Red White Green	Yellow Chocolate Orange	Chocolate Orange Navy

Table 2: Vowel-to-colour mappings for English.

	ɩː	ɪ	e	æ	ʌ	ə	ɜː	ɑː	ɒ	ɔː	ʊ	uː
chi2	105	39	57	75	30	36	67	81	44	44	76	43
df	26	27	23	26	28	27	24	25	26	25	27	23
sig	0.00	0.06	0.00	0.00	0.34	0.10	0.00	0.00	0.01	0.01	0.00	0.01
	Yellow Gold Green		Green Blue Cyan	Red Coral Gold			Grey Navy Maroon	Red Dark red Navy	Brown Orange Gold	Maroon Brown Chocolate	Grey Chocolate Olive	Navy Chocolate Maroon

The distribution of mean values of hue, saturation and lightness assigned to the Polish and English vowels is presented in Figures 1 and 2.

2.2.2. Consistency in the sound-colour mappings

The mean internal consistency of colours assigned to the English vowels in the subsequent test blocks was 43%, with the highest values for /ɑː/ 59%, /ɩː/ 53%, /ɔː/ 51% and /æ/ 47%. The consistency values for the remaining vowels ranged between 44% and 31%. In case of the Polish vowels, the mean consistency was considerably higher as it reached 57%. The individual values were above 50% for all the vowel-colour mappings, with /a/ exhibiting the highest consistency of 71%.

As far as the external consistency is concerned, the comparison with the results of the previous studies (cf. Wrembel 2009) demonstrates that the associations with specific colours evoked by the perception of L2 speech sounds follow very similar patterns of distribution across the studies. Moreover, L2 sound-colour mappings do not differ significantly from those triggered by native Polish sounds, however, the strength of the native mappings expressed in percentage values appears to be relatively higher.

2.2.3. Group comparison

In order to investigate a potential conditioning effect of the level of English proficiency and the existence of any previous phonetic training, a group comparison was performed between Groups 1 and 2. The Pearson Chi-square test revealed no statistically significant differences in the assignment of colour to sound between the two groups for all the English vowels except /ɩ:/ (p<.01). In case of Polish, some differences between general tendencies in colour mappings were observed for both groups. However, the test for equality of means demonstrated that there were no significant differences between the groups in HSL assignment to English and Polish vowels.

Figure 1: Mean values of hue, saturation and lightness for Polish vowels.

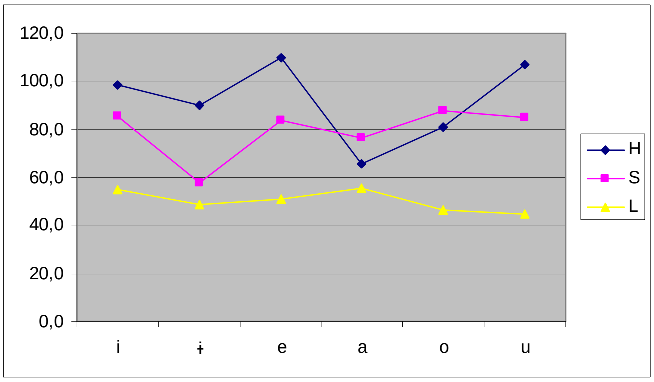
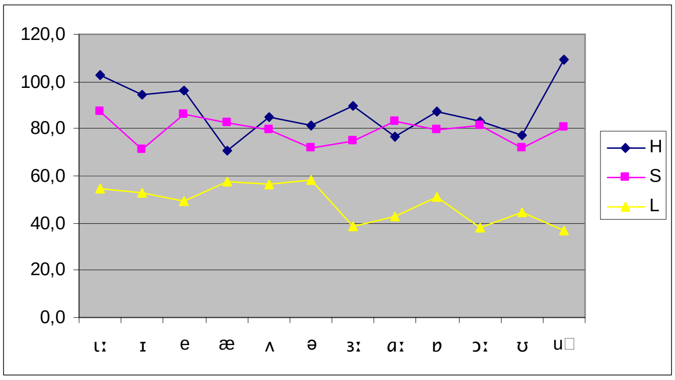


Figure 2: Mean values of hue, saturation and lightness for English vowels.



2.2.4. Formant frequencies vs. HSL

The results of the correlation analysis pointed to the existing correlations between F1 and F2 and all the HSL values, however, they were of varying strength. A strong inverse correlation was found between F1 and hue (H). Therefore, open vowels (with high F1) such as /ɑː/ or /æ/ correlated with low hue values corresponding to red hues or achromatic colours, whereas high close vowels (low F1) with high hue values (corresponding to e.g. blue, navy, magenta, orchid). There was a weak correlation between F1 and lightness and saturation. In case of the second formant, the results pointed to moderate correlations between F2 and lightness as well as F2 and hue. It indicates that the higher the F2 (i.e. the more front the vowels are) the greater the lightness of the assigned colour.

The results of the Pearson correlation analysis of the first two formants of the English vowels and RGB (red, green, blue) values demonstrated moderate correlation between F1 and Red (the higher the F1, the greater the redness) but not the other values. In case of F2, a relatively strong correlation was found between F2 and Green (the more front the vowels, the more green they were) and a moderate correlation for Blue. The analysis of the F1 to F2 ratio pointed to a moderate inverse correlation with hue.

The analysis of the Polish vowels pointed to similar trends like in English. For the Polish vowels, a moderate inverse correlation was found between F1 and hue, and a weak correlation between F1 and lightness. A strong correlation was found between F2 and lightness (i.e. the higher F2, the greater the lightness). F2 demonstrated also moderate correlations with Green and Blue. Moreover, the F1 to F2 ratio pointed to a moderate inverse correlation with hue.

2.2.5. Reaction time

The mean reaction time for sound colour mappings was 5s for English and 5.4s for Polish vowels. The ANOVA analysis of the reaction time in colour assignment to particular vowels did not point to any significant differences either between English or Polish vowels (F=1.45 p=0.1 and F=0.29 p=0.9 respectively). However, the ANOVA analysis of reaction times for particular colours yielded some significant differences between groups in case of English vowels (F=1.5, $p<0.05$). The reaction times of colour assignment was significantly lower for some colours including yellow, red, white and cyan.

2.3. Discussion

The analysis revealed that the examined vowel sounds were mapped on the visual domain of the colour spectrum in a fairly consistent and non-arbitrary manner. This was demonstrated by the internal consistency of the study, i.e. comparable patterns of vowel-colour mappings in both languages as well as the external consistency with the previous studies (Miyahara et al. 2006, Wrembel 2009, Smith et al. forthcoming). The findings reveal no effect of group on the vowel-colour interactions, i.e. the patterns of colour assignments did not depend on the participants' proficiency level (intermediate vs. advanced) nor the presence or absence of previous phonetic training in the English sound system.

Systematic relationships were found in specific parameters of speech as evoking impressions of colour vision. The analysis pointed to formant frequencies as perceptual parameters used to categorise vowel colour, as e.g. a strong correlation was found between F2 and lightness and a strong inverse correlation was observed between F1 and hue.

The results confirm most of Jakobson's (1962) and Donegan's (1985) predictions on vowel colour. Front vowels were significantly more frequently associated across languages with bright colours, as opposed to back vowels that generated darker mappings; central vowels tended to be mapped onto achromatic grey, whereas vocalic compactness was strongly associated with optimal chromaticity, i.e. red. It appeared that the sound-colour mappings are a function of the frequencies of the first two resonances, however, an unequivocal explanation for the observed patterns of sound-colour correspondence still remains to be provided.

3. REFERENCES

Day, S. 2004. Trends in synesthetically coloured graphemes and phonemes http://home.comcast.net/~sean.day/Trends2004.htm.

Donegan, P. 1985. *On the Natural Phonology of Vowels.* New York: Garland Publishing.

Grzybowski, A., R. Lewicka, T. Torlińska, B. Stelcer. 2008. "Barwy i ich znaczenie w kulturze i psychologii - rys historyczny i stan obecny", [Colours and their meaning in culture and psychology – a historical outline and contemporary status of colour vision theories]. *Klinika Oczna (Acta Ophthalmologica Polonica)*, 110 (1-3), 116-124.

Jakobson, R. 1962. *Selected Writings: I Phonological Studies* .The Hague: Mouton.

Lanthony A. The desaturated panel D-15. Documenta ophthalmologica 1978; 46: 185-189.

Marks, L. 1975. "On coloured-hearing synaesthesia: cross-modal translations of sensory dimensions", *Psychological Bulletin* 82 (3): 303-331.

Miyahara, T., T. Amemiya and R. Sekiguchi. 2006. "A psychological experiment on non-synesthetes correspondence between colours and voiced vowels". Proceedings of the First International Workshop in Kansei, Japan, 102-105.

Shimojo, S, L. Shams. 2001. "Sensory modalities are not separate modalities: plasticity and interactions". *Current Opinion in Neurobiology*, 11: 505-509.

Simner, J., J. Ward, A. Jansari, K. Noonan, D. Oakley. 2005. "Non-random associations of graphemes to colours in synaesthetic and non-synaesthetic populations". *Cognitive Neuropsychology* 22, 1-17.

Smith, R, Moos, A., Cartwright-Hignett, W., and Simmons, D. (forthcoming). Towards a phonetically-rich account of speech sound > colour synaesthesia, in: Biggam, C., C. Hough, D. Simmons and C. Kay (eds.) *New Directions in Colour Studies.*

Ward, J., B. Huckstep, E. Tsakanikos. 2006. "Sound-colour synaesthesia: to what extent does it use cross-modal mechanisms common to us all?" *Cortex*, 42, 264-280.

Wrembel, M. 2007. "Still sounds like a rainbow - a proposal for a coloured vowel chart". Proceedings of the Phonetics Teaching and Learning Conference PTLC2007 (CD edition). London: UCL, 1-4.

Wrembel, M. 2009. "*On hearing colours - cross-modal associations in vowel perception in a non-synaesthetic population". Poznań Studies in Contemporary Linguistics,* 45(4), 595-612.

Wrembel, M. & Rataj, K. 2008. "Sounds like a rainbow" - sound-colour mappings in vowel perception", in: Botinis, A. (ed.). Proceedings of the 2nd ISCA Workshop on Experimental Linguistics, ExLing 2008, Athens: University of Athens, 237-240.

Phonological Cross-linguistic Influence in Third or Additional Language Acquisition

Eva-Maria Wunder

University of Augsburg, Germany
eva-maria.wunder@phil.uni-augsburg.de

ABSTRACT

Cross-linguistic influence (CLI) in Third or Additional Language Acquisition from a non-native language is a well-documented phenomenon especially in the area of lexis, but also in morphosyntax. With regard to non-native CLI on the acquisition of an L3/Ln phonological system, however, only very few studies have been conducted so far (Hammarberg and Williams 1993; Pyun 2005; Tremblay 2007; Gut 2009; Llama et al. 2009; Wrembel 2009). This paper tries to help fill this gap by investigating the acquisition of aspiration patterns of voiceless stops. Eight L3/Ln learners of Spanish with L1 German and L2 English were recorded performing a Spanish and English read-on-your-own task. The recordings were analysed for degree of aspiration of the voiceless stops in stressed onset position. Subsequently, the L3/Ln VOT values were attributed either to L1 (low VOTs) or L2 influence (high VOTs). The mixed results point to the existence of non-native CLI on L3/Ln phonology, but also to an underlying L1 effect, as well as influence in the form of combined CLI from L1 and L2.

Keywords: phonological cross-linguistic influence, VOT, multilingualism.

1. INTRODUCTION

It was the linguist Scovel who identified the "Joseph Conrad Phenomenon"**,** named after the Polish-born British author Joseph Conrad, who acquired English only when he emigrated to England in 1878 and managed to write some famous novels with perfectly formed English syntax. His English pronunciation, however, retained a Polish accent throughout his life. For Scovel, this was proof of his hypothesis and reason enough "to offer a free dinner to anyone who can show him an individual who learned a (sic!) L2 after puberty and who now speaks that L2 with perfect native pronunciation" (Tarone 1987: 80). Apparently, the offer of April 1977 still stands.

A foreign accent in a non-native language is usually believed to be caused by influence coming mostly from the learner's mother tongue (e.g. Ringbom 1987, Pyun 2005). However, the Hammarberg-Williams study (1993) related another probable cause for foreign accent when learning a new non-native language, namely influence from other non-native languages in the learner's mind. A number of studies have already been conducted along the lines of this cross-linguistic influence in multilingual acquisition (henceforth *L3/Ln acquisition*), the majority of them undertaken in the field of lexis (e.g. Cenoz 2001; De Angelis and Selinker 2001; Müller-Lancé 2006). Moreover, an increasing number of studies on L3/Ln cross-linguistic influence in morphology and syntax are carried out (e.g. Klein 1995; Bardel and Falk 2007). Likewise, there also seems to exist cross-linguistic influence between a learner's non-native languages in the area of L3/Ln-phonology, for which only a small number of studies so far has tried to find evidence (e.g. Gut 2009; Wrembel 2009). To further explore the validity of this fascinating hypothesis, this paper will investigate whether there is phonological cross-linguistic influence between an adult multilingual speaker's non-native interlanguages. Firstly, an introduction to the concept of cross-linguistic influence and its conditioning factors will be given, followed by the presentation of the data and employed method for the present paper, as well as the discussion of the findings.

2. CROSS-LINGUISTIC INFLUENCE: TYPES AND TRIGGERING FACTORS

Compared to second language learners (henceforth *L2 learners*), learners of a third or additional language (henceforth *L3/Ln learners*) have knowledge from at least two languages stored in their minds, have gained considerable metalinguistic awareness, and are better equipped with learning strategies (e.g. Fouser 2001; Ó Laoire 2005). Multilingual learners subconsciously, or even consciously if more experienced, draw on this prior knowledge in all ensuing language learning, resulting in the key concept of this paper, i.e. *cross-linguistic influence* (henceforth *CLI*), a term coined by Sharwood Smith (1983) and Kellerman (1984), which can affect various linguistic levels. Traditionally, CLI occurs as a one-to-one type between a source and a target language (henceforth *SL* and *TL*) (cf. De Angelis 2007: 20f). This is opposed to what has been termed *combined CLI* by De Angelis, "the simultaneous influence of more than one language upon a target language, i.e. a many-to-one type" (2007: 21). However, in most cases it will be rather difficult to attribute CLI to only one specific SL amongst several that may be interacting, the common type of CLI between multiple languages in an L3/Ln learner's mind.

When acquiring a new language, depending on the linguistic level, different factors might be significant and contribute to eventually trigger CLI. For instance, the variables age of learning (e.g. Cenoz 2001), exposure to the non-native language (e.g. Fouser 2001; Vildomec 1963), order of acquisition (e.g. De Angelis 2007), language distance (e.g. De Angelis and Selinker 2001), formal similarity (e.g. Selinker and Baumgartner-Cohen 1995) or perceived language distance (e.g. Odlin 1989; Singleton 1987) have been shown to condition lexical CLI. Whether they also have an impact on L3/Ln-phonology remains to be investigated.

In studies on phonological CLI, though, to date only *proficiency*, *recency of use*, *foreign language effect* and *task relatedness* have been assessed as significant. With regard to proficiency, it is believed the more proficient learners are in a non-native language other than the one they are acquiring at the moment, the more likely CLI will occur from this SL in general (e.g. Odlin and Jarvis 2004; Ringbom 1987). With respect to proficiency in the TL, however, Hammarberg and Williams (1993, 1998) discovered that phonological CLI is most likely to occur in the initial stages of acquisition, as the learner's command of the TL is still only very rudimentary and many knowledge gaps have to be filled with previously acquired linguistic information. Cross-linguistic influence is also more likely to occur from languages the learners have made use of recently and therefore are still fresh in their minds and consequently accessed more easily, as a number of studies confirm (e.g. Hammarberg and Williams 1998; Vildomec 1963). Hammarberg and Hammarberg (1993, 2005) as well as Hammarberg and Williams (1993) delivered corroborating evidence from their subject showing CLI from her L2, which was still very vivid to her due to a recent stay abroad. Meisel's (1983) concept of foreign language effect (also *L2 status*, Cenoz 2001, or *foreign language mode*, Selinker and Baumgartner-Cohen 1995), referring to the fact that a certain language in question is categorised as a non-native language, also facilitates phonological CLI. Apparently, the similar acquisition processes and type of association established between two or more non-native languages in a learner's mind enables the learner to activate a prior non-native language more easily than the L1 when acquiring a new language (e.g. Gut 2009; Llama et al. 2009; Tremblay 2007; Wrembel 2009). Finally, Hammarberg and Williams (1993) showed in their study that phonological CLI could possibly be related to the type of task the L3/Ln learner has to perform: in a read-after-me task of an L3 text, the subject's pronunciation was coloured by the L1, whereas when faced with the more complex task of reading without a native speaker model, the L3 learner tended to rely on the L2 as a coping strategy.

3. DATA AND METHOD

To be able to tell whether CLI from the L2 of the subjects has occurred or not, the feature of aspiration in the stressed syllable-initial voiceless stops /p t k/ was chosen for the empirical study of the present paper, operationalised by measuring along the continuum of *voice onset time* (henceforth *VOT*). VOT was chosen because the presence or absence of the correct amount of aspiration is one of the reasons for speech to be perceived as accented (cf. Nathan et al. 1987: 204). Moreover, VOT is a feature which allows the establishment of a straightforward relationship between the languages investigated, namely German as L1, British English as L2, and Castilian Spanish as TL. In all three phonological systems exist the voiceless stops /p t k/. However, they are produced differently in each language. Aspiration in German and English is an allophonic feature. For instance, in English, the voiceless stops in stressed syllable-initial position followed by a syllable-nuclear voiced segment are aspirated, such as in *part* [ˈpʰaːt], *tone* [ˈtʰəʊn] or *cold* [ˈkʰɔːld] (cf. Laver 1995: 351), except for when the stops are preceded by /s/ or appear in word-internal stressed onset position (cf. Llama et al. 2009). Moreover, voiceless stops in stressed onset positions followed by a voiced syllable-marginal resonant, i.e. /r/, /l/, /j/ or /w/, are also aspirated (cf. Laver 1995: 351), for example in *cry* [ˈkʰɹaɪ], *plane* [ˈpʰleɪn], *cure* [ˈkʰjʊə] or *tweed* [ˈtʰwiːd].

For the participants' L1 German, following previous research (Jessen 1998; Mansell 1979) and the analysis of the speech of three L1 German speakers recorded as a control group, average values of between 30ms and 45ms for /p/, 40ms and 55ms for /t/, and between 45ms and 60ms for /k/ were assumed for this study. For British English, again in accordance with the literature (e.g. Lisker and Abramson 1964; Llama et al. 2009) and a control recording, average VOT values for English were set between 60ms and 100ms, with VOT for /p/ ranging between 60ms and 70ms, for /t/ between 70ms and 80ms, and for /k/ between 80ms and 100ms. In Spanish, however, there is no aspiration. For the present study, after consultation of previously established values (Llama et al. 2009; Rosner et al. 2000) and a control recording, it was decided on Spanish VOT values of between 0ms and 15ms for /p/, between 15ms and 20ms for /t/, and for /k/ between 20ms and 30ms. To find much aspiration in the subjects' Spanish would translate to CLI from their L2 English.

Eight subjects consented to be recorded and fill in a questionnaire, five of them female and three male. For comparative reasons, all of them spoke German as L1 and English as L2, except for participant 4, who lived in Romania as a child and learnt Romanian as L2, followed by English as L3 at the age of 12. Moreover, all participants were learners of the TL Spanish, perceiving their proficiency as that of beginners to advanced beginners, except for participants 2 and 5 who considered themselves advanced learners of Spanish. Apart from subjects 3 and 8, all participants had knowledge of at least one further non-native language besides English and Spanish. At the time of recording, five subjects were taking the Spanish beginners' course *Lengua española 2* taught by native speakers of Spanish. Participant 2 had attended it previously, whereas subjects 1 and 3 so far had only completed *Lengua española 1*, the four-month introductory course, which is similarly structured to *Lengua española 2*. The first intensive course of six hours per week, equalling around 90 hours per term, suffices to achieve a beginners' level of Spanish.

Each participant was recorded performing a read-on-your-own task of a text of about one minute's length firstly in their L2 English (or L3, in the case of participant 4 - PART4) and then TL Spanish. As mentioned above, this task is supposed to elicit non-L1 influence, as the learners have to rely more on their prior linguistic knowledge than only trying to imitate in a read-after-me task with native speaker model, which in turn is assumed to elicit L1 CLI. Consequently, they were expected to show influence from their strongest non-native language, which from self-evaluation, except for PART4 and PART5, was English. The recordings were made using a 24-bit Edirol audio recorder. Mono wave format was chosen at a sampling rate of 44kHz for reasons of accurate time-segmentation alignment.

To avoid intuitive, sometimes impressionistic judgment of native speaker listeners, the recordings were analysed for VOT, i.e. the interval in milliseconds between the release of the articulators and the beginning of regular vocal chord pulses (cf. Nathan et al. 1987: 205), with Praat (version 4.6.32). Firstly, in each recording the utterances containing /p t k/ in stressed syllable-initial, potentially aspirated position were transcribed orthographically on one tier. Afterwards, the voiceless stops were isolated on a second tier by setting a boundary in the waveform at the point of the release of the articulators, indicated by the highest amplitude in the waveform. The values were also included for weak releases with small amplitudes if a clear burst was still visible. Moreover, where double or multiple bursts occurred, VOT was measured with the first burst as starting point of aspiration. A second boundary was set exactly at the point of zero-crossing on the vertical axis of the first regular wave of the following sound. The interval between these two boundaries constitutes VOT or aspiration. Ambiguous tokens and individual mispronunciations resulting in stress shifts, affrication or doubtful beginnings of periodic waves were not included.

4. DISCUSSION OF RESULTS

Measurements conducted to evoke CLI on the production of voiceless stops /p t k/ in stressed onset position in the subjects' L2 English and L3/Ln Spanish yielded the mean VOT values presented in Table 1. The VOTs measured in the read-on-your-own task were compared with the previously established values for native-like aspiration and assigned to a certain L1 range, or in between, strictly according to the cut-off values (cf. section 3). All in all, the analysis yields rather heterogeneous results regarding CLI.

Table 1: Mean VOT values (in ms) for each participant's L2 English and L3/Ln Spanish voiceless stops in stressed onset position (within L1 German range / L1 English range / L1 Spanish range / hybrid values) and potential SLs for CLI on VOTs of non-native-like productions of /p t k/.

	VOT **/p/** (in ms)		VOT **/t/** (in ms)		VOT **/k/** (in ms)	
	L2 E	L3/Ln S	L2 E	L3/Ln S	L2 E	L3/Ln S
PART1	62	48	75	37	90	52
	native-like	p_S (→ G/E)	native-like	t_S (→ S/G)	native-like	k_S (→ G)
PART2	78	26	98	24	112	42
	native-like	p_S (→ S/G)	native-like	t_S (→ S/G)	native-like	k_S (→ S/G)
PART3	47	34	73	46	74	59
	p_E (→ G/E)	p_S (→ G)	native-like	t_S (→ G)	k_E (→ G/E)	k_S (→ G)
PART4	60	17	72	19	75	35
	native-like	p_S (→ S/G)	native-like	native-like	k_E (→ G/E)	k_S (→ S/G)
PART5	38	20	51	23	71	34
	p_E (→ G)	p_S (→ S/G)	t_E (→ G)	t_S (→ S/G)	k_E (→ G/E)	k_S (→ S/G)
PART6	62	54	73	60	85	66
	native-like	p_S (→ G/E)	native-like	t_S (→ G/E)	native-like	k_S (→ G/E)
PART7	50	33	73	46	83	59
	p_E (→ G/E)	p_S (→ G)	native-like	t_S (→ G)	native-like	k_S (→ G)
PART8	30	12	64	29	86	35
	p_E (→ G)	native-like	t_E (→ G/E)	t_S (→ S/G)	native-like	k_S (→ S/G)

Of the four conceivable constellations for analysing stop productions, native-like English but non-native-like Spanish productions are of the main interest regarding CLI from non-native languages investigated in this paper. According to the discussed variables, CLI from English should only occur from native-like L2 productions onto the L3/Ln, which subsequently displays non-native-like VOTs, either distinctly influenced by one specific language or situated in between two languages, showing hybrid values. Depending on whether mean VOTs in Spanish were located within the range of the L1 cut-off values established in section 3 or in between mean VOTs of L1 Spanish, German or English, it translated either to distinct influence from one or two languages, as displayed in Table 1. Generally speaking, intermediate mean L3/Ln VOTs within L1 German range would be interpreted as German influence, whereas relatively high VOTs located in the L1 English interval would be attributed to L2 influence. As can be inferred from Table 1, the analysis yields rather mixed results. For instance, in PART1's speech, L1 influence on the production of L3/Ln Spanish /k/ was measured, but values obtained for Spanish /p/ could not be assigned clearly to either German or English, and L3/Ln /t/ ranged in between Spanish and German.

A few learners, like PART4 or PART8, seem to have been able to establish phonetic categories in the sense of Flege's (1995) *Speech Learning Model* for L2 English and L3/Ln Spanish productions of /p t k/ similar to those of L1 speakers. PART4, for example, shows native-like VOT values for /t/ in both her L2 and L3/Ln as well as in L2 tokens of /p/. This might either indicate advanced proficiency in her L3/Ln Spanish, or she simply acquired native-like VOT. It could also be connected to a higher metalinguistic awareness due to the fact that PART4 had already attained native speaker proficiency in her L2 Romanian at a very early age, and consequently might approach acquisition of further non-native phonological systems more efficiently than less experienced language learners (e.g. Ó Laoire 2005; Thomas 1988).

Expressed in percentages, all subjects produce 62.5% of L2 English stops with near-native VOT values, more specifically 50% of /p/, 75% of /t/ and 62.5% of /k/, illustrating the required advanced proficiency for the L2 to potentially become SL for phonological CLI. In comparison, only 8.3% of all tokens of L3/Ln Spanish stops are produced with the correct amount of aspiration in the TL, i.e. 12.5% of /p/ and /t/ and 0% of /k/. Bearing in mind that we are dealing with beginners of Spanish, the percentage producing already near-native VOTs points to the conclusion that either the proficiency level of these participants is higher than that of beginners and their L3/Ln phonological system already beyond the stage for CLI to occur from non-native languages, or we are dealing with individual cases of native-like acquisition of VOT.

Furthermore, there seems to exist a quite evident L1 effect on both English and Spanish productions. For instance, PART5 shows mean VOTs situated between German and Spanish for her L3/Ln stops. Interesting to note is that her L2 tokens already exhibit German influence, which seems to have been transferred further onto the L3/Ln. This underlying L1 effect is also visible in the L3/Ln Spanish VOT means of /p/ and /k/ of PART3, or PART7's /p/. The reason for that might be to a certain degree inherent to the feature of VOT itself. Previous research in the field of Second Language Acquisition established that to acquire VOTs comparable to native-speaker values in the L2 is achieved only rarely (e.g. Díaz-Campos 2004; Fellbaum 1996). Instead, the L2 learners tend rather to retain some L1 features in their L2, thus creating *hybrid* or *compromise VOTs* (e.g. Flege and Eefting 1987; Laufer 1996). This can be seen, for example, in the mean values of PART5, where L1 influence leads to the creation of intermediate VOTs in her productions of English /k/, which exhibit means between L1 German and L1 English cut-off values.

Applied to L3/Ln acquisition, firstly, it is conceivable in the case of PART5 that as her L2 mean VOT for /k/ is already influenced by the L1 German, thus creating a hybrid L2 value, it could have been transferred further into the L3/Ln TL Spanish. The third language would have been added to the feature concoction and would have created another compromise value between Spanish and German. This would signify that to some degree CLI from the L2 has occurred, with the compromise L2 value instigating the creation of a hybrid TL form. In reality, looking at PART5's mean VOTs for /k/ across both L2 and L3/Ln, they show a difference of 37ms. However, in the case of PART3's values for /p/, they only differ by 13ms, with the similar mean values pointing to transferred hybrid VOT. Secondly, some of these compromise VOTs could also indicate combined influence from the L1 and L2 on L3/Ln productions (cf. De Angelis 2007). PART1 and PART6 are the only subjects to display values that cannot be clearly attributed to one SL. Whereas PART1 displays a VOT mean of 48ms for her L3/Ln /p/, VOTs for all of PART6's L3/Ln stops are located in between cut-off values for L1 German and L1 English, with 54ms for /p/, 60ms for /t/, and 66ms for /k/ respectively, both participants thus indicating aforementioned combined CLI.

With regard to potential source languages for CLI on the TL phonological system, similar to Llama et al.'s (2009) findings, my subjects also seem to exhibit an L1 effect on the TL, with 29.5% of the non-native VOT values for Spanish /p t k/ situated within L1 German range. None clearly demonstrates CLI only from the L2, contradicting findings of previous research (e.g. Hammarberg and Williams 1993; Llama et al. 2009). Influence of the L2 on the TL Spanish only occurs together with L1 as aforementioned combined CLI in 16.6% of the cases. Moreover, hybrid values constitute the largest group (45.8%), i.e. instances in which it cannot be determined whether VOTs were influenced by L1 German or native-like Spanish.

Concerning any influence coming from the L2 English, as already mentioned it was only visible in combined CLI. However, it is also conceivable that there might have been L2 influence on the L3/Ln productions located within German L1 values, as will be argued in the following. Firstly, looking at several L2 tokens (PART3's /p k/, PART4's /k/, PART5's /k/, PART7's /p/, PART8's /t/), they are all produced with compromise VOT situated in between German and English L1 means. Analogous to the aforementioned L1 influence transferred from these L2 hybrid values onto the L3/Ln, resulting in mean VOTs located in between L1 Spanish and German, it is probable that existing influence from the L2 features of the compromise values was also transferred. This transfer of L2 features would result in higher mean L3/Ln values compared to those influenced only by the L1. So, it could be hypothesised that these L3/Ln VOTs situated within L1 German range and consequently showing higher VOTs than Spanish-German intermediate ones, might either be due solely to L1 influence or maybe to hybrid L2 influence, which could have had some effect in that it has raised the Spanish aspiration rate, though only up to L1 German level.

Looking at the means, PART3 seems to be the only subject to possibly have also incorporated L2 influence in his L3/Ln productions of /k/, which display VOT means of 74ms in the L2 and 59ms in the L3/Ln. All others with L2 hybrid values either show Spanish-German L3/Ln compromise values, or already present L2 hybrid VOTs situated closer to L1 German than English. Yet, this is only a hypothesis and with the present data or existing methodologies it is impossible to determine exactly where the influence came from, or whether it occurred at all. Perhaps the hybrid values displayed by the participants simply reflect chance results or idiosyncratic pronunciations.

5. CONCLUSION

Taking everything into consideration, it has to be acknowledged that this paper contradicts rather than corroborates previous findings of non-native languages exerting phonological CLI on an L3/Ln TL (e.g. Hammarberg and Williams 1993; Llama et al. 2009; Pyun 2005; Tremblay 2007). It was not possible to produce clear evidence for L2 English influence on the acquisition of L3/Ln Spanish aspiration patterns. This may be due to certain limitations of this study.

Firstly, due to the fact that only eight learners consented to be recorded, the number of analysed tokens was quite limited. Moreover, despite efforts to record a homogeneous population, the participants of this study possibly deviated from this ideal. For instance, some subjects might have already been beyond the stage for phonological CLI to occur from a non-native language; they might have been too proficient in the TL, as they exhibited considerable L1 influence, which is supposedly characteristic for advanced stages of the acquisition of a new phonological system (cf. Hammarberg and Williams 1993).

Although the correct perception of sounds, which according to Flege (1995) seems to be the basic premise for preventing foreign accent, unfortunately cannot be taught (yet), it is feasible to raise metalinguistic awareness. Besides, incorporating prior linguistic knowledge of learners and trying to channel it in the classroom by showing up what can and what cannot be transferred could also help facilitate the acquisition of a new phonological system. There is still a lot of work to be done. However, if effective learning strategies and teaching methods can be developed and successfully implemented, Mr. Scovel will, in due course, have to host a dinner banquet.

6. REFERENCES

De Angelis, G. 2007. *Third or Additional Language Acquisition*. Clevedon: Multilingual Matters.

De Angelis, G., Selinker, L. 2001. Interlanguage transfer and competing linguistic systems in the multilingual mind. In: Hufeisen, B., Fouser, R. (eds.), *Introductory readings in L3*. Tübingen: Stauffenburg. 42-58.

Flege, J. 1995. Second language speech learning: Theory, findings, and problems. In: Strange, W. (ed.), *Speech perception and linguistic experience: Issues in cross-linguistic research*. Baltimore: York Press. 233-277.

Gut, U. 2009. Cross-linguistic influence in L3 phonological acquisition. *International Journal of Multilingualism* 1479-0718.

Hammarberg, B., Hammarberg, B. 1993. Articulatory re-setting in the acquisition of new languages. *Phonum* 2, 61-67.

Hammarberg, B., Williams, S. 1993. A study of Third Language Acquisition. In: Hammarberg, B. (ed.), *Problem, process, product in language learning*. Stockholm: Stockholm University Press. 60-69.

Kellerman, E. 1984. The empirical evidence for the influence of the L1 in interlanguage. In: Davies, A., Criper, C., Howatt, A. (eds.), *Interlanguage*. Edinburgh: Edinburgh University Press. 98-122.

Llama, R., Cardoso, W., Collins, L. 2009. The roles of typology and L2 status in the acquisition of L3 phonology: The influence of previously learnt languages on L3 speech production. *International Journal of Multilingualism* 1479-0718.

Nathan, G., Anderson, W., Budsayamongkon, B. 1987. On the acquisition of aspiration. In: Ioup, G., Weinberger, S. (eds.), *Interlanguage phonology: The acquisition of a second language sound system.* Cambridge, MA: Newbury House. 204-228.

Odlin, T. 1989. Language transfer. Cross-linguistic influence in language learning. Cambridge: Cambridge University Press.

Selinker, L., Baumgartner-Cohen, B. 1995. Multiple language acquisition: "Damn it, why can't I keep these two languages apart?" *Language, Culture and Curriculum* 8(2), 115-121.

Sharwood Smith, M. 1983. On first language loss in the second language acquirer: Problems of transfer. In: Gass, S., Selinker, L. (eds.), *Language transfer in language learning*. Rowley, MA: Newbury House. 222-231.

Tremblay, M.-C. 2007. L2 influence on L3 pronunciation: Native-like VOT in the L3 Japanese of English-French bilinguals. Paper presented at the L3 phonology workshop at Freiburg, Germany, August 2007.

Wrembel, M. 2009. L2-accented speech in L3 production. *International Journal of Multilingualism* 1479-0718.

For full references please contact me via the abovementioned email-address.

Development in L2 phonology: Let's take a fresh look at the age issue

Martha Young-Scholten

Newcastle University
Martha.young-scholten@newcastle.ac.uk

ABSTRACT

We know that both the learner's L1 phonology and various universals affect the acquisition of an L2 phonology, and that younger learners ultimately surpass older L2 learners. However, the question of whether we are dealing with fundamental differences between older and younger learners remains unresolved. When the context of acquisition is roughly comparable (learners receive equal amounts of exposure to native speaker input from the start), do we find similar stages of development regardless of age of initial exposure? I discuss how longitudinal studies of morphosyntactic development have moved SLA theory forward, and while this is also true for L1 phonology, little known about similarities or differences between child and adult L2 phonological development. I argue that while Optimality Theoretic work has led to important insights, we still require longitudinal data to fully investigate development. Such data can also shed light on currently debated issues in L2 morphosyntax.

Keywords: Longitudinal, age, naturalistic, prosodic structure, morphosyntax

1. INTRODUCTION

In their recent review of longitudinal studies published between 2002 and 2004 in major journals, Ortega and Iberri-Shea (2005:26) present the case that 'many, if not all, fundamental problems about L2 learning that SLA researchers investigate are in part problems about 'time', and that any claims about 'learning' [...] can be most meaningfully interpreted only from a longitudinal perspective. They further point out that longitudinal studies can explore relationships among variables, uncovering causes and effects with respect to various phenomena.

Researchers investigating the acquisition of morphosyntax have since the 1980s given equal attention to the kind of knowledge acquired and the way in which it is acquired (e.g. Clahsen and Muysken 1986; Meisel 2003). If the same innate mechanisms are employed across the lifespan, there should be neither fundamental differences in the processes of acquisition nor in the resulting mental representations. The matter is far from settled, and this dual emphasis has led to the continued collection of longitudinal data whose management and sharing has undoubtedly been facilitated by CHILDES data management and sharing. We now have longitudinal data from various learner types: from typically developing and atypically developing children acquiring a single L1, two L1s simultaneously, two L1s successively but prior to schooling, from children acquiring an L2 after the L1 has been acquired, between age five and puberty, and in a range of contexts which now include heritage language communities– as well as from adults acquiring an L2 in both instructional and naturalistic contexts. An existing gap with respect to child L2 acquisition is slowly being filled – typically by PhD students (and their supervisors); see e.g. Haberzettl (2005), Haznedar and Schwartz (1997), Tracy and Thoma (2007). Direct comparisons such as Snow and Hoefnagel-Höhle's (1982) and Unsworth's (2005) of L1 and L2 children and L2 adults remain rare.

Despite a tradition of longitudinal study in L1 phonology (see below), this trend has not taken hold in L2 phonology; Gut's (2009) survey of studies on L2 phonology over the past 39 years unearthed only 17 such studies. Not only is supra-segmental phonology underrepresented and the relationship between linguistic phenomena and non-linguistic phenomena little explored, but researchers' procedures often yield unnatural speech. Gut further notes lack of work empirically testing models (e.g. Flege's Speech Learning; Eckman's Markedness/Structural Conformity; Major's Ontogeny Phylogeny; Archibald's parameter resetting, Dziubalska-Kołaczyk's Natural Phonology; Escudero and Boersma's Optimality Theory –based Functional Phonology).

In what follows, I show that longitudinal data from naturalistic adult L2 learners has been instrumental in the evolution of theorizing on the acquisition of morphosyntax, and I argue that L2 phonology should tred the same path. One unexplained phenomenon – that of variable production of inflectional morphology – would benefit considerably from multi-factor longitudinal studies of naturalistic adults. More importantly for us is the possibility of addressing the age factor from a developmental perspective. Since Ioup and Weinberger (1987) it has become apparent that adult L2 phonology goes beyond L1 transfer; Optimality Theory (Prince and Smolenksy 1993) now allows us to state that a set of universal constraints is available across the lifespan. L1 acquisition means converging on the relevant adult constraint ranking; L2 acquisition then involves constraint reranking. Under OT, interlanguage phonologies are possible phonologies even if a given ranking is unattested in any language. Yet this view does sit easily with the evidence that older L2 learners rarely converge on the target phonology. In 1990, Long pegged the start of the closure of a critical period for phonology at age six, a position to which Kramsch (2003) attributes declining interest in L2 phonology, at least in classroom-relevant studies. Because current models of L2 phonology no longer assume age differences involve knowledge type, we must instead look at how L2 knowledge is acquired to identify what distinguishes older and younger L2 learners. This is best done by looking at longitudinal data. As Hancin-Bhatt notes in her review of OT L2 research, the learner's developing phonology can 'finally be considered as central to theory construction' (2008:141). She goes on to stress that without longitudinal data 'claims about acquisition within an OT model are weak.' (2008:142). Longitudinal studies go beyond accuracy; as Eckman (2008:101) notes 'it has been recognized since the early days of Error Analysis that learner errors are not the only measure of difficulty [...] rate of acquisition [...] is a more insightful measure of learning'

2. THE MODERN HISTORY OF THE STUDY OF L2 MORPHOSYNTAX

2.1. What we know about L2 morphosyntax

Debates regarding the L2 acquisition of morphosyntax have largely taken place on a longitudinal battlefield. The most important studies have been of naturalistic adult learners living in the target language country. Certainly if one wishes to compare L1 and L2 acquisition, if it is possible to remove a variable that distinguishes the two – and thus restrict adult learners' focus on forms – this is highly desirable (see e.g. Wode who studied his children when few such studies of even children, observed 'without the benefit or obstruction from schoolroom instruction.' (1978:102). Table 1 gives the main features of the most influential studies of uninstructed adult learners, who happened to be immigrants; this is not an exhaustive list of early such studies; see e.g. Huebner (1983). Among these the ZISA study of Italian, Portuguese and Spanish migrant workers acquiring L2 German has provoked the most heated debate. Since the 1980s, this has shifted from UG access to L1 transfer and functional categories at the initial state in the 1990s to the status of inflectional morphology in the 2000s, and has insured the value of longitudinal studies. While some (Vainikka and Young-Scholten 1994) propose that adults, regardless of their L1, systematically construct similar grammars, this and related issue continue to be debated.

Table 1. Longitudinal studies of L2 morphosyntax by immersed /naturalistic adult learners

Study	L1 and L2	Subjects	Duration
Cazden et al. 1970s	L1 Spanish/L2 English	2 children, 2 teenagers, 2 adults	10 months
ZISA 1970s/80s	L1 Spanish, Portuguese, Italian/ L2 German	12 adults	2 years
ESF 1980s	6 L1s/5 European L2s	40 adults	2 ½ years
VYSA 1990s	L1 English/L2 German	Three teenagers	1 year

2.2. What we don’t know about L2 morphosyntax

One current focus is learners’ variable production of inflectional morphology and the question of whether inflectional morphology and syntax are tightly coupled in both L1 child and child and adult L2 acquisition (see Prévost and White 2000; Schwartz and Sprouse 1997). To account for L2 learners’ production over time, i.e. on the basis of longitudinal studies, Haznedar and Schwartz (1997) and Lardiere (1998) propose the Missing Surface Inflection Hypothesis under which L2 learners acquires the features – for example agreement and tense – which morphology spells out, yet have mapping problems. What phonology contributes to morphosyntactic development in explaining how the learner moves from reception of a continous speech stream to construction of a mental grammar is, according to Carroll (2001:1) ‘one of the most under-researched and under-theorized aspects of second language acquisition’. Where the assignment of phonological representations to acoustic input is preliminary step in morphosyntactic processing, it is unsurprising that studies are starting to show that learners’ L1 phonology – particularly prosodic phonology - (see e.g. Goad, White and Steele 2003; Oldenkamp 2010) is implicated. We now turn to the role longitudinal studies have played thus far in research on phonological acquisition.

3. DEVELOPMENT IN PHONOLOGY

3.1. Studies in first language acquisition

Longitudinal studies of children date back to the 19th century, when the only option was to conduct diary studies whose documentation took the form of parental notes. Problems with such studies notwithstanding, Ingram (1989:10) stresses their enormous value in ‘providing a database for the field’ and their provision of ‘a comprehensiveness that is impossible to replicate’. Table 2 represents the progression presented in Ingram from diary through large sample through early longitudinal sampling, where studies examine a range of aspects of development, including phonology. With behaviourism came quantitative studies aiming to determine group norms (of vocabulary growth, pronunciation accuracy and sentence length) and a deemphasis on language as a system of rules. With the advent of Chomskyan linguistics and modern recording techniques, longitudinal studies – not necessarily of the investigator’s children, and of at least three children to ensure typicality, became the standard; depending on the phenomena under investigation, these run for between 5 months and 2 years with collection at weekly, fortnightly or monthly intervals.

Table 2. A time-line of studies on children’s acquisition

Diary Studies	Large sample studies	Longitudinal sampling
Preyer 1889	Late 1920s-1950s: groups of between 72 (Fisher 1934) and 430 (Templin 1957) children from 1;6 to 8;0 on their articulation, vocabulary, sentence length	Braine 1963: 3 children
Stern 1907,1924		Miller and Ervin 1964: 5 children
Leopold 1939-1949		Brown 1973: 3 children
Grégoire 1937,1947		Bloom 1970 3 children
Gvozdev (1949)		Others (usually one child) e.g. Velten 1943, Waterson 1971, Menn 1973, Smith 1973, Ingram 1974
Zarębina 1965		Shvachkin 1973 study of 19 Russian children’s perception of phonemic distinctions from 0;10 to 1;6

Studies focusing just on children’s phonology have ranged from Smith’s (1973) comprehensive study of a single child’s development of English to larger but more narrowly focused (on prosodic structure) studies of e.g. Dutch, German and French Canadian children’s prosodic structure (see Fikkert 1994; Lleó and Prinz 1997; Rose 2000. Interestingly, in highly similar Dutch and German, language—specific developmental differences emerge with respect to children’s rhyme elaboration.

3.2. Studies in L2 phonology

Pater's (1997) earlier attempts to tell a developmental story from cross-sectional data from French speakers' acquisition of English stress were foiled: although he divided learners into groups based on a general proficiency test, this failed to reveal patterns in the parameter resetting he examined. Yet simply because such a procedure fails to show any patterns does not mean development is asystematic. Longitudinal data indeed reveal systematicity. However, patterns are not necessarily linear, where studies, e.g. Carlisle's (1998) and Abrahamsson's (2003) of syllable structure, may point instead to U-shaped development.

Table 3. A selection of longitudinal studies of L2 phonology

Study	Description	Features investigated
Akita 1998	Three Japanese students' English over 1 year of immersion	stress, rhythm and syllable structure
Abrahamsson 2003	One Chinese adult learner of Swedish 20 months	syllable structure
Carlisle 1998	Four Spanish students learning English 4 years	syllable structure
Derwing, Munro and Thompson 2008	32 Mandarin, Ukrainian and Russian adults' English over 2 years	fluency and comprehensibility
Dickerson 1974	Five Japanese students' English over 12 months	/r/ and /l/
Edwards 2006	Two Vietnamese adults' English over 10 months	syllable structure
Hecht and Mulford 1979	One Icelandic child's English over 8 months	fricatives
Oldenkamp 2010	Eight Turkish, eight Arabic and eight Chinese speaking adults' Dutch over 15-18 months	phonology and morphosyntax
Rankin 1985	32 Spanish university students' English over 5 years	segmental perception and production
Sato 1984	Two Vietnamese boys' English over 10 months	syllable structure
Snow and Hoefnagel-Höhle 1982	33 children, teenagers and adults' Dutch over a year	phonology, morphology, vocabulary, syntax
Winitz, Gillespie and Starcev 1995	One Polish boy's English over 6 years, 8 months	accent
Wode 1978	Four German children's English over six months	phonology and morphology

L2 studies vary in length from less than a year to nearly seven, and data collection points vary considerably, with a clear trade-off for frequency of collection and study length. That is, the longer the study, the fewer the sessions (the Winitz et al. study involved five sessions). Large sample size also translates into a reduction of session frequency, e.g. Snow and Hoefnagel-Höhle collected data thrice during their large-sample study. With respect to data analysis, rather than transcription and tabulation concerning non-target-likeness, some studies (e.g. Winitz et al. and Derwing et al.) use native-speaker accent rating. Studies do not invariably start upon learners' initial exposure, i.e. at the initial state to capture the earliest stages; for example Sato's study began after the boys had been exposed to English for half a year, and Edwards' study started after the couple had been exposed to English for a year. Nor are studies are always of naturalistic learners, yet as in morphosyntax a strong case can be made for the exclusion of instruction. In addition to encouraging a focus on forms not experienced by L1 children and younger L2 learners, instruction can mean extensive, though not necessarily intensive (during foreign language learning at school) exposure to non-native input along with considerable exposure to written input. Winitz et al. (1995:124) argue that age differences reported in the literature can be attributed to older L2 learners' pressure to produce language from the start as well as to their pre-immigration exposure, which according to them, is underreported despite 'every likelihood that learners will have picked up their teachers' non-native accents'. There is too little known about these effects and it therefore makes good research sense to exclude these instructional variables when feasible. However, when it comes to adults, this has typically meant studying immigrants whose input is often insufficient.

4. THE NEXT STEPS IN THE LONGITUDINAL STUDY OF L2 PHONOLOGY

4.1. OT

Under one current theory of phonology, Optimality Theory (Prince and Smolensky 1993), treatments of L2 phonology assume the existence of a universal set of constraints that remain available throughout the lifespan but whose ranking is initially that of the adult L1 phonology. This set then comes to be reranked as a result of target language input. In the process of constraint reranking there will – as has been attested in numerous (non-OT-based) studies over the last several decades – be interlanguage phonologies, i.e. rankings, that neither resemble the learner's L1 nor are that of the target language. Note that this assumption distinguishes discussion in L2 phonology from that in L2 morphosyntax where debate regarding access to UG is unrelenting. In fact, this is an unexpected stance for L2 phonologists to take: most SLA researchers assume that attaining native competence in an L2 in adulthood is least likely in phonology, and one would therefore expect the source of problems to relate to UG (non) access. This is where longitudinal studies are needed, a point recently noted by Hancin-Bhatt (2008:119): while empirical evidence for developmental effects is now substantial, 'linguistic-theorical analyses providing an account of these effects and *how they interact over time* are few, thereby limiting these analyses' predictability' (MYS' italics).

4.2. Comprehensive studies

Continued interest in longitudinal studies in morphosyntax indicates the readiness and the need for a replication (or perhaps re-examination of the data) of Snow and Hoefnagel-Höhle's (1982) large sample study of 33 Dutch-immersed children, teenagers and adults, which remains unique in its comprehensiveness. The test battery covered auditory discrimination, pronunciation, morphology, vocabulary and syntax, and the results showed a steady progression, with older learners initially faster in morphology and phonology. Age differences in L2 phonology levelled out by the end of the year, and although younger arrivals overtook the older ones in pronouncing some sounds, when 11 were tested after 1 ½ years, all but one teenager still had an L1 (English) accent. Where data exist on a range of phenomena, one would now want to look for developmental relationships among them. Using a large-sample synchronic corpus of L2 English and L2 German learners whose initial exposure ranged from age three to 33, Gut (2009) looks at relationships between morphology and phonology and more, using spontaneous speech, a reading passage, nonsense word reading and questions on extralinguistic factors (exposure, motivation, attitude, musical and acting ability). Gut concludes that variation is highly systematic, that acquisition involves a set of mutually dependent factors, that age is no barrier to phonological acquisition and that a set of mutually dependent factors is involved. Edwards (2006) and Oldenkamp (2010) also look at phonology (syllable structure) and its relationship to inflectional morphology in their longitudinal studies, and also consider external factors. There has been little attention, however, to the relationship between vocabulary and phonology. For children, a vocabulary spurt coincides with acceleration in phonological development (Ingram 1976). This may relate to variability in syllable structure simplification processes. In his longitudinal study, Abrahamsson (2003) considers the recoverability explanation proposed by Weinberger (1987) who then noted that while L1 children exhibit little of the epenthesis which predominates in adult L2A, L1 early talkers exhibit much more epenthesis – and have larger vocabularies - than typically developing children. This points to the usefulness of investigating the density of L2 adults' phonological neighborhoods.

4.3. Naturalistic learners who receive sufficient input from the start

In my work with Anne Vainikka, we have attempted to address the problem of input insufficiency for naturalistic learners by identifying uninstructed post-puberty learners most likely to receive considerable input; these turned out to be secondary school exchange students on an American programme which expected no prior knowledge of any foreign language. Three L2 learners of German (aged 15, 16 and 17 at the start of the study) were selected and 11 half-day sessions took place with them over 12 months. A range of data were collected using interviewing techniques, elicited production, judgement and comprehension tasks. The learners also provided monthly self-evaluations of their German listening, speaking, reading and writing while they lived with host families and attended German secondary schools as matriculated students (without German-as-a-second-language classes). Only some of the data (from a final devoicing task) have thus far been analysed with respect to phonology, but a full phonological analysis will empirically ground the proposal in Vainikka and Young-Scholten (1998) that adults' in inability to use inflectional suffixes to trigger syntactic development is what distinguishes them from younger learners. Morphosyntactic development supports this proposal, but more needs to be said about the contribution of phonology.

4.4. Do L2 adults resemble atypically developing L1 children?

One line of inquiry in L2 research on morphosyntax has involved asking whether L2 acquisition by older learners resembles that of atypically developing L1 children (most recently see Marinis to appear). In L2 phonology, the question is whether developmental processes resemble that of the atypically developing children discussed for example in Ingram (1976) who either exhibited prolonged use of typically developing children's processes or who followed unique processes.

4.5. A final note on methodology

PHON is MacWhinney and Rose's CHILDES-inspired creation of data management tools (leading to an eventual databank, PhonBank) for phonology. As one might imagine, it currently overwhelmingly covers L1 phonology. What would be the ideal type of longitudinal study that might appear in PhonBank? Lest the researcher (still) be deterred, in their survey of longitudinal studies in SLA, Ortega and Iberri-Shea (2005) emphasize the value of gaining a sufficiently lengthy picture of L2 acquisition to observe development. Length of investigation and the frequency of sessions will depend on the expected time-scale over which phenomena under investigation unfold. The authors urge researchers to exhaustively analyse their longitudinal data and in so doing also apply statistics that avoid the trap of treating a longitidinal study like repeated cross-sectional comparisons.

5. REFERENCES

Akita, M. 1998. A longitudinal study of Japanese EFL learners' interlanguage phonology. Paper presented at EUROSLA 8.

Abrahamsson, N. 2003. Development and recoverability of L2 codas: A longitudinal study of Chinese-Swedish interphonology. *Studies in Second Language Acquisition* 25:313-349.

Carlisle, R. 1998. The acquisition of onsets in a markedness relationship: A longitudinal study. *Studies in Second Language Acquisition* 30:245-260

Carroll, S. 2001. *Input and Evidence. The Raw Material of SLA*. Amsterdam:Benjamins.

Clahsen, H., Muysken, P. 1986. The availability of Universal Grammar to adult and child learners- A study of the acquisition of German word order. *Second Language Research* 2:93-119.

Derwing, T. M., Munro, J., Thomson, R. I. 2008. A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics* 29: 359-380.

Dickerson, W. 1974. *Internal and external patterning of phonological variability in the speech of Japanese learners of English*. Unpublished PhD dissertation, University of Illinois.

Eckman, F. 2008. Typological markedness and second language phonology. In Edwards, J. G. H., Zampini, M. L. (eds.) *Phonology and Second Language Acquisition*. Amsterdam: Benjamins.

Edwards, J. G. H. 2006. *Acquiring a non-native phonology. Linguistic Constraints and Social Barriers*. London: Continuum:

Fikkert, P. 1994. *On the Acquisition of Prosodic Structure*. Amsterdam: Holland Institute of Generative Linguistics.

Goad, H., White, L., Steele, J. 2003. Missing surface inflection in L2 acquisition: A prosodic account. In Beachley, B, Brown, A, Conlin, F. *21st BUCLD*. Somerville, MA: Cascadilla Press. Pp. 264-275.

Gut, U. 2009. Non-native Speech. A Corpus-based Analysis of Phonological and Phonetic Properties ofL2 English and German. Frankfurt am Main: Peter Lang.

Haberzettl, S. 2005. *Der Erwerb der Verbstellungsregeln in der Zweitsprache Deutsch durch Kinder mitrussischer and türkischer Muttersprache*. Tübingen: Niemeyer.

Hancin-Bhatt, B. 2008. Second language phonology in optimality theory. In Edwards, H. J., Zampini, M. L. (eds.) *Phonology and Second Language Acquisition*. Amsterdam: Benjamins. Pp. 117-146.

Haznedar, B., Schwartz, B.D. 1997. Are there optional infinitives in child L2 acquisition? In Hughes, E., Hughes, M., Greenhill, A. (eds.). *21st BUCLD.* Somerville, MA: Cascadilla Press. Pp. 257-268

Hecht, B., Mulford, F. 1979/1987. The acquisition of second language phonology: Interaction of transfer and developmental factors. Paper presented at BUCLD 4; published in Ioup and Weinberger 1987.

Huebner, T. 1983. *A Longitudinal Analysis of the Acquisition of English*. Ann Arbor, MI: Karoma.

Ingram, D. 1976. *Phonology Disability in Children*. London: Whurr.

Ingram, D. 1989. *Child Language*. Cambridge: CUP.

Ioup, G., Weinberger, S. 1987. *Interlanguage Phonology*. Rowley, MA: Newbury House.

Kramsch, C. 2003. Second language acquisition, applied linguistics and the teaching of foreign languages.*Language Learning Journal* 27:66-73.

Lardiere, D. 1998. Dissociating syntax from morphology in divergent end-state grammars. *Second Language Research* 14:359-375.

Lleó, C., Prinz, M. 1997. Syllable structure parameters and the acquisition of affricates. In Hannahs, S.J., Young-Scholten, M. (eds.) *Focus on Phonological Acquisition*. Amsterdam:Benjamins.

Long, M. 1990. Maturational constraints on language development. *Studies in Second Language Acquisition* 12:251-285.

Lust, B. C., Foley, C. 2004. *First Language Acquisition: The Essential Readings*. London: Blackwell.

Marinis. T. (in press). Syntactic processing in developmental and acquired language disorders. In: Ball, M. J. Perkins, M, Müller, M., Howard, S. (eds.), *The Handbook of Clinical Linguistics*. Blackwell:Oxford.

Meisel, J. 2003. Results in language acquisition research. The developmental problem. *Lingue e Linguaggio* II: 321-339.

Ortega, L, Iberri-Shea, G. 2005. Longitudinal research in second language acquisition. Recent trends and future directions. *Annual Review of Applied Linguistics* 25:26-45.

Pater, J. 1997. Metrical parater missetting in second language acquisition. In SJ Hannahs, Young-Scholten, M. (eds.) *Focus on Phonological Acquisition*. Amsterdam:Benjamins. pp. 235-261.

Pharr, A. B., Bernstein Ratner, N., Rescorla, L. 2000. Syllable structure development of toddlers with expressive specific language impairment. *Applied Psycho- linguistics* 21:429-449.

Pienemann, M. 1981. *Der Zweitspracherwerb ausländischer Arbeiterkinder*. Bonn: Bouvier Verlag Herbert Grundmann.

Prévost, P., White, L. 2000. Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research* 16:103-134.

Prince, A, Smolensky, P. 1993. *Optimality Theory: Constraint Interaction in a Generative Grammar*. Rutgers University Archive.

Rankin, I. A. G. 1985. Error Analysis at the Phonological Level and its Pedagogical Implications. Mendoza, Argentina: Universidad Nacional de Cuyo.

Rose, Y. 2000. *Headedness and prosodic licensing in the first language acquisiton of phonology*.Unpublished PhD thesis, McGill.

Sato, C. J. 1984. Phonological processes in second language acquisition: Another look at interlanguage syllable structure. *Language Learning* 34:43-57.

Smith, N. V. 1973. *The Acquisition of Phonology: A Case Study*. Cambridge: CUP.

Snow, C. E., Hoefnagel-Höhle, M. 1982. Age differences in the pronunciation of foreign sounds. In Krashen, S., Long, M., Scarcella, R. (eds.) *Child-adult Differences in Second Language Acquisition.* Rowley, MA: Newbury House.

Tracy, R., Thoma, D. 2007. Convergence on finite clauses in L1, bilingual and early L2 acquisition. Paper presented at the DGfS conference, Siegen.

Unsworth, S. 2005. *Child L2, Adult L2, Child L1: Differences and Similarities. A study on the acquisition of direct object scrambling in Dutch*. Leiden: LOT. .

Vainikka, A., Young-Scholten, M. 1994. Direct access to X'-Theory: Evidence from Korean and Turkish adults learning German. In Hoekstra, T, Schwartz , B. D. (eds.) *Language Acquisition Studies in Generative Grammar:* Amsterdam: Benjamins.

Vainikka, A., Young-Scholten, M. 1998. Morphosyntactic triggers in adult SLA. In Beck, M-L. (ed.) *Morphology and its Interfaces.* Amsterdam: Benjamins. pp. 89-113.

Weinberger, S. 1987. Functional and phonetic constraints on syllable structure simplification. In Ioup, G., Weinberger, S. (eds.) *Interlanguage Phonology*. Rowley,MA: Newbury House. Pp. 401-417.

Winitz, H., Gillespie, B, Starcev, J. 1995. The development of English speech patterns of a 7-year-old Polish-speaking child. *Journal of Psycholinguistic Research* 24:117-143.

Wode, H. 1978. Developmental sequences in naturalistic L2 acquisition. In Hatch, E. (ed.) *Second Language Acquisition*. Rowley, MA: Newbury House.

http://childes.psy.cmu.edu/phon/ accessed 4 April 2010

Unexpected Effects of the Second Language on the First

Noriaki Yusa, Kuniya Nasukawa**, Masatoshi Koizumi***, Kim Jungho***, Naoki Kimura*** and Kensuke Emura****

*Miyagi Gakuin Women's University, **Tohoku Gakuin University, ***Tohoku University
n_yusa@me.com, nasukawa@tscc.tohoku-gakuin.ac.jp

ABSTRACT

This paper discusses the developmental paths of laryngeal-source contrasts in native Japanese children learning L2 English. In particular, we focus on unexpected effects of the second language (L2) on the first (L1) in VOT production. Bilinguals who are proficient in both languages and spend considerable time using their L2s tend to produce "compromise" VOT values: L2 VOT values "gravitate" toward L1 values, while L1 VOT values do so toward L2 VOT values. In our longitudinal study of Japanese-speaking children learning English, however, we have found a stage unreported in previous literature where L1 values do not move toward L2 values but shift in a reverse way. This supports the hypothesis that two languages affect each other, but even L2 learners who do not produce compromise VOT values due to lack of amount of L2 exposure unconsciously modify their L1 VOT values, thereby enhancing the phonetic contrast of the two languages in a clever way.

Keywords: Voice Onset Time (VOT), voiceless stops, L2 effects on the L1

1. LARYNGEAL SOURCE CONTRASTS: VOT

This section presents an overview of cross-linguistic laryngeal-source contrasts between English and Japanese. Both English and Japanese show a two-way laryngeal-source contrast between 'voiceless' and 'voiced'. However, these terms alone are insufficient for describing the true nature of the phonetic differences between 'voicing' contrasts in English and Japanese. These differences can be more properly expressed in terms of voice onset time (VOT), which describes the interval between the release of a stop closure and the onset of vocal fold vibration (Lisker and Abramson 1964; Abramson and Lisker 1970).

TABLE I. The Mean VOT in English and Japanese word-initial stops by monolingual adults (Harada, 2007: 372)

	closure			release					
English				*b* (7)	*d* (19)	*g* (22)	*p* (68)	*t* (80)	*k* (88)
Japanese	*b* (–27)	*d* (–34)	*g* (1)	*p* (24)	*t* (26)	*k* (42)			
	voicing-lead –VOT			neutral			voicing-lag +VOT		

Table I lists mean VOT values (ms) in English and Japanese word-initial stops by monolingual adults. Because the English voiceless stops /p, t, k/ display a relatively long time lag between closure release and the onset of voicing, they are usually termed voiceless aspirated. On the other hand, the English voiced stops /b, d, g/ are termed voiceless unaspirated because the time lag between closure release and the onset of voicing is relatively short. By contrast, Japanese voiceless stops /p, t, k/ are produced with VOT values similar to those found in the English 'voiced' stops, while the Japanese voiced stops /b, d, g/, which are described as truly voiced, display a relatively long lead time between the onset of voicing and stop release. (In precise terms, English voiced stops and Japanese voiceless stops show different time lags; however, they are often considered to belong to the same VOT region because they show similar phonological behaviors.)

2. L2 Effects on the L1

Knowledge of language is a property of an individual brain, which underlies language production and comprehension (Chomsky 1957). Once this biolinguistic perspective is taken on bilingualism research, knowledge of two languages of a bilingual is also assumed to be a property of the individual brain. It is from this biolinguistic perspective that one can most clearly see a convergence between linguistic theory (generative grammar) and second language acquisition (SLA) research. The biolinguistic research program of second language acquisition suggests that cross-linguistic influences (*aka* transfers) are not exclusively from L1 to L2 but should be bidirectional between L1 and L2 (for bidirectional cross-linguistic influences manifest in many domains of language and cognition, see Cook 2003; Jarvis and Pavlenko 2008). With respect to VOT values, for example, proficient bilinguals tend to exhibit intermediate or "compromise" VOT values between L1 and L2 (for key findings in the VOT literature, see Zampini 2008): L2 VOT values tend to be attracted toward L1 VOT values, while L1 VOT values of bilinguals toward L2 values (Birdsong 2006: 22). However, inexperienced L2 learners do not differentiate L1 and L2, equating these two languages (Flege 1987a): their VOT values are like those of monolinguals. It is worth noting here that most studies on L2 effects on the L1 (or L1 attrition) have focused on L2 speakers in L2 migrant settings who spend considerable time interacting in the L2. To the best of our knowledge, few studies have been made on the reverse cross-linguistic effects in child L2 learners residing in the L1 community.

3. Experiment

3.1. Participants

We report a new finding on the unexpected effects of L2 on L1 when L2 learners are NOT proficient: the production of VOT for /p, t, k/ in English and Japanese by Japanese-speaking children first exposed to English at the age of four in Japan. The children were divided into three groups, according to the amount of English as a foreign language (EFL) in their kindergartens: the immersion group (n=40; mean age 5.48 yrs) had almost all instruction conducted in English; the regular-exposure group (n=18; mean age 5.55 yrs) had a 15-to-45-minute English lesson every day by a native speaker of English; the occasional-exposure group (n=49; mean age 5.50 yrs) had a native English speaker play with them for limited time. Thirty children (mean age 5.41 yrs) without exposure to English participated as a control group to investigate Japanese VOT values.

3.2. Tasks

In order to examine how language experience (early exposure to English) in different EFL settings influences Japanese children's phonetic norms of English and Japanese VOT values, we measured their VOT values after 20-month (1st measurement) and 32-month (2nd measurement) exposure to English. The test words were all selected on the basis of the following criteria from Harada (2007: 354): "(1) the vowel quality ([a] for Japanese words or [æ] for English words), (2) disyllabic words, (3) the same pitch accent or stress pattern (HL for Japanese VOT data, LH for singletons and LHH for geminates, and stress on the first syllable for English VOT data) and (4) concrete words." Using a picture card depicting what a target word refers to, an experimenter asked a participant to repeat three times each word starting with stops listed in the VOT corpus (Table II):

Experimenter (showing a cue picture depicting a panda): “What's this?”

Participant (looking at the picture): “Panda, panda, panda”

The VOT value of the target word in the second repetition out of the three repetitions was used for VOT measurements.

TABLE II: VOT Corpus (Harada 1999, 2007)

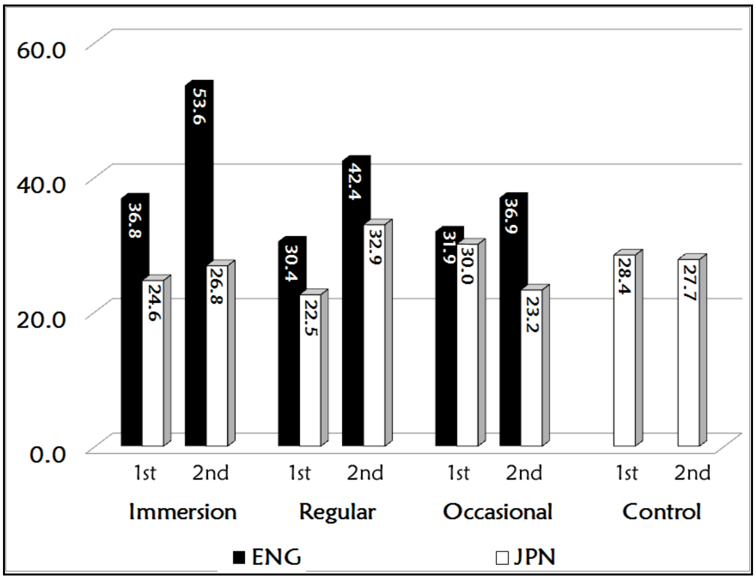
English			Japanese		
/p/	/t/	/k/	/p/	/t/	/k/
panda	tablet	carrot	pari	tako	kame
parrot	tadpole	camel	papa (x2)	tane	kasa
package	taxi	candy		tate	kata

3.3. Results

The data were analyzed using Multi-speech (KayPENTAX, Model 3700) software. The VOT of stops was measured by finding the nearest millisecond from the beginning of the release burst to the onset of voicing energy in F2 formants with the occasional use of the waveform.

The repeated measures of 3x2x2x3 ANOVA (Period of Exposure, Years, Languages, Places of Articulation) showed significant interaction between languages and the amount of English exposure (Figure I) ($F(2,104) = 4.492$, $p < .01$). At the 1st measurement the immersion children produced English stops with *marginally longer* VOT ($F(2,208) = 2.632$, $p=.074$), but at the second measurement significantly longer VOT than the regular and occasional-exposure groups ($F(2,208) = 11.683$, $p<.001$), which suggests that it takes more time and exposure to influence L2 English VOT (Flege 1987a). At the 1st measurement the Japanese VOT for the regular-exposure group was *shorter* than that for the occasional-exposure and control groups ($F(2,416) = 3.881$, $p<.05$), but their English VOT was the same as the Japanese VOT for the occasional-exposure group. Interestingly, in the first measurement the immersion group produced Japanese stops with *significantly shorter* VOT ($F(2,416) = 18.966$, $p<.001$) while their English VOT was significantly longer than Japanese VOT for the occasional-exposure group ($F(2,208) = 4.644$, $p<.05$).

FIGURE I: The mean VOT in English and Japanese word-initial stops by four groups



4. Discussion and Conclusion

At the first measurement, the participants in the immersion and regular-exposure groups produced significantly longer VOT values for English stops than Japanese ones ($F(1,208) = 25.92$, $p<.001$, $F(1,208) = 11.03$, $p<.05$, respectively), but no difference between English and Japanese was observed in the occasional-exposure group ($F(1,208) = 0.65$, $p=.42$). The results suggest that the former groups made phonetic distinctions between two languages, but the occasional-exposure group employed L1 VOT values for both L1 and L2. Regarding English VOT values, the immersion and regular-exposure groups, rather than using compromise Japanese VOT values, produced Japanese stops with significantly shorter VOT values than monolingual Japanese speakers. This finding cannot be explained by any L2 phonology literature implying that exposure to English with long-lag VOT *lengthens* Japanese VOT (Flege 1987a; Harada 1999). We propose that frequent and regular-exposure to English made the Japanese children sensitive to phonetic differences between two languages and that in the initial phase they made clearer cross-linguistic distinctions by shortening L1 Japanese VOT rather than lengthening L2 English VOT. This suggests that in the initial phase of bilingualism where L2 learners are more dependent on L1 than L2, L1 may be more subject to modifications than L2. The same tendency can be seen in the 2nd measurement results of the occasional-group, whose English VOT was marginally longer ($F(1,208) = 3.07$, $p=.081$), but Japanese VOT was *significantly shorter* than those of the 1st measurement ($F(1,208) = 5.85$, $p<.05$). Our results suggest that L2 affects the phonetic production of L1 even when users are not proficient: inexperienced L2 try to dissimilate L1 VOT values from L2 VOT values and form a new category for the L1. In other words, they seem to resort to an overriding principle in L2 phonology (Eckman et al. 2003: 190), which prevents "the complete neutralization of a contrast." The results presented here support the biolinguistic hypothesis that L2 learners exhibit bidirectional transfer effects in their two languages, and our data show that even in a stage where L2 effects are not apparent, L2 affects L1 in an interesting way.

5. REFERENCES

Abramson, A.S., Lisker, L. 1970. Discriminability along the voicing continuum: Cross-language tests. In: B. Hála, M. Romportl M., Janota, P. (eds.), *Proceedings of Speech Communication*. London: Academic Press. 127-134.

Birdsong, D. 2006. Age and second language acquisition: A selective overview. In: M. Gullberg, P. Indefrey. (eds.), *The Cognitive Neuroscience of Second Language Acquisition*. Oxford: Blackwell. 9-49.

Chomsky, N. 1957. *The Syntactic Structures*. The Hague: Mouton.

Cook, V. (ed.). 2003. *Effects of the Second Language on the First*. Clevedon: Multilingual Matters.

Eckman, F.R., Elreyes, A., Iverson G. K. 2003. Some principles of second language phonology. *Second Language Acquisition* 19. 162-208.

Flege, J. 1987a. The production of "new' and "similar' phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of the Acoustic Society of America* 76. 692-702.

Flege, J. 1987b. Effects of equivalence classification on the production of foreign language sounds. In: A. James, J. Leather. (eds.), *Sound Patterns in Second Language Acquisition*. Dordrecht: Foris. 9-39.

Flege, J., Hillenbrand, J. 1984. Limits on phonetic accuracy in foreign language production. *Journal of the Acoustic Society of America* 76. 708-721.

Harada, T. 1999. *The Acquisition of Segmental Timing by Children in a Japanese Immersion Program*. PhD dissertation, UCLA.

Harada, T. 2007. The production of voice onset time (VOT) by English-speaking children in a Japanese immersion program. *International Review of Applied Linguistics in Language Teaching* (*IRAL*) 45. 353–378.

Jarvis, S., Pavlenko A. 2008. *Crosslinguistic Influence in Language and Cognition*. New York/London: Routledge.

Lisker, L. , Abramson, A.S. 1964. A cross-language study of voicing in initial stops: Acoustic measurement. *Word* 20. 384-422.

Lord, G. 2008. Second language acquisition and first language phonological modification. *Selected Proceedings of the 10th Hispanic Linguistic Symposium*. 184-193.

Zampini, M.L. 2008. L2 speech production research: Findings, issues, and advances. In: J. G. H. , M. L. Zamini. (eds.), *Phonology and Second Language Acquisition*. Amsterdam/Philadelphia: John Benjamins. 219-249.

Acknowledgements

This work has been supported by a Grant-in-Aid for promotion of “Brain Science and Education, Type II” from RISTEX (Research Institute of Science and Technology for Society) of JST (the Japan Science and Technology Agency) to Hiroko Hagiwara. We are grateful to Sanae Yamaguchi for helping us with the data analysis.

The retrieval of potentially morphologically complex clusters in English and Polish

Paulina Zydorowicz

Adam Mickiewicz University
zpaula@ifa.amu.edu.pl

ABSTRACT

The aim of the present study is to examine the retrieval of potentially morphologically complex clusters in the native production of Polish and non-native production of English. The study has been inspired by the proposal of morphonotactics, developed by Dressler and Dziubalska-Kołaczyk (2006). Dressler and Dziubalska-Kołaczyk place consonant clusters on a continuum between purely morphonotactic (morphologically driven) and lexical ones (monomorphemic). Thus the following groups of clusters can be distinguished: clusters which occur only across morpheme boundaries, clusters which by default occur at morpheme boundaries (one can find exceptions), clusters which occur both across morpheme boundaries and within morphemes, and clusters which occur exclusively within morphemes. The aim of the study was to check whether clusters, which have a potential of being morphonotactic in nature, tend to exist in the mind of language users as rather lexical or morphonological ones.

Keywords: phonotactics, morphonotactics, morphological boundary

1. MORPHONOTACTICS

The term morphonotactics, which has been introduced rather recently to cover the area of interaction of phonotactics and morphotactics (Dressler and Dziubalska-Kołaczyk 2006), in fact refers to the first of the three parts of morphonology introduced by Trubetzkoy (1931), that is, the study of the phonological structure of morphemes. In other words, morphonotactics can be understood as phonotactics at morpheme boundaries.

There are two sources of morphonotactic clusters, concatenative and non-concatenative (Dressler and Dziubalska-Kołaczyk 2006). The former consists in attaching suffixes to the root of the word, e.g. *cat+s* in English and *z+robić* 'to do-PERF.' This is the only mechanism of generating morphonotactic clusters in English and the default mechanism generating morphonotactic clusters in Polish. The other, non-concatenative, source of morphonotactic clusters occurring in Polish is reflected in such morphological operations as vowel ~ zero alternation, e.g. *pies* vs *psa*, or zero-Genetive-Plural formation, e.g. *miejsce* vs *miejsc*.

Dressler and Dziubalska-Kołaczyk place consonant clusters on a continuum between purely morphonotactic (morphologically driven) and lexical ones (monomorphemic). Thus the following groups of clusters have been distinguished.

1. Clusters which occur only across morpheme boundaries

   English example: final /fs vz/ in *laughs*, *loves*
   Polish example: initial /fs/ in *wsadzić* 'to put into sth'

2. Clusters which by default occur morpheme boundaries, however, one can find very few morphologically unmotivated clusters (a strong default)

   English example: final /ts dz/ in *cats, kids*
   EXCEPTION: final /ts dz/ in *quartz*, *adze*, respectively

   Polish example: initial /zr/ in *zrobić* 'to do'

EXCEPTION: initial /zr/ in *zraz* 'beef collop'

3. Clusters which by default occur morpheme boundaries, however, one can find quite a few morphologically unmotivated clusters (a weak default)

   English example: final /ks/ in *breaks*
   EXCEPTION: final /ks/ in Latinate words such as *mix, fix, six* etc.

   Polish example: initial /ɕtɕ/ in *ścinać* 'to cut down'
   EXCEPTION: *ściana* 'wall', *ścieg* 'stitch', *ścieżka* 'path' etc.

4. Clusters which occur both across morpheme boundaries and within morphemes

   English example: final /ld/ in bimorphemic *called* as well as monomorphemic *cold*
   Polish example: initial /sp/ in bimorphemic *spaść* 'to fall down' as well as monomorphemic *spacer* 'walk'

5. Clusters which occur exclusively within morphemes
   English example: final /lt/ as in *melt*
   Polish example: initial /pt/ as in *ptak* 'bird'

The aim of the study was to check whether clusters belonging to groups 2, 3, and 4 exist in the mind of language users as rather lexical or morphonological ones.

2. THE EMPIRICAL STUDY

2.1. Methodology

The subjects of the study were 40 Polish students of the first year of English Philology. They were in the process of receiving formal instruction in practical English phonetics as well as theoretical phonetic and phonological background of English and (to some extent) Polish. In the experiment the subjects were provided with an auditory presentation of word initial and final clusters, accompanied by the phonemic transcription of the clusters. The target clusters represented one of the groups listed above, i.e. strong default clusters (group 2), weak default clusters (group 3), and clusters which may function both as lexical and morphonotactic ones (group 4). Students were instructed to provide three examples of words that came to their minds when presented with a given cluster. The same procedure was applied to elicit English word final and Polish word initial and final clusters.

Table 1 below presents the stimuli used in the experiment. The status of the cluster was determined on the basis of word entries in two dictionaries: *English pronouncing dictionary* (Jones 2006) and *Uniwersalny słownik języka polskiego* (Dubisz et al 2003).

Table 1: The stimuli used in the experiment.

cluster type	English	Polish
strong default	ts dz nz lz kst nst	initial: ft zb zd zg zr zm spʃ final: sw
weak default	ps ks mpt ŋkt ŋks	initial: ɕtɕ zn final: ɲtɕ
both lexical and morphonotactic	pt kt ft st nd ld	initial: sp st sk ftʃ fʃ sf sx vr vw zw fst stʃ str zdr

		final: ɕtɕ

It is predicted that the more morphonotactic the cluster is, the more frequently it will trigger the retrieval of morphologically complex words.

2.2. Results

2.2.1. English

Figure 1 below presents the realization of individual clusters in English with dark-shaded fields denoting lexical responses and light-shaded fields denoting morphonotactic responses.

Figure 1: The reaslisation of individual clusters: English.

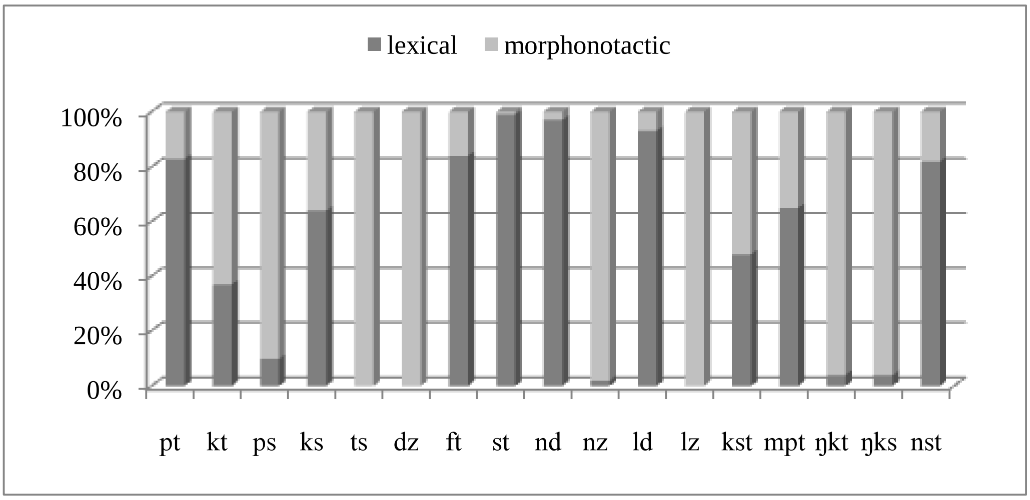
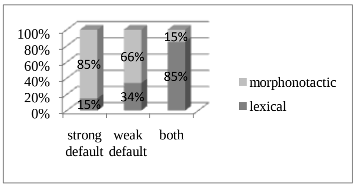


Table 2 below presents the number of the subjects' responses to the English stimuli, whereas Figure 2 presents the retrieval rates.

Table 2: The results: English.

		lexical	morpho.	missed
strong default	no.	43	252	425
weak default	no.	97	188	315
both lex. and mor.	no.	421	73	226
total	no.	561	514	965
	%	27.5%	25.5%	47%

Figure 2: The results: English.



The results show that the retrieval of clusters is dependent on the category a given cluster belongs to. Thus clusters which are morphonotactic by strong default are associated with words containing a morphological boundary in the cluster. For instance, final cluster /ts/ was always associated with words such as *cats*, *rats*, *boots* etc., and so were the remaining strong default clusters with the exception of /nst/ and /kst/, which were usually associated with the monomorphemic words, e.g. *against* and *next* respectively.

Consonant clusters which are weak defaults, i.e. one may found quite a few lexical exceptions, were rather realized as morphonotactic (66%), though the aforementioned exceptions were evoked by the subjects relatively frequently too (34%).

Finally consonant clusters which can both occur intramorphemically and intermorphemically were to large extent (85%) associated with words whose target cluster was lexical in nature. This suggests that unsurprisingly, in a lexical search monomorphemic consonant clusters are preferred (especially that the lexical search for English words was an *a tergo* one, which posed additional difficulties for the subjects).

Finally, let us analyze the number of missed answers. The task allowed for 2040 responses (17 clusters x 3 requested responses x 40 subjects), out of which the subjects provided 1075 words, i.e. 53%. These results mean that almost half of the answers was missed, i.e. the subjects were not able to produce the words or their answers did not qualify (the words did not include the target cluster). In terms of the 3 categories of clusters, the responses for strong default, weak default and both inter- and intramorphemic clusters were missed in 59%, 52%, 31% of the cases respectively. This means the retrieval of clusters gets more difficult, the more morphonotactic the cluster becomes. The explanation lies in the fact that the more morphonotactic the cluster is, the fewer lexical responses the subjects have at their disposal and the more frequently they must evoke morphological rules, which costs more effort on the part of a language user.

2.2.2. Polish

Figure 3 below presents the realization of individual clusters in Polish with dark-shaded fields denoting lexical responses and light-shaded fields denoting morphonotactic responses.

Figure 3: The realisation of individual clusters: Polish initials.

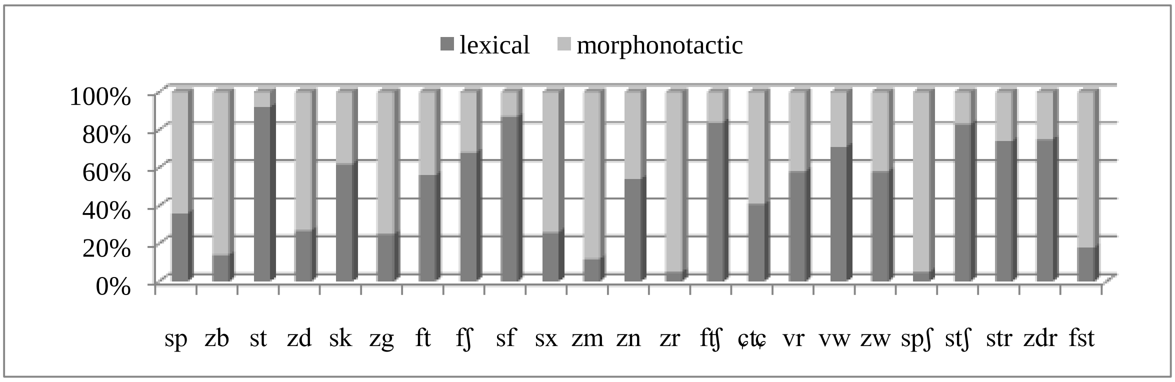
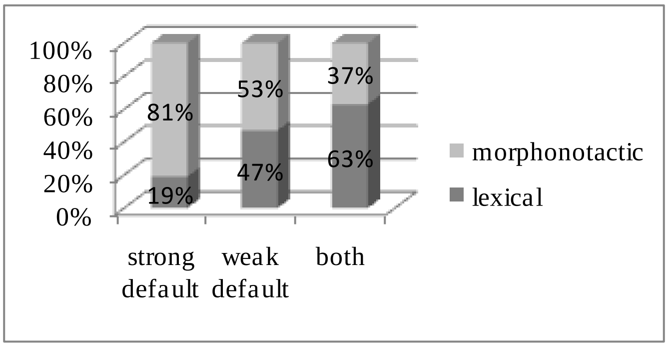


Table 3 presents the number of the subjects' responses to the Polish stimuli, whereas Figure 4 presents the retrieval rates.

Table 3: The results: Polish initials.

		lexical	morpho.	missed
strong default	no.	145	606	89
weak default	no.	99	111	30
both lex. and mor.	no.	948	547	185
total	no.	1192	1264	304
	%	43%	46%	11%

Figure 4: The results: Polish initials.



Similarly to the English data, the retrieval of words in Polish (L1) depends on the status of a cluster. Thus strong default clusters were associated with words whose target cluster structure is bimorphemic. This results in 19% of lexical responses and 81% of morphonotactic responses. The only exception in this category concerns cluster /ft/ which was realised as lexical in 56% of the cases. This can be explained by the high frequency of such words as *wtorek* 'Tuesday'or *wtem* 'suddenly' which were often retrieved by the subjects. Weak default clusters were associated with words whose target cluster is mono- or bimorphemic to a similar extent (47% of lexical responses and 53% of morphonotactic responses). Finally, clusters which are both lexical and morphonotactic were rather associated with lexical responses (63%). The exceptions within this group include cluster /fst/, and /sp/. Initial /fst/, which was associated with morphonotactic responses, occurs lexically and morphonotactically only in a couple of words (+ their derivatives). However, in the retrieval task the subjects found it easier to evoke morphonotactically complex words, possibly because they were more frequent or because the marked shape of the cluster (its size and phonetic shape) signalled a morphological boundary. The morphonotactic realisation of /sp/ is rather unexpected.

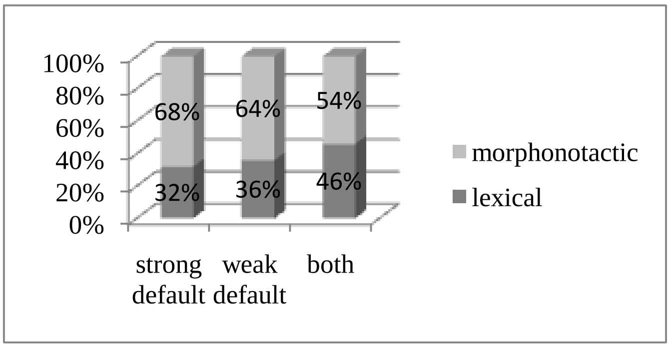
Let us now analyse the number of missed answers. The experimental task allowed for 2760 responses (23 clusters x 3 requested responses x 40 subjects), out of which the subjects provided 2456 words, i.e.89%. Thus the subjects performed much better in the retrieval of Polish initials than in the retrieval of English finals. The explanation for this difference may be twofold. Firstly, the lexical search according to words onsets is easier than an *a tergo* one (which had to be applied by the subjects in the case of English). Secondly, English final morphonotactic clusters have to be retrieved by applying inflectional rules to the word stems, whereas in Polish morphonotactic clusters may but do not have to be generated online, i.e. many morphologically complex words may be accessed directly from the lexicon, e.g. *schody* (a 3-morphemic word s+chod+y) exemplifying initial /sx/. Thus such words, though morphologically complex, are more easily accessible than, e.g. past tense forms in English. In terms of the 3 categories of clusters, the responses for strong default, weak default and both inter- and intramorphemic clusters were missed in 11%, 12.5%, 11% of the cases respectively. In the Polish lexical search of initials all groups of clusters were retrieved with equal ease.

As far as final clusters are concerned, the Polish task included three word final clusters: strong default /sw/, weak default /ɲtɕ/, and lexical and morphonotactic /ɕtɕ/. Table 4 below presents the number of the subjects' responses to the Polish stimuli, whereas Figure 5 presents the retrieval rates.

Table 4: The results: Polish finals.

		lexical	morpho.	missed
strong default /sw/	no.	23	50	47
weak default /ɲtɕ/	no.	14	25	81
both lex. and mor. /ɕtɕ/	no.	47	55	18
total	no.	84	130	146
	%	23%	36%	41%

Figure 5: The results: Polish finals.



Within the group of Polish finals, one can observe that the number of lexical responses increases slightly as the cluster moves on the continuum from strong default, through weak default, to lexical / morphonotactic one.

As regards the number of missed answers, the experimental task allowed for 360 responses (3 clusters x 3 requested responses x 40 subjects), out of which the subjects provided 214 words, i.e. 59%. In terms of the 3 categories of clusters, strong default, weak default and both inter- and intramorphemic clusters were missed in 39%, 67.5%, 19% of the cases respectively. The high reduction rates in the group of the weak default may be ascribed to the fact that the final cluster /ɲtɕ/ must be decoded from the following orthography: ą + ć, ę + ć, ą + dź, ę + dź. The decoding process is impeded by the fact that the first element of the cluster, i.e. /ɲ/ must be extracted from the vowel. Otherwise, alike in the case of English finals and Polish initials, words from the lexical / morphonotactic group were retrieved most successfully.

2.3. Conclusion

The aim of the study was verify the hypothesis that the more morphonotactic a cluster is, the more frequently it will be associated with words whose target cluster is morphologically driven. The study corroborated this assumption in all tested groups of clusters, i.e. English finals, Polish initials and Polish finals. The study also revealed that occasionally a high frequency of words whose cluster structure is morphologically simple (though the cluster itself belongs to the strong default) or difficulties with the application of a morphological rule may override the criterion of the cluster status.

What is remarkable is that clusters from group 4 (both lexical and morphonotactic) which have equal chances of triggering words containing a morphological boundary in a cluster or devoid of it, tend to trigger morphonotactically simple responses. The observation holds for Polish initials (63% of lexical activations vs 37% of morphonotactic activations) as well as for English finals (85% of lexical activations vs 15% of morphonotactic activations).

3. REFERENCES

Dressler, W. U., Dziubalska-Kołaczyk, K. 2006. Proposing Morphonotactics. *Rivista di Linguistica* 18.2: 249-266.

Dubisz, S., Drabik, L., Sobol, E., Grzegółka-Maciejewska, A. 2003. *Uniwersalny słownik języka polskiego* [A universal Polish dictionary]. CD-ROM. Warszawa: Wydawnictwo Naukowe PWN.

Jones, D. 2006. *English Pronouncing Dictionary*. (17th edition, edited by Peter Roach, James Hartman and Jane Setter.) Cambridge: Cambridge University Press.

Trubetzkoy, Nikolaj S. 1931. Gedanken über Morphonologie. *TCLP*. Vol. 4: 337-342.