On Segmental Factorability in Second Language Learning.

Kenneth de Jong & Noah Silbert

Linguistics and Cognitive Science

Indiana University

Indiana University Working Papers in Linguistics – Online, 06-2

Abstract.

This paper distinguishes between two models of learning a second language phonological classification system, segment-by-segment learning, which starts with individual elements and generalizes to broader featural classification skills later on, and featural classification, which starts with featural classification schemes first and fills in specific details later on. Evidence reported here comes in two forms. First, subjects who are more accurate in distinguishing one pair of segments are also more accurate with other pairs contrasting the same feature. Second, identification patterns for unfamiliar ('new') segments approximate those expected for featural independence, as estimated using J-factor analysis. The tentative conclusion is that second language learners create generalized classification schemes first and develop more detailed and robust classification skills with increased experience

1. Introduction.

Previous research in second language acquisition has relied on two different sorts of models of how a learner develops a second language perceptual system that is able to identify phonological segments in a second language. First, in some models, second language learning is determined by general properties of the two languages. Many models rely on general linguistic facts about a language's phonological system, such as prosodic position or the presence of a featural contrast, to predict the specific abilities of second language acquirers. Such models predominate the linguistically oriented literature, so examples of these sorts of models are easy to find. For example, Brennan (2002) examines differential substitutions of the English non-sibilant fricatives in terms of the types of features which contrast in a learner's L1. The model tested there attempts to predict the treatment of individual segments on the basis of properties which generalize across the segments in L1. There is much other research exemplifying this type of approach is common (e.g, Broselow, *et al*., 1987; Broslow and Finer, 1991; Carlisle, 1988; Eckman, 1991; Eckman & Iverson, 1994, to name but a few).

In other models, second language learning is determined by the specifics of the learner's experience with the second language. In most models used in the experimental and quantitative literature, the units of perception are not general properties abstracted across the second language system, but are the peculiarities of allophonic variants of each segment. One very well researched model of this type is the Speech Learning Model developed by Flege and colleagues (Flege, 1987; 1988; 1995) in which linguistic items are treated on an allophone-by-allophone basis. Other recent studies have suggested even specificity in learning down to the level of individual experience (Bradlow et al, 1997).

This dichotomous nature of the second language literature parallels the more general literature in identifying segments in a first language. On one hand, Nearey (2003), building on previous work by Boothroyd & Nittrouer (1988) on speech segmentation, suggests a model in which segment identification relies on independent classification of a sound on the basis of general features distinguishing it from other sounds (though he does not report positive results supporting the model). Other models of segment identification rely on maps of individual tokens labeled with very specific information, such that each segment has its own independent signature.

There are few studies that can address how these two types of models can be reconciled in second language learning. This is because studies that focus on general properties tend to sample the learners' performance using diffuse experimental techniques, such as impressionistic transcription, and there is not sufficient gradient information about each of the listeners' performance on any one segment to quantify the generality of their behavior. On the other hand, studies that examine more specific models have much more quantified measures of the listeners' performance, but tend to focus on small numbers of segments. In order to reconcile these two approaches, studies which sample a larger number of segments with enough experimental acuity to determine the learners' performance on each segment are necessary.

One strand of previous work on second language perception and learning has sampled a large phonological system in this way. Strange *et al* (1998) sample the English vowel space in sufficient gradient detail to determine the general performance of learners' perception of each vowel. They have been developing corpora of perceptual identification and goodness judgments from learners of English from a variety of backgrounds. Strange *et al* (1998) report such data in detail for Japanese learners. Perusing the data reported there reveals relatively little in terms of parallelism between different vowels. Listener identification patterns for the front vowel system does not exhibit any obvious parallels with their treatment of the back vowel system, and the tense-lax distinction in English does not seem to be treated in any general way by the Japanese listeners.

The research reported in this paper investigates whether there is evidence for feature generality in early second language acquirers' identification of non-native consonants. Are segments learned according to classes of consonants, or do acquirers simply construct segment-specific identification skills for each consonant separately?

To this end, the current paper reports data from experiments in which native Korean listeners were asked to identify a number of consonants placed in a variety of prosodic locations. Like work by Strange and colleagues, the point of this research is to sample identification accuracy for a large enough sample of consonants to be able to examine whether the pattern for one segment is replicated with other segments which are members of the same class of segment, or whether individual segments are each treated differently. Unlike research by Strange and colleagues, this work focuses on consonant identification. Based on a variety of considerations, we might expect consonant classification to be more reliant on general classification schemes than are vowel segments. The features we examine are of two types: prosodic location, and segmental features (voicing and manner).

To distinguish a featural from a segment-specific model, we take two approaches. The first takes a segment specific model as the null hypothesis. If the identification skills for a consonant bear no relationship to the identification skills for another consonant, we expect that there will be no correlation between a listener's accuracy with one consonant and their accuracy with another segment in the same class. If, however, identification skill develops with respect to featural contrasts, we expect a listener's accuracy with one consonant to be correlated with their accuracy with another consonant that shares the same featural contrast. The presence of correlations between, for example, accuracy in perceiving voicing contrasts for coronal consonants and accuracy in perceiving voicing contrasts for labial consonants would eliminate the null-hypothesis that voicing identification is learned independently for coronal and labial consonants.

Of course, there are a number of reasons besides a featural generality to learners' identification skills that would explain a correlation between the skills for two sets of consonants. The most obvious is that some speakers are simply more experienced with the second language than are others, and so their identification skills across the board are better than are others', and so, in turn, identification accuracy for any two sets of consonants will be correlated. To test for this interpretation of correlations, we examine a variety of sets of consonants, expecting, on the basis of a featural learning model, that only sets of consonants which share a feature will exhibit correlations.

The second approach tests featural independence more directly, examining predictions for a factored model of identification taken from Nearey (2003). Nearey (2003) builds on work by Boothroyd & Nittrouer (1988) on J-factor analysis, which examines the relationship between listeners' accuracy with a complex unit (here a segment) and their accuracy with individual components of that complex unit (here the features which distinguish the segment from its neighbors). We calculate J-factors for portions of our data, and compare them to predictions from Boothroyd & Nittrouer (1988) coming from models in which linguistic identification is the combination of making independent distinctions between the components of a linguistic unit.

2. Methods.

2.1. Stimuli

Four speakers of American English who grew up and resided at the time of recording in Northern Mid-west in their late 20's were asked to produce nonsense words consisting of the vowel /a/ and a variety of consonants. The consonants examined in this study are all voiced and voiceless, coronal and labial, stops and non-sibilant fricatives, as arrayed in a 2x2x2 matrix in Table 1. Other segments were included in the perceptual protocols, but are not included in the current analyses. The consonants appeared individually either before a vowel (*Initial Condition*) or after a vowel (*Final Condition*), or between two vowels. The consonants between two vowels either had stress on the vowel preceding the consonant (making it *Post-Stress*) or on the vowel following the consonant (making it *Pre-Stress*).

Speakers were seated in a quiet room in the Indiana University Phonetics Lab and recorded onto DAT using a free-standing microphone. Recordings were digitized at 44kHz and digitally edited from their surrounding context.

Table 1. Segments Examined in Current Corpus

	Coronal		**Labial**	
	Voiced	**Voiceless**	**Voiced**	**Voiceless**
Stops	**/d/**	**/t/**	**/b/**	**/p/**
Fricatives	**/ð/**	**/θ/**	**/v/**	**/f/**

The speakers were cued orthographically to produce each of the 32 forms (8 consonants by 4 prosodic positions). The speakers were familiarized with the orthographic stimuli before recording to insure that they knew what each of the letters indicated and how the orthography cued the various prosodic locations and stress patterns, and they were monitored for accurate interpretation of each stimulus.

The eight segments examined here fall into two general classes with respect to the Korean phonological system. The stops are all *similar* to segments in Korean. Korean has labial and coronal stops, though the voicing contrast is somewhat different from that in English. While English stops in initial position exhibit a two-way contrast between aspirated and unaspirated, Korean stops exhibit a three-way contrast between 'aspirated', 'fortis', and 'lenis'. The aspirated Korean stop generally exhibits more aspiration than its English counterpart, and the 'voiced' English stop corresponds to two unaspirated Korean categories which contrast between a 'lenis' stop with a relatively soft release characteristic and low fundamental frequency on the following vowel, and a 'fortis' stop with a sharp release and high fundamental frequency in the following vowel. Though the match across the languages is not exact, the stops are close enough so that the acquisition literature generally reports very little in the way of segmental substitutions for English stops.

The non-sibilant fricatives fall into what most classification schemes would call *new* segments. Korean has no non-sibilant fricatives except for /h/, and Korean acquirers of English in most studies often substitute either stops or sibilant fricatives for them.

With respect to the prosodic positions, Korean allows all of the consonant contrasts to appear in pre-vocalic position, so the pre-vocalic stimuli are likely to be the most similar to Korean productions. Korean also allows consonants in final post-vocalic position as well, but the number of contrasts is greatly reduced by a variety of neutralization rules, which restrict productions to single voiceless stops. In addition, the rendition by Koreans of even these neutralized forms is different enough from those in English that studies of Korean learners of English find that Korean learners often put in an excrescent vowel after the coda consonant. This suggests that what English producers would call a single syllable with a final consonant is different enough from the Korean forms that it gets treated as a disyllabic form. One other detail in the Korean phonology is that lenis stops in intervocalic position are shortened and voiced throughout, much like post-stress non-coronal stops in English. Finally, Korean does not have stress, so it is difficult to evaluate how Korean listeners will treat the two intervocalic positions.

2.2. Listeners

41 listeners were recruited from the undergraduate student population at Kyonggi University, located near Seoul, Korea. Most of the listeners were in their early 20’s, and none had traveled extensively in an English speaking country. Each of them were recruited from basic level English classes. With their experience with English classes in primary and secondary school, this would mean that they have extensive experience with written English, though commensurate experience with spoken English is rare. Hence each of them would be classified as an inexperienced learner, with respect to native English productions.

2.3. Procedure

The listeners were each seated in quiet cubicles in groups of approximately 10 subjects, and presented with the stimuli over a loud-speaker. They were asked, after each presentation, to identify the consonant which was spoken from a list of 14 alternatives presented in Roman orthography. The list is given in Table 2. Along with each response alternative was a key word, which was chosen to exemplify each segment in initial position. Also, subjects were provided with the option of indicating an additional write-in response (Other:___), though the subjects very rarely used this option for the consonants we are analyzing in the current paper. A block of practice trials was run first, and the listeners were asked if they had any questions about the procedure.

Table II. Orthographic Response Alternatives.

dog	**t**ell	**th**in	**th**at	**f**all	**v**ase	**s**it	**z**ip	**p**in	**b**all	**r**ain	**l**aw	**h**all	**w**ood	**y**es
d	t	θ	ð	f	v	s	z	p	b	r	l	h	w	y

2.4. Analyses

In the current paper, generalization is examined in two ways, one that treats the segmental model as a null hypothesis, and the other which specifically examines results expected from a featural model.

In the first analysis, we examine whether the accuracy rates for one pair of segments correlate with the accuracy rates for another pair of segments that contrast in the same feature. If each segment has its own identification signature, the development of abilities for one segment pair need not parallel that of abilities for other, parallel segment pairs. However, in a featural model, all segments that share a featural contrast should develop in parallel. To test this, we regressed each listener's accuracy in a pair of segments against that listener's accuracy with other pairs of segments that contrast in the same feature. Generalized models predict that listeners who are better at a contrast in one pair of segments (e.g. the voicing contrast between /p/ and /b/) will also be better at another parallel pair (e.g. the voicing contrast between /t/ and /d/).

The second analysis follows from analyses developed by Boothroyd & Nitrouer (1988). They reason that, if the contrasts in a complex of stimuli are decomposable into separate, independent contrasts which each operate without regard for one another, the probability of identifying a complex stimulus should be the product of the probabilities of perceiving each of the separate contrasts. So, in Boothroyd & Nitrouer's research, they examined perceptual accuracy in CVC syllables consisting of three segments, and examined the relationship between overall accuracy for a syllable and accuracy in identifying each segment. They expected that the overall accuracy would be the product of the accuracy of each of the segments, and hence, this model would predict that the syllable-level accuracy should be the cube of the average segmental accuracy.

To evaluate the relationship between syllable and segment accuracy, Boothroyd & Nitrouer (1988) took observed syllable accuracy and segment accuracy, fit them into the modeling equation: syllable accuracy = (average segment accuracy)^J , and solved for J, expecting a value of roughly $J = 3$. Nearey (2003) reports the results of a variety of computational simulations of various degrees of segmental independence and obtained values of nearly 3 for independent 3-segment stimuli. As the independent segments begin to artificially interact in his simulations, the value of J tends to decrease to values below 2.

We adapt the J-factor analysis from these syllabic perception studies, as Nearey (2003) suggests, to examine part-whole relationships within the segments. The logic transfers exactly, and we expect with our 2x2x2 matrix to find J-factors of approximately 3 for featural classification schemes and much lower figures for segmental classification schemes.

3. Results I: Inter-listener Accuracy Correlations

Since the Korean learners must acquire the ability to differentially identify stops and fricatives, the most obvious featural contrast to examine is manner. We first, then, examine whether there is correlation between listeners' accuracy in distinguishing coronal stops from fricatives and their accuracy in distinguishing labial stops from fricatives. Figure 1 plots coronal accuracy against labial accuracy, showing a very clear correlation across the listeners. Thus, listeners who were better at one contrast were also better at the other. Figure 1 also shows another general tendency, and that is for error rates to be lower for coronals than for labials; more of the symbol tokens lie in the lower right portion of the figure than in the upper left.

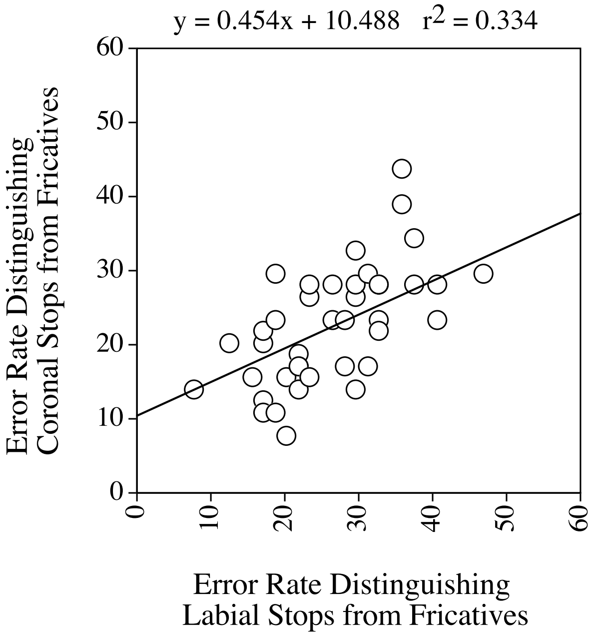

Figure 1. Error rates for distinguishing coronal stops from fricatives plotted against rates for labials. Each symbol corresponds to one listener.

Figure 1 suggests that there is a common skill that is acquired for distinguishing stops from fricatives, though the distinction is easier to make with coronals than labials. This is not particularly surprising, considering that the contrast is quite similar acoustically in coronal and labial situations, and that burst releases and fricative noise in labials tends to be weaker than in coronals.

If we examine the listeners' manner identification abilities for voiceless compared with voiced segments, we might not expect the same degree of generalization, since the contrast for voiceless segments is acoustically quite different than for that in voiced segments. For example, the distinction between fricative and stop in voiceless segments in initial position contrasts aspiration and fricative noise, while the distinction between voiced fricatives and stops contrasts silence and burst from noise.

Figure 2 plots error rates for distinguishing stops from fricatives for voiceless segments against those for voiced segments. As can readily be seen, the relationship between the accuracy rates is just as clear as it is across points of articulation in Figure 1. Here, the difference between voiced and voiceless is much bigger than the difference between coronals and labials, as indicated by the extreme down- and rightward shift of the data points. Error rates are much lower for voiceless than for voiced segments for every listener. However, again, listeners who are relatively good at distinguishing voiceless stops from fricatives tend to be relatively good at doing so for voiced stops and fricatives as well.

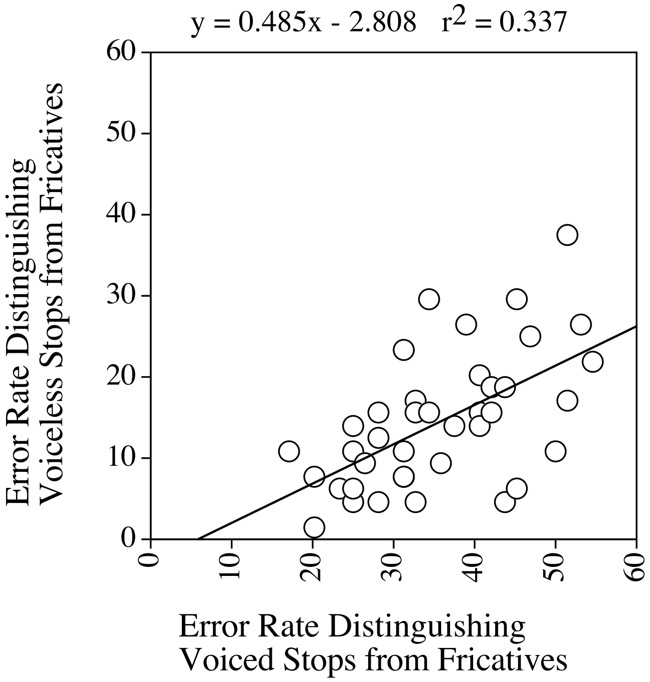

Figure 2. Error rates for distinguishing voiceless stops from fricatives plotted against rates for voiced segments. Each symbol corresponds to one listener.

To complete the picture of manner contrasts, we also examine how manner perception generalizes across prosodic positions. The relationships here are considerably more complicated, since there are four prosodic conditions to compare. Pearson r^2 values for regressions between each of the prosodic conditions are tabulated in Table 3. Each of the six correlations is statistically significant, and except for the final position, the relationships are consistently of about the same strength as those in the analyses that split by point of articulation and voicing. Thus, though the final position accuracies are less strongly connected to accuracy for the other positions, this analysis suggests that there is a general skill in distinguishing stops from fricatives that varies from listener to listener.

Table III. Pearson r^2 values for manner accuracy rates across prosodic position

	Initial	Pre-stress	Post-stress	Final
Initial	1.000	**0.283**	**0.302**	**0.196**
Pre-stress		1.000	**0.384**	**0.241**
Post-stress			1.000	**0.140**
Final				1.000

Of course, one obvious worry about this conclusion is that it might be the case that the various accuracy levels correlate with one another simply because the different listeners vary in level of experience with spoken English. Hence, their skills for any type of English performance will correlate with any other type of English performance.

While overall experience may be a factor in the results so far, it is not so strong a factor that all accuracies correlate with one another. In general, accuracy rates for various contrasts do not correlate well across these subjects. This is best illustrated with accuracy with voicing contrasts in the different prosodic conditions. Correlations between voicing accuracies in the prosodic conditions are summarized in Table IV.

Table IV. Pearson r^2 values for voicing accuracy rates across prosodic position

	Initial	Pre-stress	Post-stress	Final
Initial	1.000	**0.488**	0.074	0.027
Pre-stress		1.000	**0.325**	0.002
Post-stress			1.000	0.034
Final				1.000

Of the six comparisons, significant correlations are only obtained for two. Voicing accuracy for initial position is correlated with that in Pre-stress intervocalic position, and voicing accuracy for Pre-stress intervocalic position is correlated with accuracy for Post-stress intervocalic position. There is no hint of relationship between Final voicing accuracy and anything else or between Post-stress and Initial accuracy. These results suggest that there are three different aspects of voicing identification that are acquired at different rates by the different listeners. The first of these is an initial (or

Onset) skill which is for Initial and Pre-stress intervocalic conditions; the second is an intervocalic skill; and the third is a final position skill.

More generally, what the results reported here suggest is that a simple segmental model of segment identification, whereby individual segments have individual identification signatures, fails to explain why skills for various segments correlate in a manner which suggests generalized identification skills across segments which share a featural contrast. Especially clear is the skill of distinguishing stops and fricatives, which operates regardless of which segments are being distinguished and regardless of prosodic context.

4. Results II: Independent Feature Classification Models

While results reported so far indicate that segmental models are not complete, they provide only weak evidence for a fully independent featural model. Results for voicing accuracy indicate that some features clearly do not generalize across prosodic environments, for example. Also, regression values, while they do indicate a quite strong set of relationships in the data, are far from perfect. In this section, we report a more direct test of featural factorability by means of J-factor estimation.

J-factors for the current design are calculated with respect to the prediction that the probability of identifying a segment is the product of the marginal probabilities of identifying each featural distinction correctly. If (as discussed in Nearey, 2003) we average the featural probabilities get a single estimate of compositional accuracy (p(features)) we expect the following relationship: p(segment) = p(features)^J. Solving for J, we should thus expect: log[p(segment)]=J*log[p(features)], or: J = log[p(segment)]/log[p(features)].

Table V summarizes J-factor values estimated in this way for each of the prosodic conditions and for the entire data set combined. Each prosodic environment was examined separately, since the segmental contrasts in each prosodic environment in Korean are different due to prosodically-conditioned neutralization rules. As can be seen in Table V, J-factors are all larger than 2.6 and generally approximate the value of 3, which is what would be expected in cases of perfect feature independence. Compared to the results of Nearey's (2003) simulations, these values suggest a situation where classification relies heavily on independent featural contrasts.

Table V. J-factors Estimated for Each Prosodic Position.

	Initial	Pre-stress	Post-stress	Final	Overall
J =	2.627	3.034	2.816	2.750	2.710

However, we must also note that, as in previous work with J-factors, the values are not exactly at the expected value of 3, suggesting that there is some segment-specific components in the listeners' skills. It is interesting that the lowest J-factor values are found for segments in initial position, the prosodic form which is probably the most similar to that in Korean. If this is the case, these results match those for work by Boothroyd & Nitrouer (1988) and Benki (2003), in which they find that items whose forms are highly frequent tend to exhibit a reduction in the J-factor value. The accuracy of the perception of the overall stimulus is slightly greater than what would be expected by accuracy rates of the component parts.

One potential worry about the J-factor's however, concerns the possibility that averaging across different accuracies for the different features might distort the J-factors. To investigate this possibility, we examined the effect of averaging different featural accuracies in the ideal case in which the total accuracy really is the product of the individual feature accuracies. In general, the effect of having accuracy on each of the features diverges from one another is to inflate the J-factor. The greater the difference, the greater the inflation.

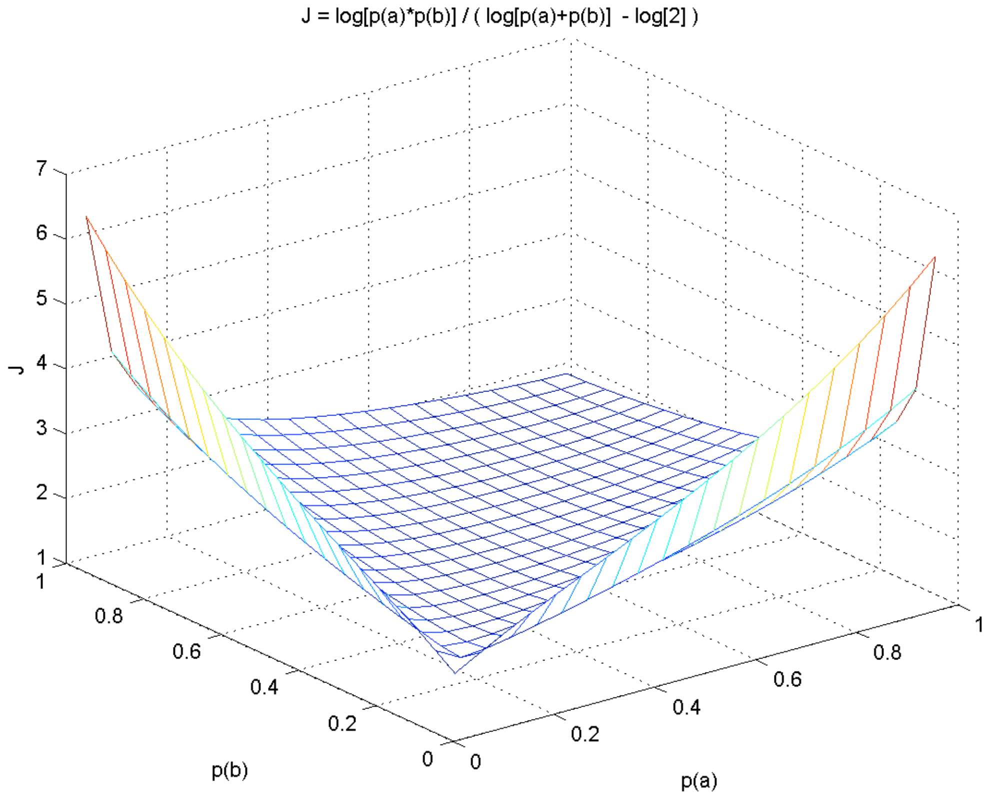

Figure 3. Values of J as a function of probabilities for two component factors (a and b), assuming independence of the factors.

This effect is plotted for a two-feature (labeled simply ‘a’ and ‘b’) case in Figure 3. In the case with two parts to the whole, as p(a) and p(b) vary from 0 to 1, the exponent applied to p(part) varies from 2 to about 7. Thus, in the case where independence does hold (and we know it), x = J, and can range from 2 to 7 (the more dissimilar the marginals, the larger the exponent). When p(a) and p(b) are equal (on the diagonal running front to back in Figure 3), J = 2, but as the two diverge either going leftward or rightward away from the diagonal, the value of J goes up, especially as one of the probabilities approaches 0.

So, for a given degree of positive dependence, if the marginals are very near one another in value, the J factor should be reliable and < 2, but if they're not, the same degree of dependence would give higher J values than expected. This inflation, combined with non-independent, segment specific perceptual criteria, which would deflate the value of J, could combine to give the value of J expected of independence.

To get more of an intuitive feel for how the J-factor models work, a more direct evaluation can be acquired by simply plotting the accuracy for each segment as a function of the product of the accuracies of each of the features that distinguish it from the other segments. An independent featural classification model predicts a one-to-one match between the segment accuracy and the product of the feature accuracies. Figure 4 is such a plot, with each symbol token indicating a particular segment in one of the four prosodic locations (for a total of 32 symbol tokens).

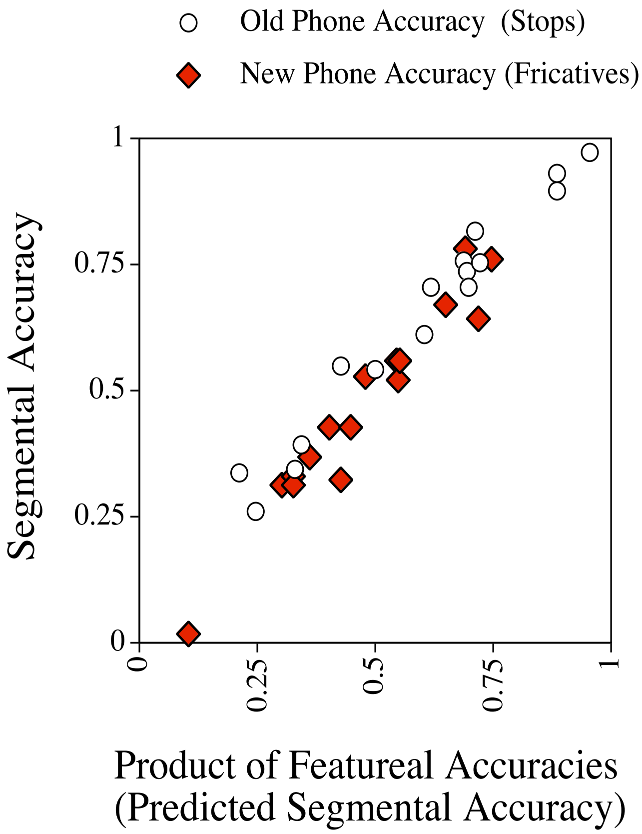

Figure 4. Accuracy for each segment in each prosodic location plotted against the product of the accuracy for each of the three features distinguishing it from the other segments.

Figure 4 clearly shows a strong, linear relationship between the accuracies predicted by an independent model and the observed accuracies. The correlations between predicted and actual are very high. Here, we have separated the stops from the fricatives, and find very high correlations for both stops: $r^2 = 0.969$, and fricatives: $r^2 = 0.944$. In addition, the slopes of the correlation functions are very nearly equal to 1, indicating an almost perfect fit. This fit is obtained across a very wide range of accuracies. Interestingly, there is, however, a very small divergence between the stops and the fricatives. The stop segment accuracies tend to drift slightly above the diagonal in Figure 3. This upward drift would correspond to a lower J-factor, since the segment accuracies are slightly higher than would be predicted from the component feature accuracies. Hence, it seems as if this divergence could also be seen as a familiarity effect of a similar kind to that noted above. Here, the stops, which are more similar to Korean segments, are identified slightly better than predicted.

5. Discussion and Conclusion.

The thrust of the current results is that a featural model of second language segment identification is more appropriate than a simple segmental model. Evidence against a strictly segmental model comes most clearly from the fact that skills necessary for particular featural contrasts are more apparent in certain learners than others. Learners who are relatively good at the skill with one pair of sounds are also relatively good with the same skill when applied to different segments. If learning a second language involves simple and isolated segmental skills, there is no explanation as to why skills that involve the same contrast in different segments should be systematically acquired.

Even more suggestive evidence for a generalized featural model is provided by the J-factor analysis, in which the expectations based on a large degree of independence of featural decisions are met. The goodness of fit between the model predictions and actual accuracy for individual segments is remarkably high. In fact, the goodness of fit is so good that we wonder if there might be some artifactual compression of the values for J. It is not the case that there is no variability in the J-factor; the J-factor does seem to vary in small amounts, whereby segments that are most analogous to what is found in Korean exhibit a slight deflation of the J-factor. Thus, it is not the case that there is no sensible patterning to the J-factor results. However, a weakness of our current approach is that we have no estimate of what J-factor values would be in a completely segmental model with random variation in segmental identifiability. Nearey's (2003) simulations suggest fairly strongly that the current values indicate a large degree of independence in the featural components of the segments, but we are unsure of just how to assess the degree of dependence indicated by the slight deflation of the J-factor.

It is clear that, though results suggest that a generalized feature model is closer to being right than a segmental model, there are definite indications of a lack of complete featural independence in the listeners' performance.

First, there is a slight deflation of the J-factor for segments that are analogous to Korean segments, suggesting a role for response bias in favor of the more familiar segments.

Second, and even more clearly, the results for voicing accuracy suggest that voicing contrasts in different prosodic positions are best considered to be basically unrelated. We are not sure at this stage whether these results say something about the target English system or about Korean expectations applied to the English system. Voicing contrasts in English are allophonically different by position. Voicing is an aspirated – unaspirated contrast in the Initial and Pre-stress positions, a change in closure dynamics mixed with voicing in Final position, and a weakened hybrid in Post-stress position. (Impressionistically, the subjects did not flap the coronal stops in intervocalic position, so the contrast was not neutralized.) Thus, it would make sense that the Korean learners would have to learn what to do with each of these different actual contrasts individually. However, this interpretation is complicated by the fact that the Korean laryngeal contrasts also are allophonically different in initial, intervocalic, and final position. Hence, the three separate components to inter-listener variation found here may correspond to the three prosodic variants in the Korean system.

With these questions of how to deal with large positional differences aside, however, the general thrust of the current results suggests a model in which second language learners begin with generalized labeling criteria which get individually tuned with familiarity with individual segments. The native language provides such familiarity, making identification criteria for 'similar phones' more specific, and less general.

Acknowledgements.

Work supported by the NIDCD (grant #R03 DC04095-01A2) and by the NSF (grant numbers BCS-9910701 and BCS-046540). We also express our appreciation to Prof. Mi-Hee Cho of Kyonggi University for help in collecting the data reported here and to Kyoko Nagao and Hanyong Park for their work on the design and processing of the data presented here, and Steve Winters and Becca Ronquest for their criticism of this ms.

References.

BENKI, J. (2003). Quantitative evaluation of lexical status, word frequency, and neighborhood density as context effects in spoken word recognition. *Journal of the Acoustical Society of America*, **113**: 1689 - 1705.

BOOTHROYD & NITTROUER, S. (1988). Mathematical treatment of context effects in phoneme and word recognition. *Journal of the Acoustical Society of America***, 84**: 101 - 114**.**

BRADLOW, A.R., D.B. PISONI, R. AKAHANE-YAMADA, & Y. TOHKURA (1997). Training Japanese listeners to identify English /r/ and /l/, IV: Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, **101**: 2299 – 2310.

BRANNAN, K (2002). The role of perception in differential substitution. *Canadian Journal of Linguistics*, **47**: 1 – 46.

BROSELOW, E., & D. FINER (1991). Parameter setting in Second Language phonology and syntax. *Second Language Research*, **7**: 25 – 59.

BROSELOW, E., HURTIG, R.R., & RINGEN, C. (1987). The perception of second language prosody. In Ioup, G., & Weinberger, S.H. (Eds.) *Interlanguage Phonology* (pp. 350-362). New York: Newbury House.

ECKMAN, F.R. (1991). The Structural Conformity Hypothesis and the acquisition of consonant clusters in the ineterlanguage of ESL learners. *Studies in Second Language Acquisition*, **13**: 23 – 41.

ECKMAN, F. AND G. IVERSON (1994). Pronunciation difficulties in ESL: Coda consonants in English Interlanguage. In M. Yavas (Ed.), *First and Second Language Phonology* (pp. 251-265). San Diego, CA: Singular.

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). Factors affecting the degree of perceived foreign accent in English sentences. *Journal of the Acoustical Society of America*, 91: 370-389

FLEGE, J.E. (1995). Second language speech learning: Theory, findings, and problems. In Strange (1995), pp. 233 - 277.

NEAREY, T (2003). On the factorability of phonological units in speech perception. In J. Local, R. Ogden, & R. Temple (eds.), *Papers in Laboratory Phonology VI: Phonetic Interpretation*. Cambridge: Cambridge University Press, pp. 197 – 221.

STRANGE, W., R. AKAHANE-YAMADA, R. KUBO, S.A. TRENT, K. NISHI, & JJ. JENKINS (1998). Perceptual assimilation of American English vowels by Japanese listeners. *Journal of Phonetics*, 26: 311 – 344.