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

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 (Halle 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 [ʃ] or retroflex [ɻ] with additional labial and pharyngeal constrictions. GE /l/ is non-velarized [l], making it phonetically similar though not identical to AE /ɻ/, which is typically velarized [ɻ].

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 /ɻ/ is phonetically different from AE /ɻ/, and because GE /ɻ/ is similar but not identical to AE /ɻ/. 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 /ʁ/ is realized differently from AE /ɾ/, and because AE /w/ is either assimilated to GE /v/, which differs from the /w/ endpoint, or heard as an unclassified 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. 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 /ɾ/-/l/ is that the GE listeners will perceive AE [ɻ] as quite different from their native [ʁ] and thus classify it (eventually) as a new phone, and that they will classify the similar AE [l] as being equivalent to their L1 [l] in spite of their phonetic differences, and will therefore perceive the /ɾ/-/l/ continuum categorically. However, the GE listeners will not perform as well as AE listeners especially on the [ɻ] endpoint because of the large phonetic distance between AE [ɻ] 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 /ɾ/-/l/ like L1 AE listeners, and they may perceive AE /w/-/ɾ/ and /w/-/ɻ/ 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/-/ɻ/, though not equal to AE, and less accurate for /w/-/ɾ/ and /ɾ/-/l/.

The experiments reported below test these predictions. We compare the results of L1 GE listeners’ identification and discrimination of stimuli from AE /ɾ/-/l/, /w/-/ɾ/, and /w/-/ɻ/ 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 /ɾ/ as [ɾ] (i.e., none of the participants realized GE /ɾ/ as the apical trill [ɾ] or came from an area where /ɾ/ is commonly realized as [ɾ]).
2.2. Stimulus Materials

We used the three 10-step synthetic continua of AE approximant contrasts - /r̚/ɑk/-/l̚/ɑk/, /w̚/ɑk/-/j̚/ɑk/, and /w̚/ɑk/-/r̚/ɑk/ - 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. /r̚/-/l̚/

Figure 1 compares the identification and the discrimination function for the /r̚/-/l̚/ 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 /l̚/ end (Figure 1, left).

ANOVARs 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, $F(6,150) = 3.21, p$
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, Fs (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 [ɻ] and GE [ʁ], whereas the difference between AE [ɻ] 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 /ʁ/.

None of the predictions correctly anticipated the GE listeners’ identification of stimuli from the /w/-/ʃ/ 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/-/ʃ/. 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/-/ʃ/ contrast, in which lip rounding is also employed contrastively.

5. REFERENCES


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.