Linguistic generalization in L2 consonant identification accuracy: a preliminary report

Noah Silbert, Kenneth de Jong, and Hanyong Park
Linguistic Speech Laboratory, Department of Linguistics, Indiana University

Abstract

Cross-language perception of phonetic features was investigated via an experiment in which native speakers of Korean and English identified speech sounds varying across voicing (voiced vs. voiceless), place of articulation (labial vs. coronal), and manner of articulation (stop vs. fricative) features as well as prosodic context (syllable initial vs. syllable final). Because Korean has no anterior non-sibilant fricatives and exhibits voicing and manner neutralization in syllable codas, the identification task served as a test of the effects of native language phonological patterns on generalization in the perception of non-native features. While some features (e.g., voicing) were identified fairly accurately and generalized across other features (e.g., manner of articulation), complex patterns of interaction between the experimental factors were also observed (e.g., coronal fricatives in coda position were almost all identified as voiceless, whereas labial stops in coda position were identified equally often as voiced or voiceless regardless of their actual voicing specification). These results are discussed in terms of previous approaches to second language speech perception. Limitations of the experimental protocol are discussed, and directions for future research are briefly outlined.

Acknowledgements

The authors would like to thank reviewers Manuel Díaz-Campos and Mark VanDam for their comments and critiques, as well as those who attended to presentation of this research at the 75th meeting of the Acoustical Society of America. This work was supported by grants from NSF and NIH: NSF-#BCS-9910701 & NIH-NIDCD R03 DC04095.

1.1. Two Approaches to Cross-Language Perception

It is conventional wisdom that perception of second language (L2) speech sounds depends, at least in part, on properties of the listener's native language (L1). Which properties of the L1 are relevant and how they impinge on L2 speech perception is less clear. Investigations of these issues tend to follow one of two general approaches to L2 speech perception, each of which embodies different assumptions about the units of analysis for determining the nature of second language phonology.

Experimental approaches typically take segments to be the fundamental units in second language speech perception. For example, the two predominant experimental

models in the study of second language speech, the Perceptual Assimilation Model, or PAM (see, e.g., Best, McRoberts, & Sithole, 1988), and the Speech Learning Model, or SLM (see, e.g., Flege, 1987), have profitably built on a segmental foundation. By analyzing same–different responses to assorted pairs of non-native phones, studies employing the PAM have elucidated a number of interesting facts concerning discrimination of nonnative sounds from one another. Similarly, by considering the role of individual L2 phones that don't have clear counterparts the L1, studies making use of the SLM have elucidated the relationships between nonnative and native speech sound categories and the perception and production of nonnative sounds.

Other, more traditional linguistic approaches, by contrast, focus on cross-category properties of phones in L2 speech, such as 'natural-class' phonological features and prosodic constraints. Researchers taking such an approach have profitably exploited such properties in producing explanations for differing patterns of phone substitution for different L1~L2 pairs (Brannen, 2002) and repair strategies in L2 production of structures disallowed in the L1 (Eckman & Iverson, 1994; Edge, 1991), among other issues.

Although both segmental and feature-based models have proven productive in examining certain aspects of cross-language speech (e.g., discriminability of non-native phones, accentedness of L2 speech), the relative utility of segmental and feature-based models has only occasionally been considered. The current project is explicitly designed to do just that, thereby beginning to bridge the gap between segmental models and models which focus on cross-segment properties. There are various precedents for bridging this gap. More recent descriptions of the PAM consider cross-category properties of consonants (e.g., PAM in relation to gestural phonology, Best, McRoberts, & Goodell, 2001), and the SLM relies on the phonetic features of speech sounds in determining what counts as 'new' or 'similar' (Flege, 1987; Flege, 1988). A small number of experimental studies of L2 speech perception explicitly compare segmental to non-segmental factors (Polka, 1991; Polka, 1992).

In addition, issues of cross-category generalization have arisen in the experimental literature as the result of methodological considerations. Sampling from a large proportion of the L1 and/or L2 phonological system allows researchers to test both individual segment level patterns of identification and discrimination as well as generalization of such patterns across categories (Strange et al., 1998; Strange, Akahane-Yamada, Kubo, Trent, & Nishi, 2001).

1.2. Case Study: Korean Perception of English Consonants

The present study focuses explicitly on the nature of cross-category generalization in L2 perception by examining the structure of segmental categorization involving a large number of consonantal categories. The strategy is to determine the extent to which effects found across a particular segmental pair will be generalized across other segmental pairs which are classified the same way according to featural composition or prosodic position. The phonological systems of Korean and English offer a good opportunity to do so.

On the one hand, there are a number of interesting matches and mismatches between individual segments in the two languages. Adopting the terminology of the

SLM, for Korean L1 speakers, English stops /p b t d/ are 'similar' (or 'old') phones, at least in onset position (Korean exhibits voicing neutralization in coda positions), while English anterior non-sibilant fricatives /f v θ ð/ are 'new' phones (in both onset and coda positions).

On the other hand, sampling a large number of English categories allows for fairly direct tests of generalization across segments. In particular, the densely populated consonant space of English allows for nicely 'factorial' stimulus sets sampling across voicing, place, and manner specifications. In addition, although prosodic location differs from these paradigmatic features, it offers another opportunity to examine the relationship between L1 Korean and L2 English and issues of cross-category generalization.

The current preliminary study examines the consonant perception system of Korean learners of English, particularly with an eye toward determining the extent to which consonants which share some attribute elicit the same sort of performance in those Korean's L2 identification. To do this, we presented the Korean learners with single-syllable utterances containing stops or non-sibilant consonants varying in voicing and prosodic position. To the extent that identification performance is determined by cross-segmental properties such as voicing and prosodic location, we expect the identification performance for one segment to generalize to that of other segments sharing that property. To the extent that segmental identification performance is idiosyncratic to particular segments, we have evidence that segmentally oriented models are the most appropriate for understanding the acquisition of second language perceptual phonology.

2.1. Methods

2.1.1. Subjects

Two groups of subjects were tested: 20 adult Korean L1 speakers with a mean time of residence in the U.S. of 4.95 years (standard deviation 2.78 years; range 1–10 years), and 9 native English speaking control subjects.

2.1.2. Stimuli

Experimental stimuli filled out four binary dimensions: voicing, place of articulation, manner of articulation (as indicated in Table 1), and prosodic location. Three repetitions of voiced and voiceless labial and coronal stops and (non-sibilant) fricatives in both onset and coda position in nonsense syllables containing the low, unrounded vowel /ɑ/ were produced by a male English L1 speaker, resulting in 48 test stimuli.

Table 1: Stimulus Consonants by Feature

	Labial		Coronal	
	voiced	voiceless	voiced	voiceless
stop	b	p	d	t
fricative	v	f	ð	θ

2.1.3. Procedure

Stimuli were presented auditorily. Subjects identified each stimulus in a pseudo-closed-set task. Response options for each stimulus (shown in Table 2) were circled on a paper answer sheet. Each response option was paired with a sample word containing that consonant at the top of each column of the response sheet.

Table 2: Stimulus Response Options

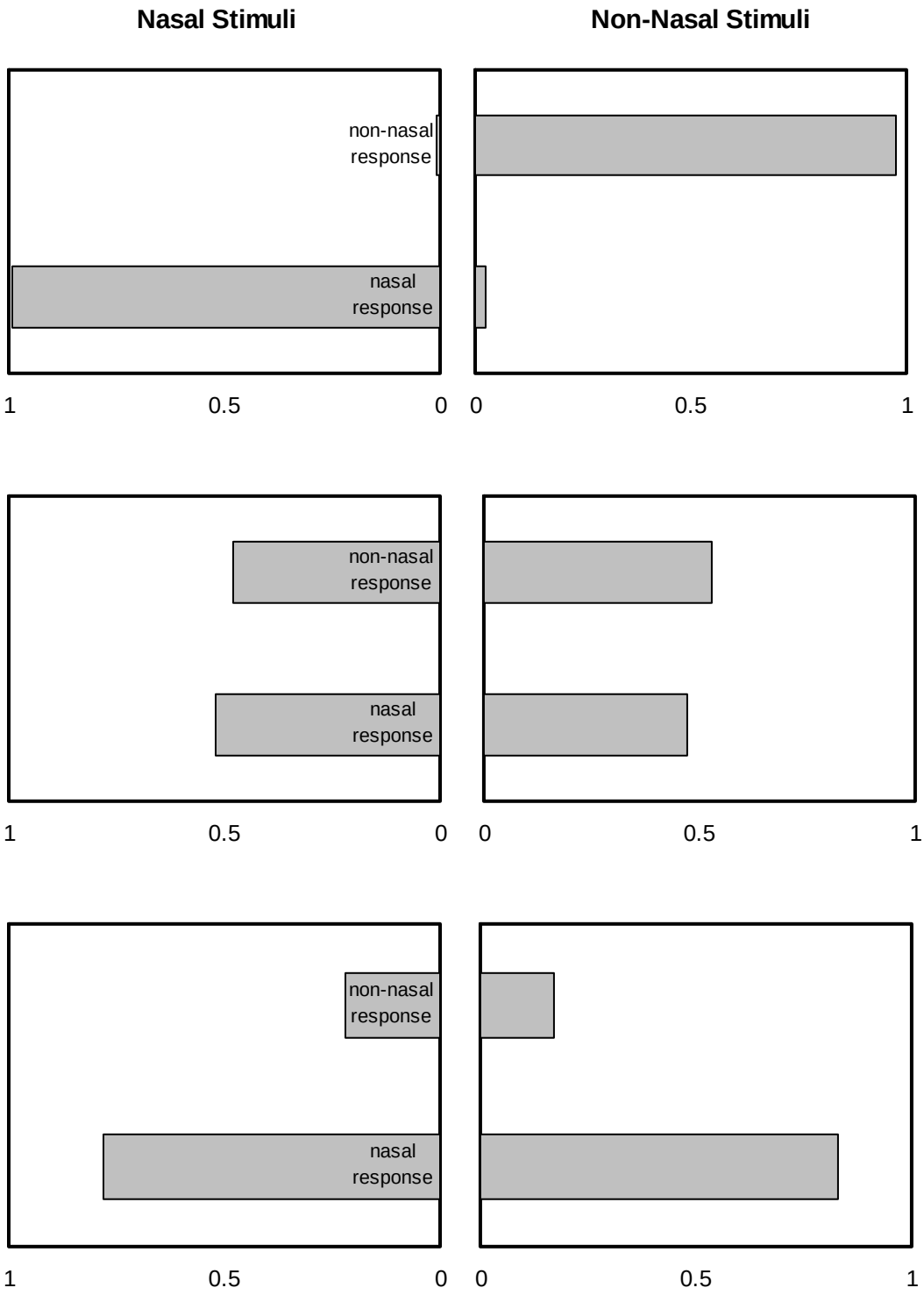
t d θ ð f v s z p b m h other__

2.2. Analysis

Of interest in the current study is the extent to which the L2 learners differentially identify each consonant in each prosodic location, with respect to other minimally distinct consonants at that location, as well as the degree of bias in the confusions. To present these results in a simple fashion, we plot matched bar graphs, indicating the proportion of times two segments were identified accurately with respect to a particular dimension of contrast. The two panels of each figure contain representations of performance according to a feature of the stimuli; the bars within each panel represent proportions (indicated along the x-axis) of responses along the feature dimension represented by each panel. Figure 1 illustrates three scenarios which might hypothetically be obtained ('nasality' is used here purely as an illustrative example). The top panel shows a case of highly accurate feature identification, in which most nasal stimuli are identified as being nasal, and most non-nasal stimuli are identified as being non-nasal. The middle panel illustrates poor identification with no bias – both nasal and non-nasal stimuli are identified as being nasal or non-nasal roughly equally often. The bottom panel illustrates poor identification with a strong bias toward identifying both nasal and non-nasal stimuli as nasal.

The features of primary interest in the present study were voicing, manner of articulation, and prosodic position. Place of articulation was treated as something of a 'replication variable,' and variation along this dimension was not expected to induce variation in feature identification. This expectation was not fulfilled, as discussed below. The English L1 control subjects consistently performed very near ceiling (i.e., always identified the stimuli accurately).

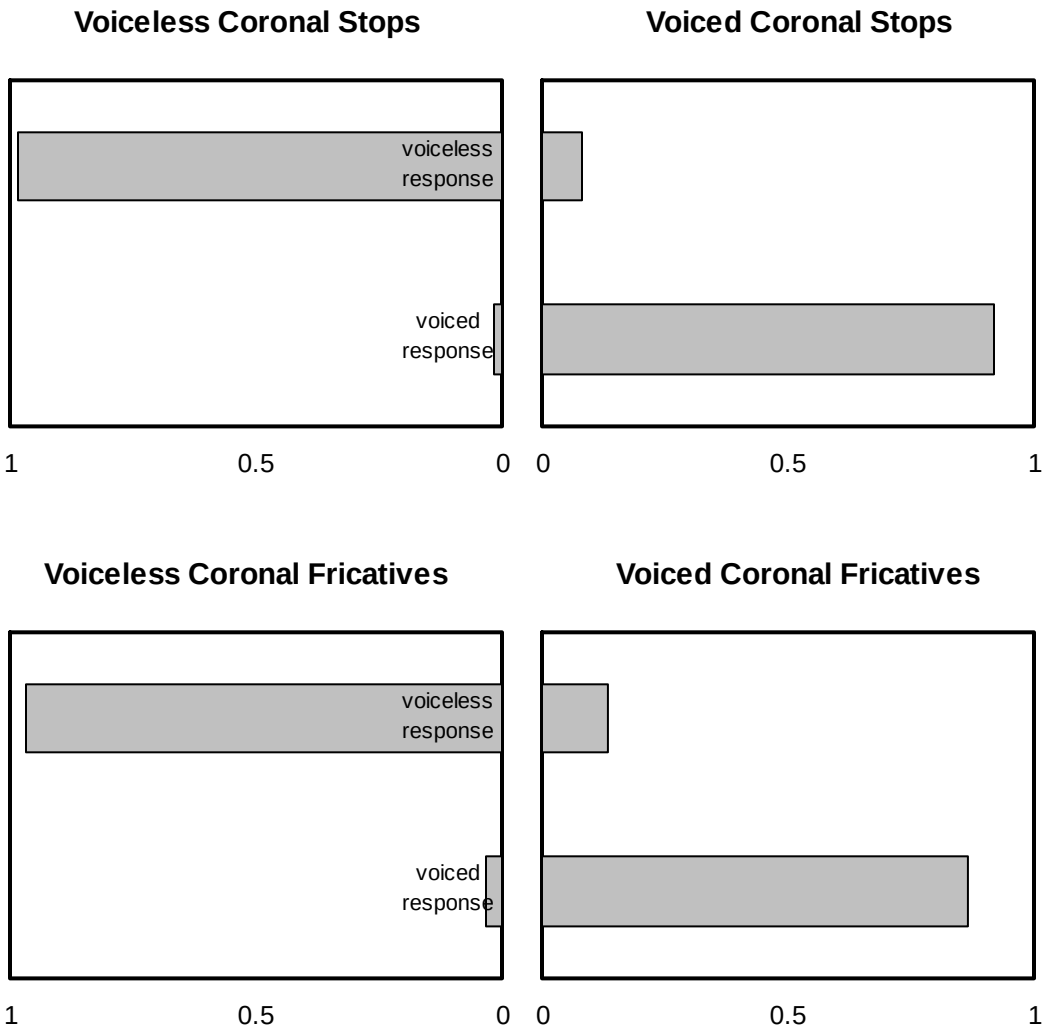
Figure 1: Possible Identification Proportion Scenarios



3.1. Results: Visual Display and Analysis

Subjects performed very well on some contrasts, particularly voicing contrasts in onset position. Figure 2 shows voicing identification for coronal stops and fricatives presented in onset position. Clearly, voicing identification for these stimuli was highly accurate. The bottom panels show that voicing identification was not only good for 'old' stop segments, but also was quite good for 'new' fricatives. In both cases, there appears to have been a small bias toward identifying the stimuli as voiceless. This bias was slightly larger for the fricatives. This pattern suggests some generalization of the good voicing performance across manner specifications, from 'old' stops to 'new' fricatives.

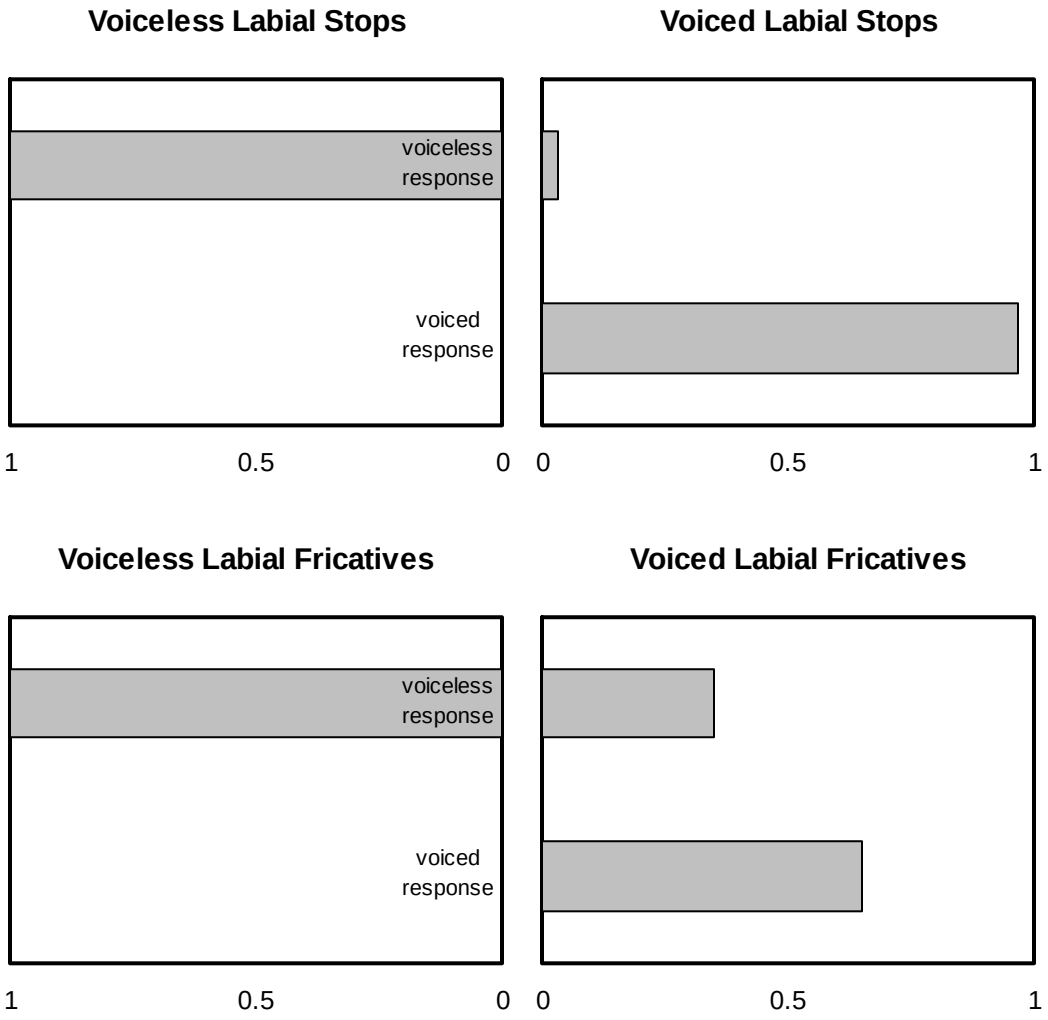
Figure 2: Voicing Identification Proportions in Onset Coronal Segments



Proportions of voiceless responses to (top left) voiceless coronal stops: 0.983; (top right) voiced coronal stops: 0.083; (bottom left) voiceless coronal fricatives: 0.967; (bottom right) voiced coronal fricatives: 0.133. Proportions of voiced responses to each stimulus type = 1 – proportion of voiceless responses.

Figure 3 shows voicing identification data for onset labial stops and fricatives. A pattern similar to that observed for coronals holds for these segments, although the difference in bias between stops and fricatives is much larger. A large number of voiced labial fricatives were identified as voiceless.

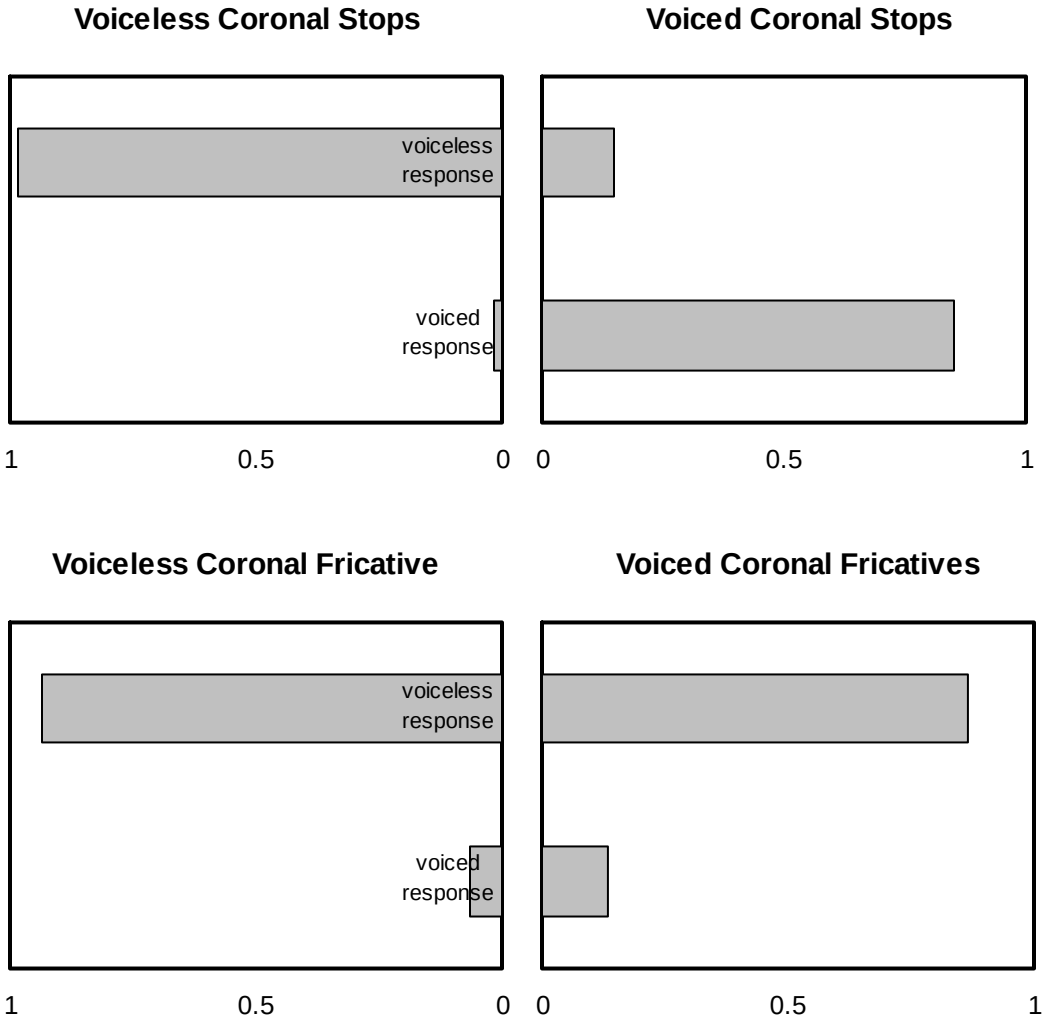
Figure 3: Voicing Identification Proportions in Onset Labial Segments



Proportions of voiceless responses to (top left) voiceless labial stops: 1.00; (top right) voiced labial stops: 0.033; (bottom left) voiceless labial fricatives: 1.00; (bottom right) voiced labial fricatives: 0.350. Proportions of voiced responses to each stimulus type = 1 – proportion of voiceless responses.

The pattern of generalization between stops and fricatives is quite a bit less clear for segments in coda position. Figure 4 shows voicing identification for coronal stops and fricatives in coda position. As with coronal stops in onset position, voicing identification of coronal stops in coda position was quite good, although there appears again to be a small bias toward voiceless identification. On the other hand, identification of voicing in coronal fricatives shows a large effect of prosodic position. The bottom panels of Figure 4 show that voicing identification was fairly inaccurate and the bias toward voicelessness was very large – both voiced and voiceless coda coronal fricatives were most often identified as voiceless.

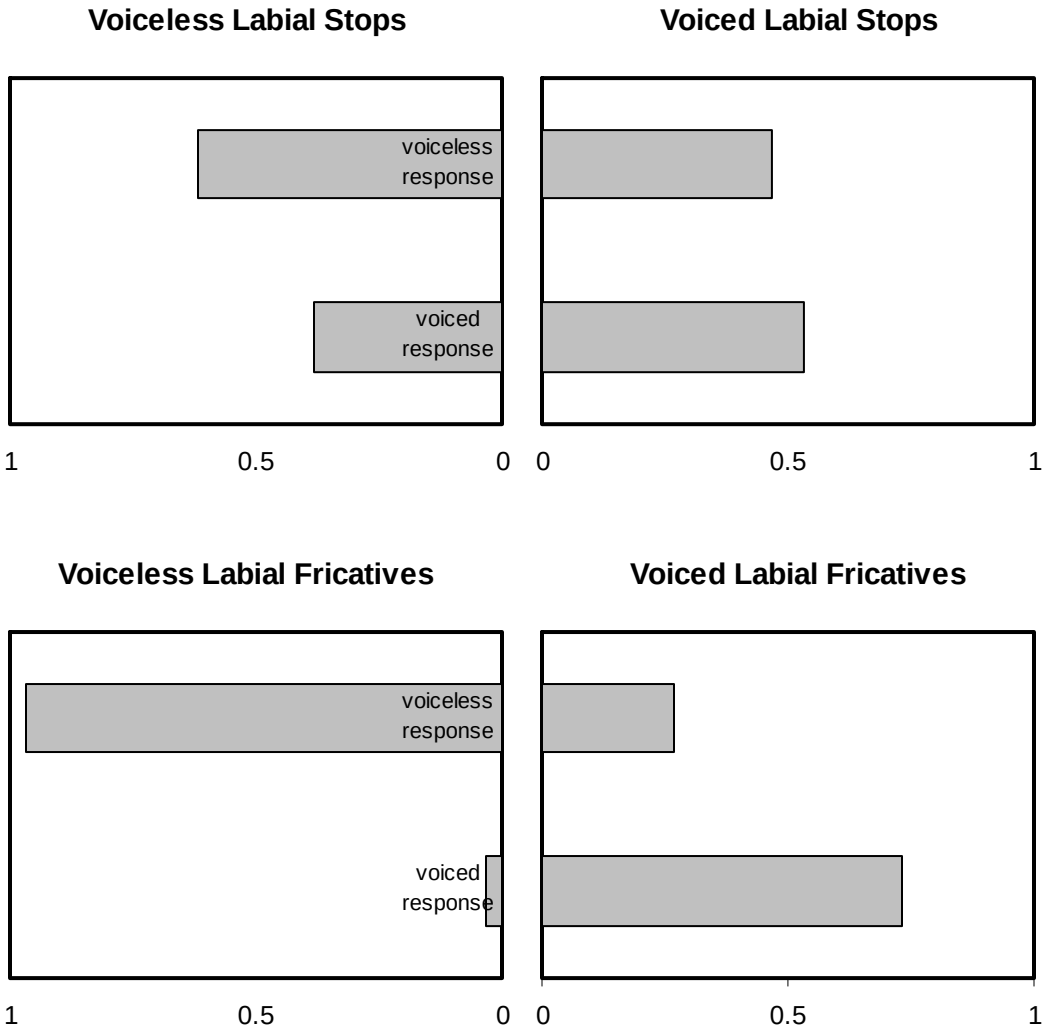
Figure 4: Voicing Identification Proportions in Coda Coronal Segments



Proportions of voiceless responses to (top left) voiceless coronal stops: 0.983; (top right) voiced coronal stops: 0.150; (bottom left) voiceless coronal fricatives: 0.933; (bottom right) voiced coronal fricatives: 0.867. Proportions of voiced responses to each stimulus type = 1 – proportion of voiceless responses.

Examining the behavior of labial consonants in coda position, we see a different pattern than we did for coronals. The top panels of Figure 5 show voicing identification proportions for labial stops in coda position, for which voicing identification was rather poor. Voicing in coda labial fricatives, on the other hand, was largely identified correctly, as shown in the bottom panels of Figure 5. For both labial stops and fricatives in coda position, there was a small bias toward voiceless identification.

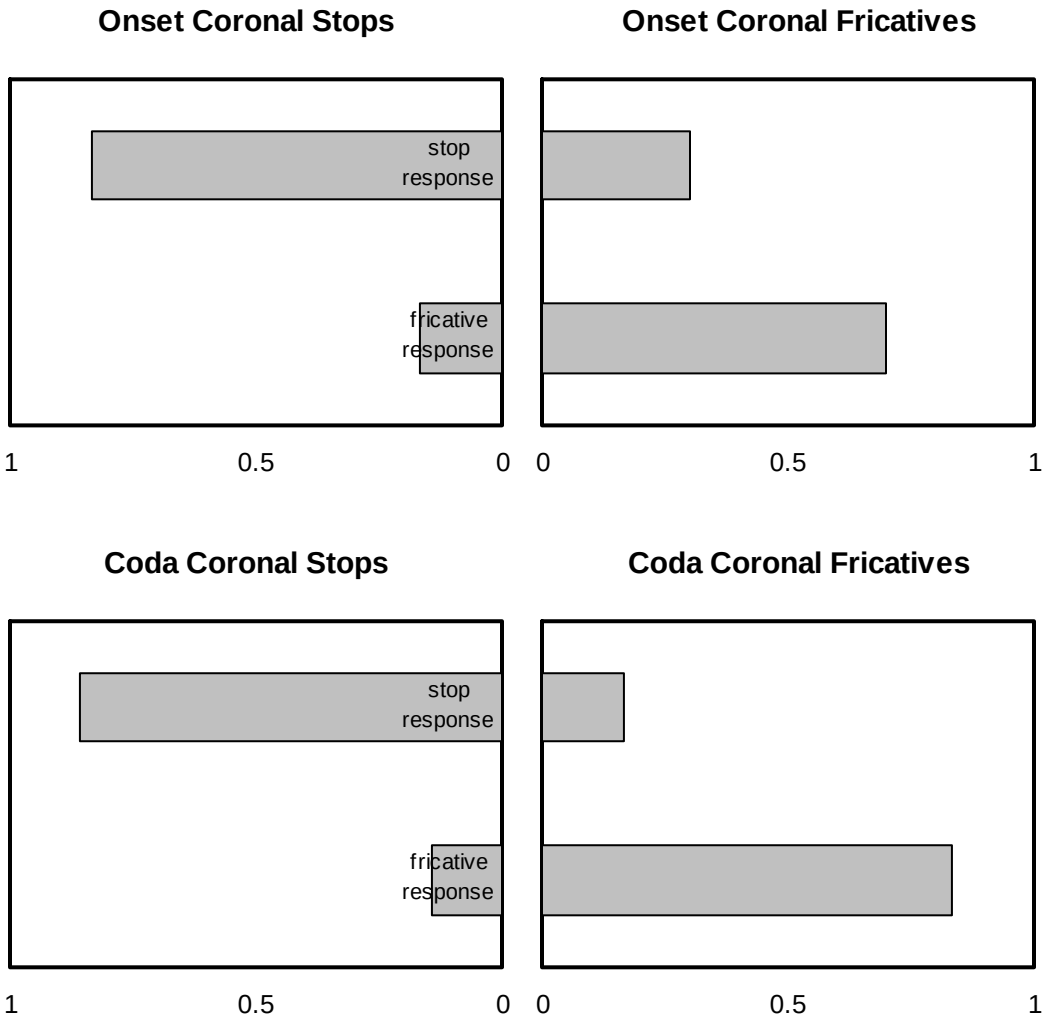
Figure 5: Voicing Identification Proportions in Coda Labial Segments



Proportions of voiceless responses to (top left) voiceless labial stops: 0.617; (top right) voiced labial stops: 0.467; (bottom left) voiceless labial fricatives: 0.967; (bottom right) voiced labial fricatives: 0.267. Proportions of voiced responses to each stimulus type = 1 – proportion of voiceless responses.

Recall that manner of articulation represents the primary difference between 'new' and 'old' phones. Figure 6 shows manner identification for coronal phones; a slight bias toward stop identification is evident, although overall accuracy was reasonably high. The bottom panels show manner identification for coronal phones in coda position; here we see both reasonably high accuracy and little discernable bias.

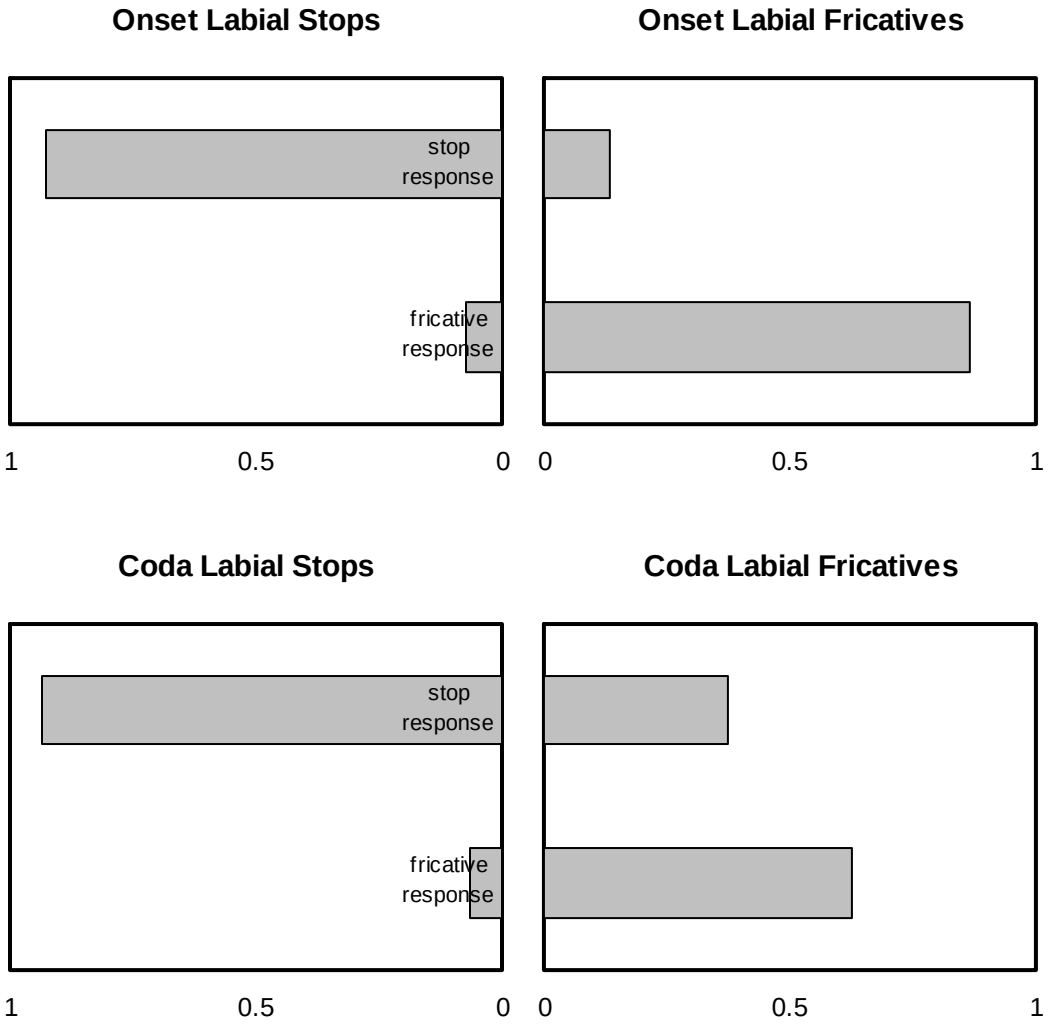
Figure 6: Manner Identification Proportions in Coronal Segments



Proportions of stop responses to (top left) onset coronal stops: 0.833; (top right) onset coronal fricatives: 0.300; (bottom left) coda coronal stops: 0.858; (bottom right) coda coronal fricatives: 0.167. Proportions of fricative responses to each stimulus type = 1 – proportion of stop responses.

The top and bottom panels of Figure 7 show manner identification proportions for labial phones in onset and coda position, respectively. Manner identification for labial phones exhibited a pattern opposite that of coronals. Manner identification in onset labials was highly accurate and a very small bias toward stop identification can be seen; manner identification in coda labials was far less accurate, and a much larger bias toward stop identification is evident.

Figure 7: Manner Identification Proportions in Labial Segments



Proportions of stop responses to (top left) onset labial stops: 0.925; (top right) onset labial fricatives: 0.133; (bottom left) coda labial stops: 0.933; (bottom right) coda labial fricatives: 0.375. Proportions of fricative responses to each stimulus type = 1 – proportion of stop responses.

3.2. Results: Statistical Analysis

To substantiate and compile the observations made above, raw frequency data was analyzed using a multi-way frequency analysis model-fitting algorithm implemented in SPSS. Multi-way frequency analysis is a higher dimensional generalization of the two-dimensional χ^2 test of independence. The algorithm begins with a saturated model that predicts cell frequencies perfectly. Parameters (i.e., terms representing, in this case, voicing, manner, place, and prosodic position) are then eliminated in a stepwise fashion; at each step, the factor contributing least to the predictive power of the model is eliminated, at which point the ability of the model to predict cell frequencies is re-tested. Parameter elimination repeats until the pruned model has statistically significantly less predictive power than the full model.

The end result of the application of this algorithm was a model with the following parameters (* indicates interaction between factors): `VOICING*MANNER*PLACE`, `VOICING*PLACE*PROSODY`, `MANNER*PLACE*PROSODY`. It is important to note that each interaction term implies the presence of any lower order term included in the interaction term, so, for example, the presence of `VOICING*MANNER*PLACE` implies the presence of `VOICING*MANNER`, `VOICING*PLACE`, and `MANNER*PLACE` as well as all the simple component factors.

The statistic used to test the predictive power of the model is the likelihood ratio G^2; in this case, $G^2 = 2.4457$, df = 2, $p = 0.294$. The high value of p (> 0.05) indicates that the simpler model is statistically indistinguishable from the full model, and hence the removed factors have no significant predictive power.

The `VOICING*MANNER*PLACE` interaction can clearly be seen by comparing Figure 2 with Figure 3 or by comparing Figure 4 with Figure 5. Comparing Figures 2 and 3, it is clear that responses to labial fricatives in onset position were strongly biased toward voiceless, whereas voicing identification for labial stops and all coronal segments in onset position was largely unbiased. Comparing Figures 4 and 5, we see that labial fricatives and coronal stops in coda position were identified fairly accurately, whereas coronal fricatives in coda position were almost all identified as voiceless and labial stops in coda position were identified as voiceless and voiced equally often.

The large difference between responses to coda coronals and labials is likely also responsible for the `VOICING*PLACE*PROSODY` interaction in the final model. Again, although voicing in both coronal and labial segments was identified fairly accurately in onset position, the patterns of identification for coronals and labials in coda position were very different. The `MANNER*PLACE*PROSODY` interaction can clearly be seen by comparing Figures 6 and 7. While both coronals and labials were identified as stops more often than as fricatives, this bias was greater for onset coronals than for coda coronals, and it was greater for coda labials than it was for onset labials. In addition, this bias toward stop identification was greatest for coda labials and least for onset labials. Lower order interactions are likewise apparent in the associated figures.

The overall picture provided by the statistical model is consistent with the overall picture provided by visual analysis of the data. While generalization of voicing from 'old' to 'new' segments is apparent in many cases, place of articulation, manner of articulation, and prosodic context interact to modulate generalization in certain cases.

4. Conclusions and Discussion

It is clear from inspection of the proportions presented above (and consideration of the statistical analysis) that features do generalize from 'old' to 'new' phones, at least in certain cases. As shown in Figures 2–5, voicing identification was very good for both coronal and labial stops in onset position, although a small bias toward voiceless responses was observed. Similarly, voicing identification was very good for coronal fricatives in onset position, and it was good, although slightly worse, for labial fricatives in onset position. A slightly larger bias toward voiceless responses was observed for these phones.

Of course, voicing identification patterns are difficult to interpret without consideration of identification of manner of articulation features. Given that voicing in stops was identified with a high degree of accuracy, if fricatives were identified as stops, it would be no surprise that voicing in fricatives was also identified accurately. Figures 6 and 7 show that manner of articulation was identified reasonably accurately for both coronal and labial consonants in onset position. Framing the onset voicing identification results in terms of featural contrast (as opposed to segmental inventories) while maintaining the terminology of the SLM, we can say that the highly accurate manner identification supports the idea that 'old' (i.e., voicing) features generalize across 'new' (i.e., manner) features.

Complicating matters is the fact that, in certain cases, prosody, place, and manner interact. Clearly, coda consonant identification patterns were rather different than those for onset consonants. Voicing identification was good for labial stops in onset position, poor for labial stops in coda position, and there was a small bias toward voiceless responses. On the other hand, voicing identification was good for coronal fricatives in onset position, poor in coda position, and there was a large bias toward voiceless responses.

The observation about generalization performance that can be garnered from the present data is that, while generalization is most clearly obtained in cases where the non-native listener is performing fairly accurately, a closer look at what happens in cases in which accuracy is not as high suggests that a property of the stimuli along one dimension may show up as biasing factors in the identification of a property along another dimension. For example, comparing Figures 4 and 5 suggests that being a fricative and being coronal both increase the likelihood that the listeners will call a segment voiceless. Both of these factors apparently contribute to the amount of consonantal noise in the signal in much the same way that being voiceless does.

With this sort of interpretation, these patterns of non-generalization of voicing identification across manner and place of articulation suggest that factoring out manner and place from the voicing judgments is not something that comes 'for free', as would be suggested by models that treat categories solely in terms of cross-segment properties such as features or prosodic constraints. Rather, independence of these factors is something which second language perceivers must develop.

It may be that segmental models (e.g., PAM, SLM) are well suited to characterizing the behavior of early acquirers of a second language and that models incorporating cross-segmental properties more accurately describe later stages of acquisition. Indeed, the research program that gave rise to the PAM was based on naïve

subjects' responses to a variety of non-native speech sounds varying in their (hypothesized) similarity to native speech sounds. Although the SLM has been successfully applied to a wider variety of populations with varying degrees of experience with a second language, including some very advanced second language acquirers, it, too, seems particularly adept at characterizing early stages of acquisition. However, because the SLM is not characterized in terms of dimensions of contrast, it is difficult to interpret SLM-based experimental findings in terms of patterns of generalization. It may be that the assumptions regarding segments and their properties that form the foundations of these models are based in part on aspects of second language acquisition particular to early-stage learners. Of course, this is very speculative. Further (cross-sectional or longitudinal) research is needed to directly test whether or not feature generalization plays an increasingly important role as second language acquisition progresses.

Also of interest is the interaction of L1 phonological processes with 'old' vs. 'new' properties of L2 speech. Recall that the anterior non-sibilant fricatives of English are 'new' segments to native speakers of Korean. However, English stops are not so clearly defined in terms of the SLM properties 'old' and 'new.' In onset position, we may posit that the English stops correspond reasonably well to native Korean consonants, so it seems reasonable to take these to be 'old' phones. In coda position, however, Korean exhibits voicing and manner neutralization.

Because Korean native speakers are accustomed to hearing a much more restricted range of speech sounds in syllable codas, it may be possible to take a subset of the English stops in coda position to be 'old' phones and take the complement to be 'new' phones. The English stops that sound more like neutralized Korean coda stops would be considered 'old,' while the other, less Korean-like stops would be considered 'new.' Although this seems intuitively correct, it does not correspond well to the original SLM conception of 'old' and 'new' phones, which is based on the segment inventories of a speaker's L1 and L2. Allophony and neutralization in either the L1 or the L2 do not typically enter into the equation, so it is difficult to determine which English stops should serve as 'old' and which as 'new' in the present case. It is also difficult to derive straightforward predictions of the effects of this kind of 'old' or 'new' status of coda stops on identification performance. The voicing identification performance for onset (Figures 2 and 3) vs. coda (Figures 4 and 5) consonants is suggestive, but not conclusive. Given the difficulties in determining the effect of coda neutralization in Korean on identification of English consonants, we can say only that it is not surprising that most of the interactions of voicing, place, manner, and prosodic position we found are due to the large differences found in the responses to coda, as opposed to onset, consonants.

Finally, it is important to note that our interpretation of both the relatively straightforward 'onset' data that support that idea of generalization and the complicated patterns of interaction found for some coda consonants must be tempered by an important limitation of this study: the stimuli were all generated by a single speaker. It is impossible to say unequivocally that the results were not due to idiosyncratic features of this individual's speech. Although it helps that the native speaker control subjects performed very near ceiling, seemingly insignificant idiosyncratic variations in the stimuli may well have been ignored by the L1 controls and magnified by the L2 subjects. For this reason, the present experimental paradigm is currently being extended and refined with multi-talker stimuli.

5. 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*(2), 775-794.

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of Perceptual Reorganization for Nonnative Speech Contrasts: Zulu Click Discrimination by English-Speaking Adults and Infants. *Journal of Experimental Psychology: Human Perception and Performance, 14*(3), 345-360.

Brannen, K. (2002). The role of perception in differential substitution. *Canadian Journal of Linguistics, 47*(1/2), 1-46.

Eckman, F., & Iverson, G. (1994). Pronunciation difficulties in ESL: Coda consonants in English interlanguage. M. Yavas (Ed.) *First and Second Language Phonology*. (pp. 251-265). San Diego, CA.: Singular.

Edge, B. A. (1991). The production of word-final voiced obstruents in English by L1 speakers of Japanese and Cantonese. *Studies in Second Language Acquisition, 13*(3), 377-393.

Flege, J. E. (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. E. (1988). The development of skill in producing word-final English stops: Kinematic parameters. *Journal of the Acoustical Society of America, 84*(5), 1639-1652.

Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America, 89*(6), 2961-2977.

Polka, L. (1992). Characterizing the influence of native language experience on adult speech perception. *Perception & Psychophysics, 52*(1), 37-52.

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*(4), 1691-1704.

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