Learning Non-Native Vowel Categories

John Kingston, Christine Bartels, José Benki, Deanna Moore, Jeremy Rice, and Rachel Thorburn

Linguistics Department, University of Massachusetts, Amherst

Neil A. Macmillan

Psychology Department, Brooklyn College, CUNY

Abstract

Two hypotheses have recently been put forward to account for listeners' ability to distinguish and learn contrasts between speech sounds in foreign languages. Best's (1994) perceptual assimilation hypothesis predicts that the ease with which a listener can tell one non-native phoneme from another varies directly with the extent to which these sounds assimilate to different native phonemes. Pisoni et al. (1994) have argued that training listeners to identify non-native phonemes teaches them sets of exemplars rather than leading to the abstraction of more general prototypes. We report here the results of four experiments examining how American English listeners learn to perceive the contrasts among the front rounded vowels of German. Their results suggest that listeners' responses are a function of the phonetic dissimilarity of the vowels themselves rather than their assimilability to American English vowels, a result incompatible with the strong phonological interpretation of Best's hypothesis, but compatible with the weaker category recognition interpretation. These results also show strong speaker effects, and are thus compatible with Pisoni et al.'s exemplars-not-prototypes interpretation of non-native category learning.

1 Introduction

In this paper, we report the results of four experiments examining the perception of non-native vowels by American English listeners, and explore issues arising from previous work on foreign speech sound perception and learning:

1. Best's (1994) perceptual assimilation hypothesis predicts that listeners' success at identifying or discriminating a foreign contrast varies directly with the extent to which the members of this contrast assimilate to contrasting native phonemes.

2. In their studies of Japanese listeners' learning of the American English /r:l/ contrast, Pisoni et al. (1994) found: (a) substantial speaker and context effects, (b) large individual differences between listeners, and (c) weak generalization to novel stimuli. These results suggest that listeners learn sets of exemplars rather than abstracting prototypes.

3. Best & Strange (1992) have shown that Japanese listeners' success in distinguishing the contrasts among the American English non-nasal sonorants /w,j,r,l/ varies directly with probable differences in the phonetic similarity of these sounds to Japanese phonemes. Flege (1990) also argues for effects of phonetic similarity between foreign and native phonemes.

The size of phonetic differences between the foreign sounds themselves may also influence their discriminability. This possibility is likely for the type of foreign sounds examined in our experiments, front rounded vowels. These sounds occur in no more 6% of the UPSID sample (Maddieson & Precoda, 1992), and this low frequency may arise because in violating the usual redundancy between the features [back] and [round], these vowels are less distinct from other vowels.

2 Methods

All four experiments adapted a method introduced by Logan et al. (1991, et seq.) in assessing the learning of the American English /r:l/ contrast by Japanese listeners. Our experiments examined the learning of contrasts among the German front rounded vowels by American English listeners who had never heard such vowels before. The experiments all have three phases:

1. In "pre-training," listeners are assessed for their ability to identify and categorize the stimuli of interest.

2. Pre-training is followed by multiple days of training on identification.

3. Finally, in "post-training & generalization," listeners are presented with new tokens of vowels they have been trained on, for identification and categorization.

The stimuli in our experiments were:

1. German words and pseudo-words of the form CVC̚,

2. Produced by 5 native German speakers, 3 female and 2 male, referred to as A-E,

3. In which the preceding C could be any of /b,d,g,p,t,k/, and the following C could be any of /p,t,k/.

The vowel was one of the four front rounded German vowels, which contrast for the features [tense] and [high], as shown in Table 1:

German vowels	+tense (long)	-tense (short)
+ high	y	ʏ
- high	ø	œ

Table 1: German front rounded vowels, showing contrasts for [tense] and [high].

Both front unrounded and back rounded vowels contrast for tenseness and height in English, as shown in Table 2 -- all these vowels also occur in German.

English vowels	+ tense	- tense
+ high	i, u	ɪ, ʊ
- high	e, o	ɛ, ɔ

Table 2. English front unrounded and back rounded vowels, contrasting for [tense] and [high].

Thus, both the tenseness and height contrasts between the German front rounded vowels are "two-category" contrasts in Best's terms, i.e. each member of each contrast can assimilate to a distinct phoneme in English. A strong version of this hypothesis predicts that each instance of both contrasts should assimilate as easily as the other.

In all three phases of the experiment, listeners performed complete identification of the 2x2 set of front rounded vowels defined by the features [tense] and [high]. In pre- and post-training, listeners also had to categorize the four vowels with respect to one distinctive feature or the other.

Four groups of four listeners each were tested under different conditions:

1. Mixed Speaker ABC listeners heard speakers A, B, and C together in training, and a representative 2/3s of the consonantal contexts.

2. Mixed Speaker BDE listeners heard speakers B, D, and E together in training, and a different but still representative 2/3s of the consonantal contexts.

3. Fixed Speaker listeners heard blocks of trials in which the speaker was fixed in alternation with blocks in which the speaker varied, across speakers A, B, and C. Stimuli varied across all 18 possible consonantal contexts.

4. Fixed Context listeners heard blocks of trials in which the initial consonant was fixed to one of /b,g,t/ in alternation with blocks in which the initial consonant was varied across these three consonants. Stimuli varied across all 5 speakers and 3 following consonants.

In all 4 experiments, the reserved speakers or contexts were presented separately, and mixed with training stimuli, in post-training generalization.

Note that listeners in the Mixed Speaker ABC and the Fixed Speaker and Context experiments heard speakers A, B, and C in training.

For all four groups, pre-training lasted one day followed by 4 days of training, and then a day of post-training generalization. All four members of a group were run together. They sat is semi-isolation in a sound-treated room and heard the stimuli binaurally over TDH-49 headphones. Responses were given by pressing buttons on a button box. In categorization, listeners made a confidence judgment, on a 1-4 scale, after making their response. In all phases of the experiment, immediately after the slowest subject responded on a particular trial, a feedback light would go on over the button corresponding to the correct answer. A block of trials began with 16 practice trials in which the stimuli alternated systematically among the responses listeners were to give during that run. 96 randomized test trials followed. Performance is assessed here with the perceptual distance measure *d'*, calculated using the constant ratio rule (Macmillan & Creelman, 1991) for all pairs of vowels in the identification tasks, and from ROC curves for the categorization tasks.

3 Results

Straight lines were fit to successive training blocks in each experiment to assess the rate of learning; the slopes of these lines vary across conditions, but all are modestly positive, ranging from 0.04 to 0.10 *d'* units per training block.

Figure 1 breaks down training and generalization performance on complete identification for the two Mixed Speaker experiments, the BDE experiment at the top and the ABC experiment at the bottom. The bottom two panels in each figure display mean performance across listeners in the first three and last three training blocks; the top four panels display mean performance in post training generalization tasks: the first three of these panels show performance when new stimuli were mixed together with the old stimuli used in training, for old+new stimuli together and old vs new separately; and the fourth panel, at the top of each figure, shows performance when new stimuli were presented by themselves. Different plotting figures show *d'* values for the six possible vowel contrasts: squares for tenseness contrasts, circles for height contrasts, and diamonds for the correlated contrasts; open plotting figures represent instances of these contrasts in which the mid lax vowel /œ/ participates, closed plotting figures comparable contrasts not involving this vowel.

Performance improves across training for all vowel contrasts. Listeners also generalize training to identifying new stimuli, performing on average at least as well as on the last training block. Finally, all contrasts involving the mid lax vowel /œ/ (open figures) were uniformly easier than comparable contrasts (filled figures of the same shape) in all phases of both experiments, although this difference is more pronounced for the ABC than BDE listeners.

Figure 2 displays performance in the Fixed Speaker experiment (at the top) and Fixed Context experiment (at the bottom). At the bottom of each figure, performance on fixed speaker or context training blocks is distinguished by the letter "F" from performance on mixed speaker or context training blocks, indicated by the letter "M". Performance is poorer on the more variable mixed than fixed blocks in both experiments. Generalization occurred in these two experiments, too. Finally, in both these experiments, contrasts involving the mid lax vowel /œ/ are always easier than comparable contrasts.

Marascuilo's (1970) method was used to test the significance of differences in *d'* values between contrasts involving the mid lax vowel /œ/ vs comparable contrasts, as laid out in Table 3.

[tense]		[high]		correlated
+ high	y : ʏ	+ tense	y : ø	ø : ʏ
- high	ø : œ	- tense	ʏ : œ	œ : y

Table 3. Contrasts compared in assessing the distnguishability of the mid lax vowel /œ/.

In exactly half of the comparisons (177/354) in the Mixed Speaker ABC and Fixed Speaker and Context experiments, contrasts involving mid lax /œ/ (in the bottom half of Table 3) were significantly easier than those that did not, and of the half that were not significant, only 9 reversed the trend in favor of contrasts involving this vowel, none significantly. The trend in the Mixed Speaker BDE experiment is similar, but not nearly so strong: only 7 of 75 comparisons significantly favored contrasts involving the mid lax vowel /œ/, and there were 16 non-significant reversals of this trend.

The Mixed Speaker ABC and Fixed Speaker and Context results show that listeners perform very differently on one instance of a particular phonological contrast than on another. The very noticeable attenuation of this difference in the Mixed Speaker BDE experiment shows that the difference comes from who said the vowels in training, for in all experiments but this one, the training stimuli included speakers A, B, and C.

Figure 3 shows performance in the two Mixed Speaker experiments on the tasks in which listeners had to categorize the four vowels into two sets of two each for tenseness or height contrasts. Pre-training categorization is displayed at the bottom and post-training generalization at the top, with the latter broken down between mixed old+new blocks and new only blocks. Tenseness results are displayed at the top and height results at the bottom. Individual listeners' performance is indicated by open plotting figures; mean performance across listeners by an "X". Individual listeners clearly differ in how much they generalize, and generalization is slightly greater for height than tenseness differences. Figure 4 compares the Fixed Speaker and Context experiments. Again, listeners differed in the extent to which training led to generalization on categorization, and only the tenseness contrast (at the top) in the Fixed Speaker experiment shows a (modest) overall improvement.

4 Discussion

In summary:

1. Training improves performance on identification and leads to generalization to novel stimuli on this task.

2. But only some listeners generalize complete identification training to tenseness or height categorization.

3. Contrasts involving the mid lax vowel /œ/ were easier in identification than comparable contrasts not involving this vowel, especially if training involved stimuli produced by speakers A, B, and C.

The last result disconfirms the prediction of the strong version of Best's hypothesis: that all instances of the same phonological contrast should assimilate equally well.

That performance should depend so much on who said the vowels in training prompted us to examine whether the different speakers' vowels differ in how physically distant they are from one another. The effects of three measures of acoustic distance on identification performance were evaluated: the Euclidean distances between center frequencies and bandwidths of the vowels' first three formants and between their durations.

When combined with a factor representing improvement over the course of training, these distance measures account for 43.8% of the identification variance in Mixed Speaker ABC experiment and 58.7% of the variance in the Mixed Speaker BDE experiment. These proportions are substantial enough to suggest that it isn't a vowel contrast's assimilability to a native contrast that predicts its distinguishability, but rather how physically different its members are from one another. This interpretation is compatible with Best's (1994) proposal that 10-12 month-old infants only approximate adult categories because the infants recognize only some but not all physcial differences between contrasting sounds. Our adult American English listeners may similarly approximate the German vowel categories by responding to the physical dissimilarity among the German vowels themselves rather than by gauging the physical similarity between the German and American English vowels.

5 References

Best, C.T. "The emergence of native-language phonological influences in infants: A perceptual assimilation hypothesis," J.C. Goodman & H.C. Nusbaum (eds.) *The Development of Speech Perception*, Cambridge, MA: MIT Press, 167-224, 1994.

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

Flege, J.E. "Perception and production: The relevance of phonetic input to L2 phonological learning," C. Ferguson & T Huebner (eds.) *Crosscurrents in Second Language Acquisition and Linguistic Theories*, Philadelphia, PA: John Benjamins, 1990.

Logan, J.S., Lively, S.E., and Pisoni, D.B. "Training Japanese listeners to identify /r/ and /l/: A first report," *J. Acoust. Soc. Am.* 89: 874-886, 1991.

Macmillan, N.A. and Creelman, C.D. *Detection Theory: a User's Guide*, Cambridge, UK: Cambridge University Press, 1991.

Maddieson, I. and Precoda, K. "UCLA phonological segment inventory database," Los Angeles, CA: UCLA Phonetics Laboratory, 1992.

Marascuilo, L.A. "Extensions of the significance test for one-parameter signal detection hypotheses," *Psychometrika* 35: 237-243, 1970.

Pisoni, D.B., Lively, S.E., and Logan, J.S. " Perceptual learning of nonnative speech contrasts: Implications for theories of speech perception," J.C. Goodman & H.C. Nusbaum (eds.) *The Development of Speech Perception*, Cambridge, MA: MIT Press, 121-166, 1994.

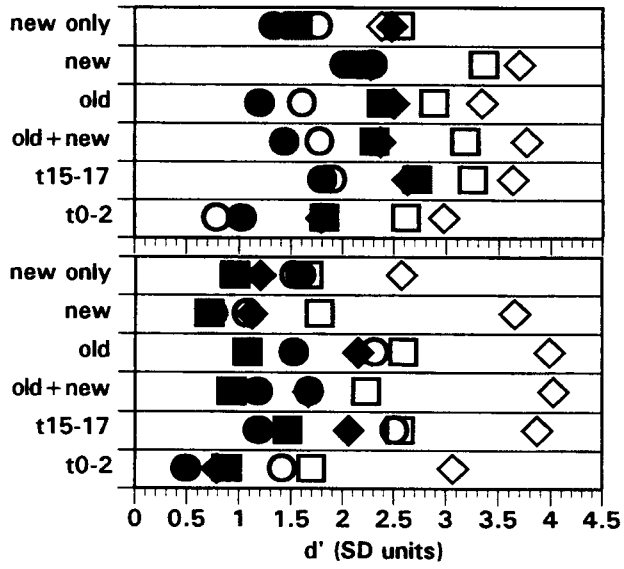

Figure 1. Mean d' values in Mixed Speaker BDE (top) and ABC (bottom) training and post-training identification: squares for [tense] contrasts, circles for [high] contrasts, diamonds for correlated contrasts; open figures for mid lax /œ/ contrasts (see Table 3).

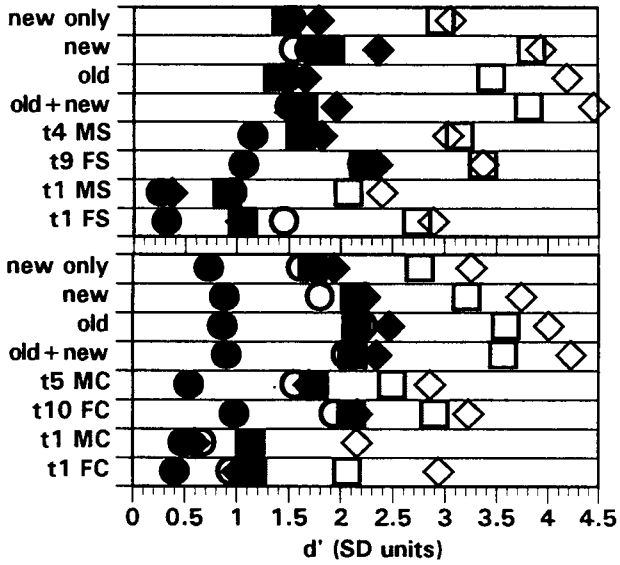

Figure 2. As in Fig. 1 for Fixed Speaker (top) and Fixed Context (bottom) identifcation; "F" vs "M" indicates fixed vs mixed training blocks.

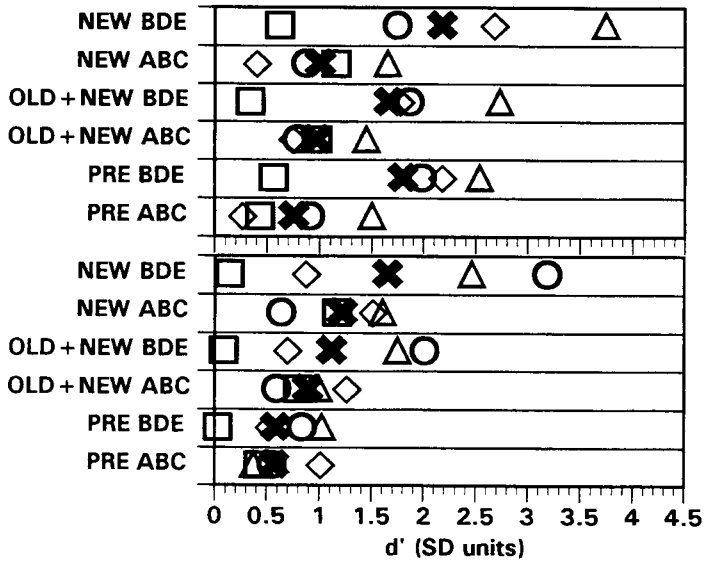

Figure 3. Individual (open figures) and mean (X) categorization on [tense] (top) vs [high] (bottom) contrasts by Mixed Speaker BDE and ABC listeners in pre- and post-training (old+new vs new only).

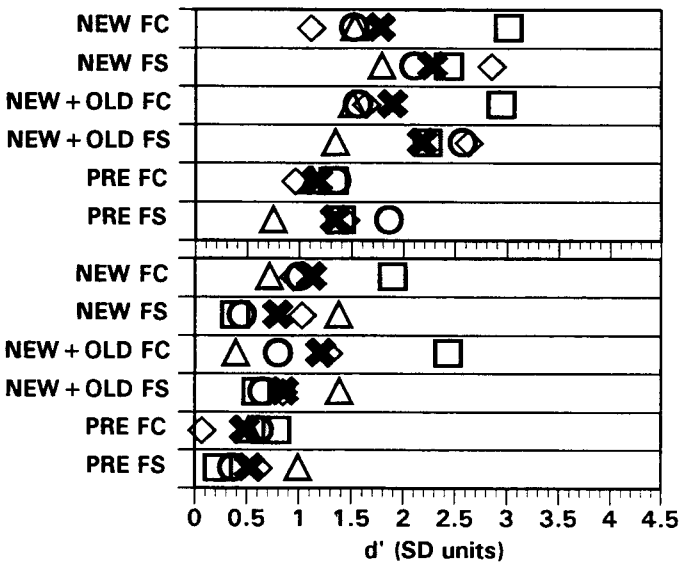

Figure 4. As in Fig. 3 for Fixed Speaker and Context listeners.