LARYNGEAL FEATURE STRUCTURE IN 1st AND 2nd LANGUAGE SPEECH PERCEPTION

Noah H. Silbert & Kenneth J. de Jong

Department of Linguistics, Department of Cognitive Science, Indiana University
nosilber@indiana.edu, kdejong@indiana.edu

ABSTRACT

This study reports an analysis of confusion data in Cutler, et al. [2] designed to probe interactions between distinctive features in English consonant identification by English and Dutch native listeners. While both listener groups exhibit extensive interaction between features, the Dutch listeners' interactions deviate systematically from the English listeners'. In the original analysis, coda voicing neutralization in Dutch was invoked to account for the lower identification accuracy and information transmission rates for coda voicing contrasts in Dutch listeners [2]. The present study augments these findings, analyzing consonant pair similarity measures, finding evidence for different laryngeal feature structure in both language groups in both onset and coda positions. This is not accounted for by a general neutralization rule.

Keywords: Cross-language Perception, Perceptual Similarity, MDS, Perceptual Space

1. INTRODUCTION

A Listener's ability to identify speech sounds depend crucially both on properties of the speech signal and properties of the listener's perceptual system. Native English speakers' obstruent identification confusions show two general patterns. First, listeners are very consistent with one another. Second, phonological features interact pervasively such that perception of one contrast (e.g., voicing) often varies with levels of a second feature (e.g., place of articulation).

Such results strongly suggest that confusion patterns arise from relations between speech actions and their encoding in the signal. Because the degree to which phonetic actions impinge on the acoustic signal is variable, and because experimental speech sound identification typically occurs in noisy contexts, there are systematic patterns in the degree of noise-induced information loss in linguistically relevant aspects of the signal. Thus, different features exhibit consistent differences in information transmission rates [1, 2, 5, 10] and psychological proximity [7, 8, 9].

Investigations of cross-language speech perception can hold signal properties constant while the properties of listeners' perceptual systems vary. Reliable differences in, e.g., phoneme identification performance across listener language groups must, then, be attributed to differences in linguistic experience and perceptual expectations, not variation in the encoding of phonetic actions. This approach has proven fruitful, for example, in the study of perception of English vowels by native English and native Spanish/non-native English listeners [3].

The current study furthers this approach by employing the Similarity Choice Model (SCM) [4, 6] to obtain a bias-free measure of the perceptual similarity between English anterior obstruents for individual English and Dutch native speakers. Individual-differences multidimensional scaling (MDS) analyses show that, consistent with previous analyses of information transmission rates and identification accuracy [3], place and manner of articulation features capture much of the overall structure of the perceptual space for onset and coda obstruents for both language groups, but that voicing distinctions collapse in coda position for the Dutch native speakers.

However, analyses of relationships between similarity parameters for pairs of minimally contrastive segments indicate that laryngeal contrasts still provide structure in the similarity space indicated by consonant identifications. This is true even when the size of the voicing contrast in coda position is severely reduced in the non-native listeners.

2. CORPUS AND METHODS

Individual subject data from a previously published study [2] were submitted to two analyses. One fit individual differences MDS models to all pairwise similarities between 10 anterior English obstruents. The other examined consistency across subjects in

the relationships between similarities for a subset of minimally contrastive pairs of these consonants.

2.1. The corpus

Consonant and vowel confusion data were collected from 16 native American English and 16 native Dutch listeners identifying 15 vowels and 24 consonants in onset and coda positions. Stimuli were presented at three signal-to-noise (six-talker babble) ratios (0, 8, and 16 dB).

We restrict our analyses to confusions of the 10 anterior obstruents [p], [b], [f], [v], [θ], [ð], [t], [d], [s], and [z] in both onset and coda position. This set contains multiple minimal contrasts on voicing, place, and manner of articulation dimensions, and it enables assessment of the perceptual effects of coda laryngeal neutralization in in Dutch.

2.2. Analyses

The SCM defines the probability of response *j* given presentation of stimulus *i* as the product of *b*(*j*), the bias to respond *j*, and *s*(*i*,*j*), the similarity between stimuli *i* and *j*, normalized by the sum of all such products across all response options:

$$(1) \qquad p(j \mid i) = \frac{b(j)s(i,j)}{\sum_k b(k)s(i,k)}$$

For *n* stimuli, the SCM has *n*(*n*-1)/2 free similarity parameters and *n*-1 free bias parameters, so for the subset of 10 consonants considered here, the full SCM has 45 free similarity parameters and 9 free bias parameters. For each of the 32 subjects, maximum likelihood estimates of the similarity and bias parameters were found. Only the similarity parameters will be considered here.

2.2.1. Individual Differences MDS

In order to visualize the overall structure of the similarity data, matrices of similarity parameters served as input to individual differences MDS analyses for each prosodic context and language group. Individual differences MDS employs two sets of parameters: a set of coordinates shared by all the subjects, and individual listener weights that scale the coordinates in the group space.

2.2.2. Ordinal Tests of Feature Interaction

Similarity parameters were also subjected to a series of ordinal comparisons to test for *feature equivalence* (FE) and *feature independence* (FI). FE comparisons examine whether similarity is constant across minimal contrasts, while FI comparisons examine whether similarity between segments contrasting in one feature (e.g., voicing) varies as a function of another feature (e.g., place).

Because we are interested in the effects of linguistic experience on the general relationships that hold among listeners from a given language group, and because the ordinal relationship between any two similarity parameters may vary randomly from listener to listener, we tally the number of listeners within each language group for each prosodic context and calculate the proportion of listeners showing the same ordinal relationships between pairs of similarity parameters. We then test the difference in proportion of Dutch and English listeners that exhibit a particular ordinal relationship. The present results are a subset of a larger analysis involving 68 tests per prosodic context. Results are presented with and without Bonferroni adjustments to criteria for statistical significance.

2.3. Predictions

If the different linguistic experience of the Dutch listeners reduces sensitivity to features not used in the native language, distances in the MDS group space should reflect this. Thus, we expect voicing distances to be systematically reduced in coda position for the Dutch listeners, and FE should fail. Also, insofar as voicing is irrelevant in coda position in Dutch, it should not be affected by other features, and FI should hold.

Failures of FI can be either directly or indirectly evident. It is directly evident if, e.g., similarity between voiced and voiceless segments is consistently modulated by another feature. In addition, it is indirectly evident if, e.g., voicing modulates place and manner interactions. Finally, a feature may provide perceptual structure on which an interaction between other features may take place, e.g., interactions between place and manner may appear in parallel sets of voiced and voiceless consonants.

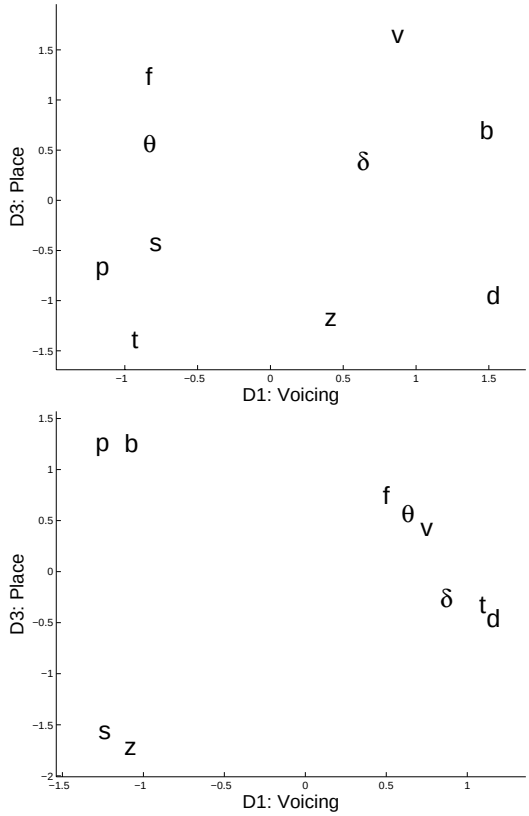
In addition, missing segments (as opposed to features) in the non-native language may cause failure of FE and FI, as well. For example, Dutch lacks dental non-sibilant fricatives, so minimally contrastive pairs involving these segments should exhibit systematically different relative similarity for Dutch listeners than for English listeners.

3. RESULTS

3.1. Individual Differences MDS analysis

Three dimensions (corresponding to voicing, place, and manner) provided the best fit for both language groups for both positions. Figure 1 plots the coda group space for the dimensions corresponding to voicing and place.

Figure 1: Coda voicing x place group spaces for English (top) and Dutch (bottom) listeners.



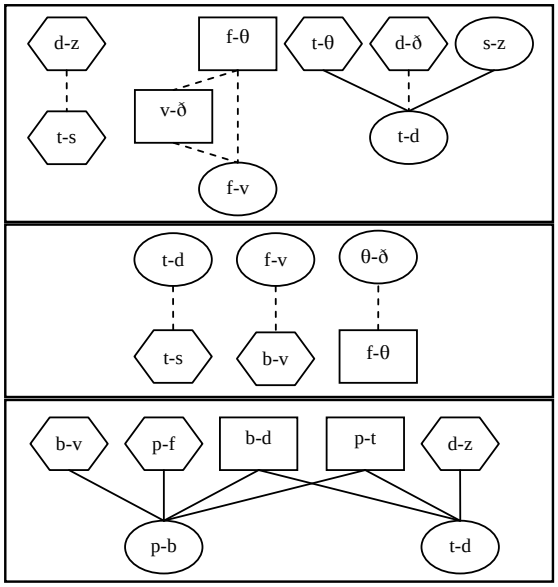
Voicing reduction is clearly evident in the Dutch data in Figure 1. Although the voiced-voiceless pairs differ systematically in the voicing dimension, distances between these segments are very small compared to the corresponding, very large distances in the English data. Evidence of a missing segment effect is also apparent; distances between dental non-sibilants and their neighbors are smaller for the Dutch listeners. Voiceless dentals group with labial fricatives, and voiced dentals group with coronal stops. Note also that place distinctions differ for the two listener groups.

3.2. Proportion Tests

The proportion of listeners exhibiting particular failures of FE and FI was largely consistent across the language groups. In 68 tests, FE and FI failed 57 (onset) and 51 (coda) times for English and 52 (onset) and 49 (coda) times for Dutch listeners, but only 17 of 68 (coda) and 14 of 68 (onset) proportion tests were statistically significant.

Figure 2 summarizes the proportion-test results for codas. The top panel depicts cases in which FE or FI fails for the English listeners but not for the Dutch listeners; the middle panel cases in which FE or FI fails for the Dutch listeners but not for the English listeners; the bottom panel cases in which FE or FI fail for the English and Dutch listeners in opposite directions. In the top two panels, more similar pairs of consonants are above and connected to less similar pairs. In the bottom panel, this relationship holds only for the English listeners. Solid lines indicate statistical significance after the Bonferroni adjustment.

Figure 2: Coda ordinal similarity relationships for English and not Dutch (top); Dutch and not English (middle); English and Dutch opposite (bottom).



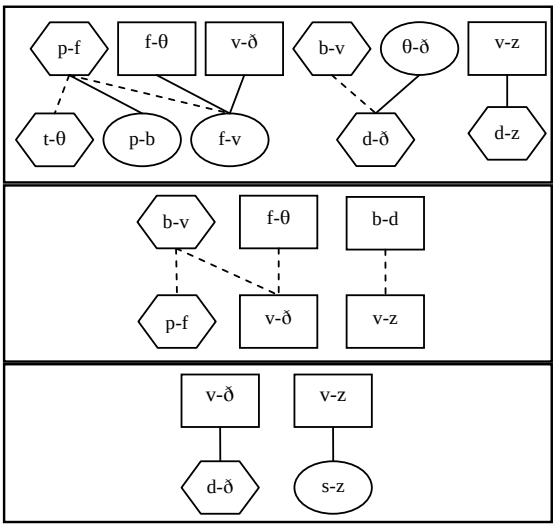
The coda proportion tests provide evidence of voicing reduction. Each voiced-voiceless pair has higher similarity than at least one other pair for the Dutch listeners, but not for the English listeners. In addition, some other relationships are modulated by laryngeal features. For example, for the Dutch listeners, *s*(f,v) is greater than *s*(b,v) but not *s*(p,f),

and *s*(θ,ð) is greater than *s*(f,θ) but not *s*(v,ð), whereas *s*(t,d) is greater than both *s*(t,s) and *s*(d,z).

There is also evidence of language specific laryngeal feature structure in onset positions, as illustrated in Figure 3. English listeners exhibit a number of interactions between voicing, place, and manner that Dutch listeners do not (top panel). Dutch listeners exhibit a different set of voicing-modulated interactions. For example, for Dutch listeners, *s*(f,θ) is greater than *s*(v,ð), but *s*(b,v) is greater than *s*(p,f). English listeners exhibit no consistent inequalities between these pairs.

In other cases, voicing modulates patterns for each language group. For example, for Dutch listeners, *s*(s,z) is greater than *s*(v,z), whereas for English listeners the opposite holds. In neither case does *s*(f,s) play a role. Similarly, both groups exhibit consistent inequalities between *s*(v,ð) and *s*(d,ð) but not *s*(t,θ) and *s*(f,θ).

Figure 3: Onset ordinal similarity relationships for English and not Dutch (top); Dutch and not English (middle); English and Dutch opposite (bottom).



There is little evidence of a missing segment effect in the proportion test results. Contrasts with dental non-sibilants are more similar than a number of other contrasts for both language groups. There are, however, unexpected segment-specific differences between the Dutch and English listeners. In onset position, *s*(b,v) and *s*(b,d) are both relatively more similar in the Dutch data than in the English data. Also the relationship of *s*(s,z) to various other contrasts differs in the Dutch and English FE and FI test results.

4. DISCUSSION

The results of both analyses are consistent with previous findings. Voicing in coda position involves low information transmission rates, low accuracy, and small perceptual distances for Dutch listeners, but not for English listeners. In addition, tests of relationships between similarities for minimally contrastive consonant pairs reveals feature structure that is unique and systematic within each language group, but not available in first-order feature analyses.

Some of this structure is likely due to properties of the speech signal itself: listeners from both language groups exhibited the same patterns of failure of FE and FI in most cases. Nonetheless, the presence of reliable language group differences involving voicing indicates that the English laryngeal contrast operates differently for native and non-native listeners, even in prosodic contexts in which it is typically found to be severely reduced or nonexistent.

5. REFERENCES

[1] Benkí, J. R. (2003). Analysis of English nonsense syllable recognition in noise. *Phonetica*, 60, 129-157.

[2] Cutler, A., Weber, A., Smits, R., Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *JASA*, 116 (6), 3668-3678.

[3] Fox, R. A., Flege, J. E., Munro, M. J. (1995). The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis. *JASA*, 97 (4), 2540-2551.

[4] Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of Mathematical Psychology* (pp. 103-189). New York: John Wiley & Sons.

[5] Miller, G. A., Nicely, P. E. (1955). An analysis of perceptual confusion among some English consonants. *JASA*, 27 (2), 338-352.

[6] Nosofsky, R. M. (1991). Stimulus bias, asymmetric similarity, and classification. *Cognitive Psychology*, 23, 94-140.

[7] Shepard, R. N. (1972). Psychological representation of speech sounds. In E. E. David, Jr., & P. B. Denes (Eds.), *Human Communication: A Unified View* (pp. 67-113). New York: McGraw-Hill.

[8] Shepard, R. N., Arabie, P. (1979). Additive clustering: Representation of similarities as combinations of discrete overlapping properties. *Psychological Review*, 86 (2), 87-123.

[9] Soli, S. D., Arabie, P., Carroll, J. D. (1986). Discrete representation of perceptual structure underlying consonant confusions. *JASA*, 79 (3), 826-837.

[10] Wang, M. D., & Bilger, R. C. (1973). Consonant confusions in noise: A study of perceptual features. *JASA*, 54 (5), 1248-1266.