DO YOU HEAR WHAT I HEAR?: A COMPARISON OF PHONEME PERCEPTION IN NATIVE AND SAARLANDIAN GERMAN NONNATIVE SPEAKERS OF ENGLISH

Carrie A. Ankerstein, Raphael Morschett, Anglistik, Saarland University

Second language (L2) phonological development is strongly influenced by the first language (L1), with L2 sounds being assimilated into the L1 phonological inventory. It has been argued that phones that are similar in the L1 and L2 will be most difficult to acquire whereas sounds that are not present in the L1 will be easier. In addition, commonly used substitutions in production can affect perception of L2 phonemes. In a forced-choice phoneme discrimination task, we explored how accurately Saarlandian German speakers of L2 English discriminate between /v/ and /w/; dental fricatives versus alveolar stops and fricatives; and initial position consonant cluster voicing. Nonnative speakers performed worst for the /v/-/w/ distinction followed by the dental/alveolar fricative and initial position consonant cluster voicing distinction. We conclude that different assimilation processes occur for these phonological phenomena.

KEYWORDS: second language phonology, assimilation, th-substitutions, phoneme perception

Die phonologische Entwicklung der Zweitsprache (L2) ist stark durch die Erstsprache (L1) beeinflusst, sodass L2 Laute in das phonologische Inventar der Erstsprache assimiliert werden. Es wurde argumentiert, dass Laute, die in der L1 und in der L2 ähnlich sind, am schwierigsten zu erwerben sind, wohingegen Laute, die in der L1 nicht auftauchen, einfacher sind. Des Weiteren können häufig verwendete Substitutionen in der Produktion die Wahrnehmung von L2 Phonemen beeinflussen. In einem Phonemdiskriminationsexperiment mit forcierter Entscheidung haben wir untersucht, wie gut saarländische Deutschmuttersprachler mit Englisch als L2 zwischen /v/ und /w/; dentalen Frikativlauten und Alveolarplosiven und -frikativen; und der Stimmhaftigkeit bei initialen Konsonantclustern unterscheiden können. Nichtmuttersprachler schnitten bei der /v/-/w/-Unterscheidung am schlechtesten ab, gefolgt von der Dental-/Alveolarfrikativunterscheidung und der Stimmhaftigkeitsunterscheidung bei initialen Konsonantclustern. Wir kommen zu dem Schluss, dass bei diesen phonologischen Phänomenen unterschiedliche Assimilationsprozesse stattfinden.

SCHLAGWÖRTER: Phonologie der Zweitsprache, Assimilation, th-Substitution, Phonemwahrnehmung

1 INTRODUCTION

In the course of phonological development, infants must learn which sounds in their native language (L1) are used phonemically, as meaning-distinguishing sounds, and which are not (Kuhl et al. 2006). From birth on (and maybe even in the womb, Gomez et al. 2000), infants have perceptual capacities that enable them to discriminate between different sounds and other properties of speech such as intonation and stress and to perceptually discriminate between all the phonetic units that exist in the world's languages (Eimas et al. 1971; Streeter 1976; Werker et al. 1984).

In the second half-year of life, the L1 begins to exert influence on perceptual processing (Kingston 2003; Kuhl et al. 2006; Werker et al. 1984). The perceptual apparatus is generally thought to be arranged in a way that promotes a shift of attention away from irrelevant information towards the detection of higher-level, functionally relevant structures (Gibson 1966). What this means in the phonetic context is that perceptual sensitivity increases at the category boundaries of phonemes in the L1 and decreases for those contrasts that are not present in the L1 (Best 1994).

Second language (L2) phonological development is strongly influenced by the L1 (Best et al. 2001; Flege 1991, 1995; Kuhl et al. 2006). Broadly speaking, quality of perception and production of L2 sounds depends on how similar the L2 phone is to an L1 phone. It has been argued that phones that are not present in the L1 will be easier to acquire and that phones that are similar in the L1 and L2 will be most difficult to acquire (Best 1994; Best et al. 2001; Flege 1991, 1995). For example, Flege (1987) found that English L1 speakers' production of the French high front rounded vowel /y/, absent in English phonology, was more authentic than the French high back rounded vowel /u/, a phone similar in English. Flege argued that English speakers recognized /y/ as a new vowel but assimilated the similar /u/ vowel into their native category, though it is not acoustically identical. Thus, they more easily acquired the novel phoneme in contrast to the similar phoneme.

Best et al. (2001) suggested a hierarchy of difficulty for the discrimination of two L2 phones depending on how they are assimilated with respect to the L1 phoneme inventory. They argued that phones that are assimilated into two different L1 phonemic categories are easiest to discriminate. Phones that are assimilated into one L1 phonemic category are most difficult. A third type of assimilation depends on "goodness of fit" in which a L2 phone is assimilated to a L1 phoneme that is close enough (see Best et al. 2001: 777). For example American listeners categorized both voiceless aspirated velar stops /k**H**/ and ejective velar stops /k**ɘ**/ as similar sounds on the basis of their goodness of fit to their native voiceless velar stop /k/ (Best et al. 2001). In a series of experiments, Best et al. (2001) tested the proposed hierarchy and found results to support it. Accuracy in phoneme discrimination was best for two-category assimilation and worst for single-category assimilation. Accuracy scores for "goodness of fit" assimilations fell between the single and two-category assimilation patterns.

Perception of L2 sounds is likely to affect production. It seems clear that L2 speakers might be able to perceive a phoneme in a native-like way without being able to produce it correctly (Flege 1991) and there is even some suggestion that production can affect perception (Hanulikova et al. 2012). Hanulikova et al. (2012) found different substitution preferences in production of the dental fricative /**T**/ in German and

Dutch speakers, /s/ and /t/ respectively, and found that this also carried over to their perception of these substitutes. In an eye-tracking study, they presented participants with real word stimuli, such as *theft*, *thrill* and *third*, and stimuli with substitutions, such as *teft*, *srill* and *fird*. When they heard the stimuli produced with substitutions, German and Dutch L2 speakers fixated longest on the correct target, produced with /T/, but they also fixated on words containing their preferred substitute longer than on any other distracters, including the perceptually more similar /f/, indicating successful lexical access for words produced with substitutions, e.g., Germans recognized *srill* as *thrill*. Hanulikova et al. (2012) argued the fixation preferences were determined by the frequency of the substitutions in the respective L2 speech.

The present study focuses on three differences between German and English phonology: dental fricatives; the /v/ - /w/ distinction; and initial position consonant cluster voicing. The dental fricatives were chosen in order to replicate the findings of Hanulikova et al. (2012) in a different task. The /v/-/w/ and dental fricative contrasts were chosen because there are clear problems in perception and production of these sounds (Eckert et al. 2005; Hamann et al. 2005; Iverson et al. 2008). German has a voiced labiodental fricative /v/ but lacks the approximant /w/. However, German speakers often produce little friction in their production of /v/ and produce something that sounds more like an approximant to English speakers (Iverson et al. 2008).

English and German also differ in their employment of consonant voicing as a function of position. Voicing contrasts exist in initial, medial and final positions in English. Standard German contrasts voicing in word-initial and word-medial positions, but lacks voicing contrasts in the word-final position for stops and fricatives. Thus, the German words for *wheel* and *advice*, *Rad* and *Rat*, respectively, are homophones. In terms of production, it has been found that German speakers of English do not produce the full acoustic features of the voicing contrast in the final position (Smith et al. 2009). In southwestern Germany, including the Saarland, it is a common practice to additionally devoice the consonant clusters /br/, /gr/, /bl/ and /gl/ in word-initial positions. Thus, Saarlandian articulations of the words *Blatt* (English: *leaf*/*sheet*) and *platt* (English: *flat*) are homophonic.

We explored how accurately Saarlandian German speakers of L2 English discriminate between /v/ and /w/; dental fricatives versus alveolar stops and fricatives; and initial position consonant cluster voicing in a forced-choice phoneme discrimination task roughly based on that used by Iverson et al. (2008). Based on personal experience, we expected discrimination to be poor for the /v/-/w/ contrast and for the initial voiced cluster contrast as many Saarlandian Germans, in informal tests, have indicated that they hear no difference between pairs such as *village - willage*; and *bride - pride*. According to Best et al. (2001) this would fall under single category assimilation and should yield high error rates. Based on Hanulikova et al.'s (2012) findings, we expected there to be a difference in discrimination accuracy between dental fricatives and alveolar fricatives, and dental fricatives and alveolar stops, with the fricative contrast being less accurate.

2 PARTICIPANTS

Twenty-one German students (12 male, nine female) between the ages of 18-28 years (mean age = 22) took part. A native speaker group functioned as the control group and consisted of eight English native speakers (five male and three female) between the ages of 18-26 years (mean age = 20). All German participants were recruited from Saarland University[1]. English native speakers were recruited from Saarland University and Leeds University. All participants gave written informed consent prior to taking part in the experiment.

3 STIMULI

Stimuli were spoken syllables, consisting of a consonant or consonant cluster followed by a low, back vowel /A/. All syllables were digitally recorded by a female native speaker of American English three times and the middle token was used. The stimuli were divided into conditions: the /vA/ - /wA/ distinction; dental and stop distinction: /ΤA/ - /tA/ and /ΔA/ - /dA/; dental fricative and alveolar fricative distinction: /ΤA/ - /sA/ and /ΔA/ - /zA/; voicing: /bA/ - /pA/, /dA/ - /tA/, /gA/ - /kA/, /fA/ - /vA/; and voicing for consonant clusters: /grA/ - /krA/, /brA/ - /prA/, /glA/ - /klA/ and /blA/ - /plA/. There were also filler items contrasting other manner distinctions, e.g., /mA/ - /bA/, and other place distinctions, e.g., /rA/ - /lA/, and sounds that contrasted manner and place, such as /kA/ - /hA/; /fA/ - /lA/ and /nA/ - /lA/. Each stimulus pair was repeated four times.

[1] Participants will from hereon be referred to as 'Saarlandian'. By virtue of studying at Saarland University, they are speakers of, or have likely heard, the feature of initial position consonant cluster devoicing of Saarlandian dialects of German. This assumption was upheld in the data – the nonnative speaker group had significantly higher ERs for the voicing cluster condition than the native speaker group. Future studies will contain a background questionnaire to formally establish the participants' language background.

4 APPARATUS AND PROCEDURE

Participants were tested individually in a quiet room. One syllable, e.g., /TA/, was played in stereo over headphones, while two orthographic syllables, e.g. 'tha' and 'ta', were shown on screen. The participants were asked to indicate as quickly and as accurately as possible which one of the syllables was heard by pressing either the left shift key (syllable on the left was heard) or the right shift key (syllable on the right was heard). Participants heard each syllable once per trial. The stimuli were presented in two blocks with a break at the half-way point. Each block contained 56 items. There was a short practice phase with eight stimuli before the full experiment started. Instructions were given in English. The experiment was run using DMDX 3.2.6.4 (Forster et al. 2003).

5 RESULTS

Mean error rate was calculated per condition. Results are presented in Table 1.

Table 1: Mean error rates as a function of condition and participant group. Standard deviations are given in brackets.

	/v/ - /w/	**<th> – fricative**	**<th> - stop**	**voicing cluster**	**voicing**	**filler**
Nonnative (N = 21)	40.9 (21.2)	26.1 (21.4)	4.0 (7.1)	19.6 (17.1)	4.2 (5.2)	2.8 (3.8)
Native (N = 8)	0 (0)	9.4 (14.6)	1.5 (4.4)	0.8 (2.2)	2.3 (3.2)	1.0 (3.0)

Error rate data were entered into a repeated measures ANOVA with Condition (/v/-/w/, <th>-fricative, <th>-stop, voicing, voicing cluster and filler) as a within-subjects variables and Group (native, nonnative) as a between-subjects variable. There was a significant effect of Condition ($F(5, 140) = 10.021$, $p < 0.001$), indicating significant differences in error rates across conditions. There was a significant interact effect between Condition and Group ($F(5, 140) = 9.001$, $p < 0.001$). There was a significant effect of Group ($F(1,28) = 25.579$, $p < 0.001$), indicating that native speakers had significantly lower error rates (mean = 2.5%) than nonnative speakers (mean = 16.3%).

In order to explore the Condition x Group effect, a separate repeated measures ANOVA was run for the native speaker group in order to explore differences in their performance across the five conditions. The effect for Condition approached significance ($F(5, 35) = 2.632$, $p = 0.040$). Post hoc pair-wise t-tests yielded no significant effects, all $p > 0.05$, indicating statistically similar error rates across all conditions. Thus, the significant effect found for Condition was due to the nonnative group.

Condition effects for the nonnative group were explored through a series of pair-wise t-tests. There was no significant difference in the nonnative group for the <th>-stop, voicing and filler conditions, all $p > 0.05$. There was also no significant difference in the nonnative group for the <th>-fricative and voicing cluster conditions ($t(21) = 1.322$, $p = 0.200$). These results indicate statistically similar error rates for these conditions. These conditions were grouped, <th>-stop/voicing/filler and th-fricative/voicing cluster for subsequent analysis to reduce the number of comparisons run.

There was a significant difference between the /v/-/w/ and <th>-stop/voicing/filler conditions ($t(21) = 7.982$, $p < 0.001$) and the /v/-/w/ and <th>-fricative/voicing cluster conditions ($t(21) = 3.650$, $p < 0.001$) indicating significantly higher ERs for the /v/-/w/ condition in contrast to all other conditions. There was also a significant difference between the <th>-stop/voicing/filler and <th>-fricative/voicing cluster conditions ($t(21) = 5.936$, $p < 0.001$) indicating significantly higher ERs for the <th>-fricative/voicing cluster conditions. ERs for the nonnative speaker group were not significantly different to those of native speaker controls for the <th>-stop/voicing/filler condition ($t(28) = 1.440$, $p = 0.161$), indicating native-like performance for these distinctions.

6 DISCUSSION

A number of phonological differences in German and English were contrasted to investigate perception performance in Saarlandian German speakers of L2 English. It was hypothesized that the /v/-/w/ distinction and initial voiced consonant cluster conditions would be most difficult and that discrimination between dental fricatives and alveolar fricatives would be less accurate than for dental fricatives and alveolar stops.

Some of these hypotheses were upheld, such as greater accuracy for the discrimination between dental fricatives and alveolar stops in contrast to dental fricatives and alveolar fricatives, confirming Hanulikova et al.'s (2012) findings. The initial voiced consonant cluster condition was similarly difficult for the nonnative group and error rates were not significantly different for this condition in contrast to the dental and alveolar frication distinction. Hanulikova et al. (2012) argued that frequency of substitution can play a role in perception. It is possible that we have the same effect here. Saarlandian German speakers of English frequently substitute /s/ for /**T**/ in English and also transfer the devoicing of initial consonant clusters, e.g., /br/ goes to /pr/. It is possible that after hearing this in their input, Saarlandian Germans have difficulty perceiving a difference between these sounds, although this requires empirical investigation, in addition to the present anecdotal evidence, to confirm the use of consonant cluster devoicing

substitutions in production. It is also possible to interpret the findings according to Best et al.'s (2001: 777) "goodness of fit" assimilation pattern. The poor performance for the dental and alveolar fricative and initial voiced consonant cluster distinction could also be due to a "goodness of fit" assimilation in which sounds such as /TA/ and /ɣɒA/ are assimilated as good enough exemplars of /sA/ and /krA/, respectively. The /v/-/w/ distinction proved most difficult for the nonnative group with error rates of around 40%, higher than any other condition. Best et al. (2001) proposed a hierarchy of difficulty in L2 phoneme recognition, arguing that performance should be worst for sounds that are assimilated into a single L1 category. On the basis of this argument, our findings indicate that poor discrimination of /v/ and /w/ is due to assimilation into a single L1 phoneme /v/, made with less frication than English /v/.

The identification of the source of perception and production problems is important for teaching of L2 phonology. Similar sounds and new sounds may require different teaching activities to draw students' awareness to the new phonological contrasts. It is also important to take into account commonly used substitutions as these may affect the perception of new phones.

REFERENCES

Best, Catherine T. 1994. The emergence of native-language phonological influences in infants: A perceptual assimilation model. In Judith C. Goodman & Howard C. Nusbaum (eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167–224). Cambridge, MA: MIT Press.

Best, Catherine T, Gerald W. McRoberts & Elizabeth Goodell. 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.

Eckert, Hartwig & William Barry. 2005. *The Phonetics and Phonology of English Pronunciation*. Trier: Wissenschaftlicher Verlag Trier.

Eimas, Peter D., Einar R. Siqueland, Peter Jusczyk & James Vigorito. 1971. Speech perception in infants. *Science* 171. 303–306.

Flege, James. 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, James. 1991. Perception and production: The relevance of phonetic input to L2 phonological learning. In Thom Heubner & Charles Ferguson (eds) *Crosscurrents in Second Language Acquisition and Linguistic Theory* (pp. 249–289). Philadelphia: John Benjamins.

Flege, James. 1995. Second-language speech learning: Theory, findings and problems. In Winifred Strange (ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 229–273). Timonium, MD; York Press.

Forster, Kenneth I. & Johnathan C. Forster. 2003. DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers* 35. 116–124.

Gibson, James J. 1966. The senses considered as perceptual systems. Boston, MA: Houghton Mifflin.

Gomez, Rebecca L. & LouAnn Gerken. 2000. Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences* 4. 178–186.

Hamann, Silke & Anke Sennema. 2005. Voiced labiodental fricatives or glides—all the same to Germans? In *Proceedings of the ISCA Workshop on Plasticity in Speech Perception* (pp. 164–167). London: University College London.

Hanulikova, Adriana & Andrea Weber. 2012. Sink positive: linguistic experience with th substitutions influences nonnative word recognition. *Attention, Perception & Psychophysics* 74. 613-629.

Iverson, Paul, Ekanayake Dulika, Silke Hamann, Anke Sennema & Bronwen G. Evans. 2008. Category and perceptual interference in second-language phoneme learning: An examination of English /w/-/v/ learning by Sinhala, German, and Dutch speakers. *Journal of Experimental Psychology: Human Perception and Performance* 34. 1305–1316.

Kingston, John. 2003. Learning foreign vowels. *Language and Speech* 46(2-3). 295-349.

Kuhl, Patricia K., Erica Stevens, Akiko Hayashi, Toshisada Deguchi, Shigeru Kiritani & Paul Iverson. 2006. Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science* 9. 13–21.

Smith, Bruce L., Rachel Hayes-Harb, Michael Bruss & Amy Harker. 2009. Production and perception of voicing and devoicing in similar German and English word pairs by native speakers of German. *Journal of Phonetics,* 37. 257-275.

Streeter, Lynn A. 1976. Language perception of 2-month old infants shows effects of both innate mechanisms and experience. *Nature* 259. 39–41.

Werker, Janet F. & Richard C. Tees. 1984. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7. 49–63.

CONTACT INFORMATION

Carrie A. Ankerstein
FR 4.3 Anglistik, Amerikanistik und Anglophone Kulturen
Universität des Saarlandes
PF 15 11 50
D-66041 Saarbrücken
c.ankerstein@mx.uni-saarland.de

Raphael Morschett
ra.morschett@freenet.de