

Cross-Language Perception of Syllable Affiliation: Effects of Voicing and Language Background*

BYUNG-JIN LIM¹, KEN DE JONG², and KYOKO NAGAO²
The University of Wisconsin-Madison¹ and Indiana University²

0. Introduction

This paper examines the effect of linguistic experience in listeners' perception of syllabic affiliation. While there is a large amount that we know about the second language acquisition of paradigmatic contrasts, there is relatively little parallel literature concerning the acquisition of syntagmatic contrasts. To do this, we compared first language and second language perceptions of rate-varied stimuli, which have been noted in previous studies to vary in perceived syllabic structure. We also in parallel compared first and second language perception of 'voiced' and 'voiceless' stops in the same corpus, in order to ascertain how well the listeners would do with a phonemic contrast in the same type of data. In general, this study finds that second language learners are very much like native listeners in the perception of syllabic affiliation, but not in the perception of voicing.

Stetson (1951, also partially documented in Tuller and Kelso 1991) noted that singleton coda consonants (VC) appear to modulate into onset consonants (CV) when repeated at relatively fast rates. Thus, with CV forms, rate increases seem to generate a continuum ranging from clearly perceived codas to clearly perceived onsets. Hence the current study examines perceptual responses to such repetitive productions to ascertain the degree to which speakers of languages with different syllabic inventories actually do perceive such resyllabification.

In addition, in order to examine the generalizability of the resyllabification phenomenon across segments with different temporal characteristics, the production corpus varied in the 'voicing' of the resyllabified consonants. Hence, we are able to examine the same listeners' responses to voicing contrasts in the same data. Voice onset time (VOT) is a well-known acoustic attribute distinguishing the voicing contrast in initial stops (Lisker and Abramson 1964). Changes in speech rate affect the range of VOT values, particularly for voiceless stops (Miller et al. 1986; Miller and Volaitis 1989; Volaitis and Miller 1992; Pind 1995; Kessinger

* Research supported by the NIDCD (grant R03 DC04095-01A2) and by the NSF (grant BCS-9910701).

and Blumstein 1997, 1998). Either due to some relational property of the speech itself, or due to listeners' learning of these rate effects, the perceptual boundary between voiced and voiceless stop consonants accordingly shifts toward shorter VOT values as syllable duration decreases (Summerfield 1981, Miller et al. 1986, Miller and Volaitis 1989, Volaitis and Miller 1992).

Non-native listeners were speakers of languages for which either voicing or syllabification is grammatically restricted. One group of listeners was native speakers of Japanese, a language which does not allow post-vocalic obstruents without a following vowel. Japanese does allow geminate consonants which are generally treated as acting as the coda of the preceding syllable, and among younger speakers, these geminates can contrast in voicing. Hence, Japanese seems to be a case of a language for which an obstruent appearing as a singly affiliated coda is not native, but in which, nevertheless, voicing contrasts in post-vocalic position are possible. The second group of listeners was native speakers of Korean. Korean, in contrast to Japanese, does allow words ending in an obstruent. However, also in contrast with Japanese, consonants which are co-syllabic with the preceding vowel do not contrast in voicing, due to a phonological neutralization rule. A third group of native English listeners acted as a control group. All three languages have a voicing contrast in onset position, though the exact contrast differs across all three languages. Table 1 summarizes these facts.

Table 1. Cross-language comparison: syllable affiliation and voicing contrast

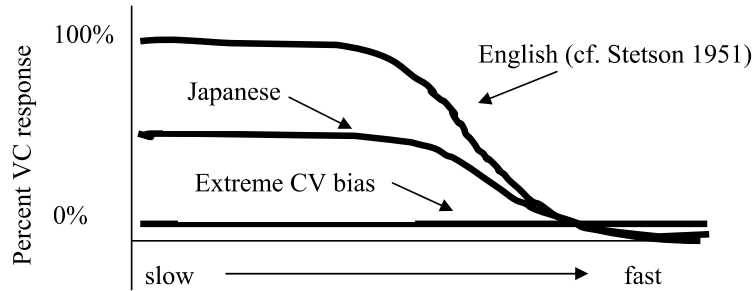
	English		Japanese		Korean	
	voiced	voiceless	voiced	voiceless	voiced	voiceless
Onset	b	p	b	p	b	p
Coda	b	p	geminate only		neutralized	

1. Hypotheses

1.1. Syllabification Perception

We hypothesize that listeners' perception of syllable affiliation depends on the syllabic inventory of their native language. Specifically, since Japanese does not allow syllable final stops in coda positions while English and Korean do, Japanese listeners will exhibit a bias toward perceiving syllable structures as CV. In the extreme case, Japanese listeners might not hear VC forms at all. Another possibility would be that Japanese listeners are unsure about how to categorize English VC forms and have identification functions which start around 50% at slow rates, and shift as stimuli sound more like canonical CV's. Coupled with Stetson's observations, then, we might expect the pattern of results in Figure 1.

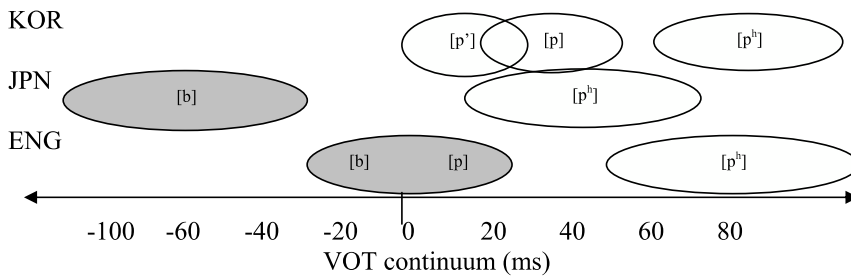
Figure 1. Prediction for perception of syllable affiliation



1.2. Voicing Perception

We hypothesize that voicing perception will be influenced by the perceivers' native voicing categories. For onsets, Korean has a three-way distinction in voiceless stops.¹ Here, the boundary between aspirated and unaspirated stops is higher in VOT than that in English. Japanese has a two-way contrast whose boundary seems to be lower in VOT value than that in English. The categories in VOT value are illustrated in Figure 2. Given the direction of the boundary differences from English, we expect a bias in Japanese listeners against hearing items as voiced, and a bias in Koreans against hearing items as aspirated.

Figure 2. Schematic phonetic categorization for voicing on the VOT continuum (Japanese and Korean VOT values from Shimizu 1996)

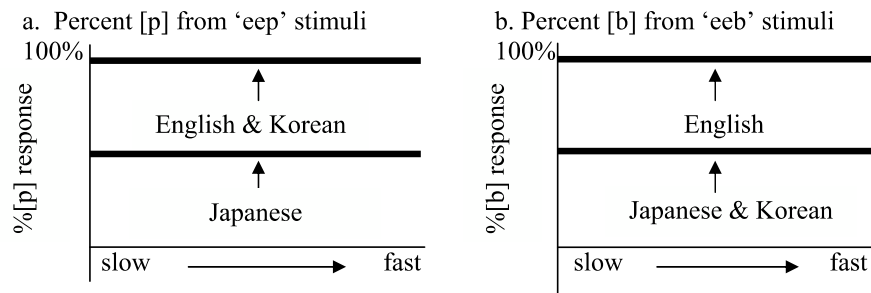


Predictions with respect to the voicing of coda consonants are more complicated. Since Japanese does not have singleton coda consonants, we might expect that Japanese natives would have difficulty perceiving both voiced and voiceless stops, and be simply guessing as to the voicing of coda stops. Also, since Korean exhibits the neutralization of coda voicing toward voicelessness, we might expect a bias toward voiceless labeling. These combined predictions can be schematized as in Figure 3. However, we should also note that younger Japanese speakers are

¹ Intervocally, however, lax stops can be voiced in Korean.

familiar with nativized foreign words which contrast in the voicing of geminate stops, and Korean does contrast voiced (lenis) and voiceless (fortis or aspirated) stops in intervocalic position. Hence, we might find native-like perception of voicing.

Figure 3. Predictions for voicing perception



2. The Perception Experiment

2.1. Methods

2.1.1. Subjects

The subjects were 21 native speakers of American English, 17 native speakers of Japanese, and 13 native speakers of Korean. They were all of university age with normal hearing and no linguistic training. All were recruited through responses to posted advertisements and were paid \$10 each for their participation.

2.1.2. Stimuli

Stimuli were extracted from a repetitive production experiment in which four English talkers produced voiced and voiceless labial stop onsets and codas at gradual increasing tempi from 450 to 150 ms/syllable controlled by metronomes (for details concerning the production of the forms, see de Jong 2001). Each stimulus included three repetitions of one of each of the four monosyllables, 'pea', 'bee', 'eep', and 'eeb'. 404 stimuli were presented to each listener group.

2.1.3. Procedures

After listening to recordings through headphones, subjects were asked to identify what they heard by clicking on a computer monitor. There were four buttons, labeled 'pea', 'bee', 'eep', and 'eeb'. Prior to experimentation, a practice run of 8 trials was given. During experimentation, a five-minute break was given after every 100 trials.

2.2. Results

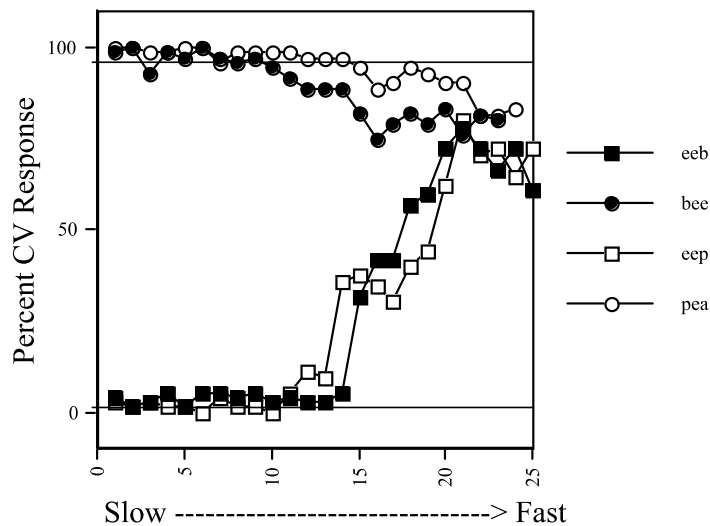
2.2.1. Perception of Syllable Affiliation

To replicate Stetson's observations, we first turn to the English subjects. Figure 4 shows the results of perception of syllable affiliation for CV and VC structures by

Cross-Language Perception of Syllable Affiliation

English listeners. As shown in the top of Figure 4, there is little effect of speech rate on the perception of CV syllable structure. Listeners identified CV productions as CV forms. We do note a slight decrease in CV responses for CV stimuli, though the amount of change is quite small (less than 20%). In the perception of VC tokens, as shown in the bottom of Figure 4, we note a clear shift in identification from VC to CV as speech rate increases. This pattern agrees with Stetson (1951). Voicing differences do not seem to affect the pattern of identification. Both ‘eeb’ and ‘eep’ stimuli elicit ogival functions, though there is an earlier and more gradual shift in ‘eep’ stimuli.

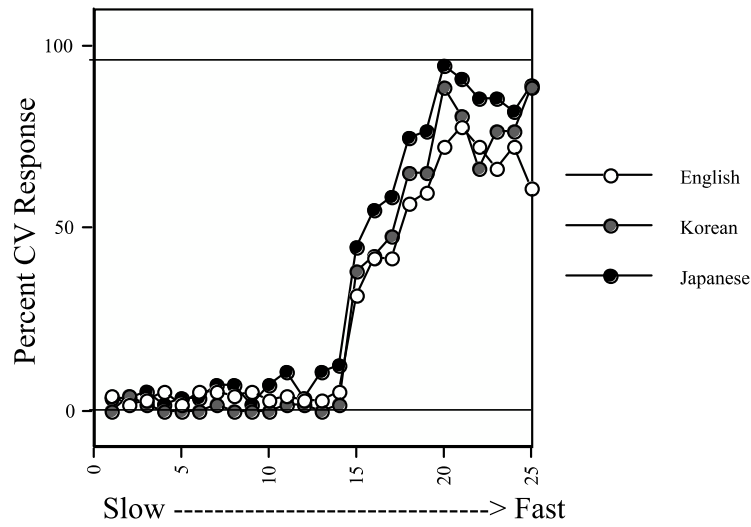
Figure 4. Identification functions for CV (circles) and VC (squares) stimuli. The horizontal scale indicates the location of the stimulus in the trial from the slow rate beginning (left) to the fast rate ending (right).



2.2.2. Cross-Language Perception of Syllable Affiliation

Figure 5 shows the identification of syllable affiliation for non-native listeners. Responses plotted here are only for ‘eeb’ stimuli. All three groups exhibit resyllabification, and the category shift seems to occur at nearly the same location in the continuum. Thus, all three response functions are very similar in shape and location. There is a difference in the three groups’ responses to faster rates, however. Especially Japanese listeners were more likely to label fast rate tokens as CV’s. Hence, there is an apparent bias toward CV labeling, but only at fast repetition rates. Responses to ‘eep’ tokens were similar.

Figure 5. Identification functions for VC tokens for native speakers of English, Japanese and Korean



2.2.3. Cross-Language Perception of Voicing Contrast

For comparison, we turn to the perception of the voicing contrast. For reasons noted above, we consider voicing perception for stops produced as onset stops separately from stops produced as codas. The results of voicing perception for stops produced as CV's are shown in Figure 6.² Examining the English listeners first (unfilled symbols), one notes that native listeners are generally very good at identifying the voiceless consonants, while voiced consonants at fast rates tend to be misidentified as voiceless. These general patterns are visible in the non-native listeners as well, though with some notable differences. Korean listeners (gray symbols) tended to hear some voiceless tokens as voiced, and Japanese listeners (black symbols) tended to hear voiced tokens as voiceless. These effects are true even of slow rate stimuli. These are what one would expect as biases induced by the native categories as illustrated in Figure 2. Koreans tend to require more aspiration for voiceless stops, and Japanese tend to require more voicing for voiced stops.

² Note the results shown here only include three of the four speakers, since a fourth speaker elicited very different responses.

Cross-Language Perception of Syllable Affiliation

Figure 6. Proportion of voiceless responses to phonemically voiced and voiceless stimuli by the three language groups

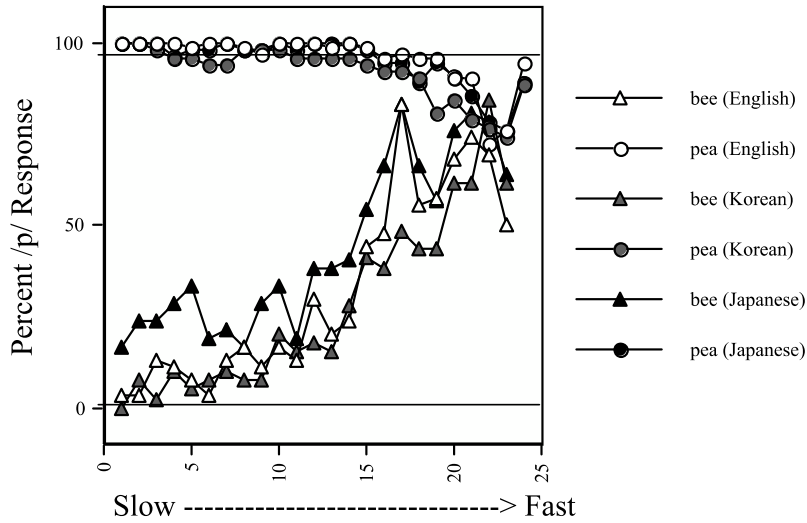
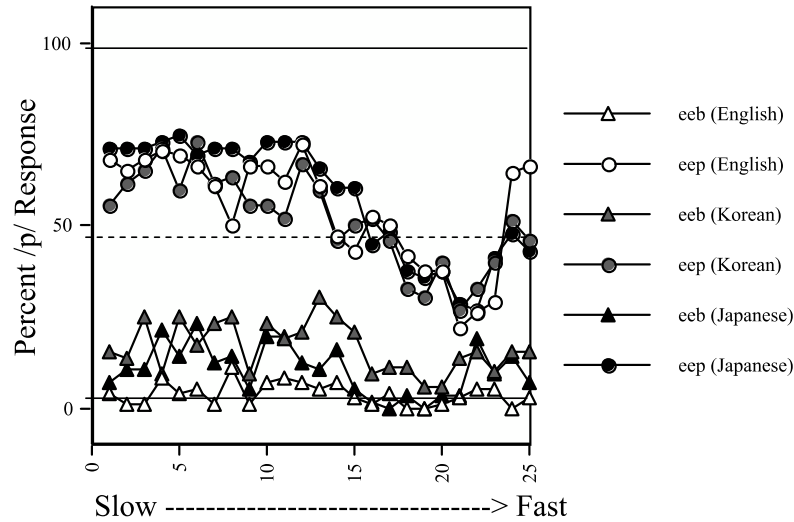


Figure 7 plots the identification of voicing for stops produced as codas. Here, English listeners were generally good at identifying voiced stops, but were very poor at identifying voiceless stops. Fast rates tended to induce even more mis-identifications of voiceless as voiced stops. Non-native listeners showed the same asymmetry, and were, in general, very similar to native listeners. However, as with onset stops, we also find differences. While native listeners were very good at identifying voiced coda stops, both non-native groups were less consistent, especially Korean listeners. Korean listeners exhibited identification functions shifted toward chance with respect to native listeners.

Figure 7. Proportion of voiceless responses to phonemically voiced and voiceless stimuli by the three language groups.



3. Discussion and Conclusions

3.1. Perception of Resyllabification

Regardless of the speakers' native language syllabic inventory, all listener groups showed very similar patterns of perceptual resyllabification as a function of speech rate change. Currently, it is unclear whether this similarity is due more to categorical changes in the articulation of the stimuli, or due entirely to similarities in the perceptual boundaries for the listeners. However, it is remarkable not only that English listeners bear out Stetson's observations, but also that acquirers whose native languages do not have the same inventory of structures exhibit essentially the same perceptual patterns. Thus, it seems that the perception of syllable affiliation seems to be, if not a cross-language universal skill, at least an aspect of production which is very accessible to acquirers.

Beyond this, it is not the case that non-native perceivers are identical to native perceivers. Particularly interesting differences begin to arise in the perception of fast rate utterances. There is a tendency for the non-native perceivers to actually be more likely to label VC's as being resyllabified as CV's. This very subtle bias away from the non-native VC structure seems to fit nicely with a model in which their perception of forms in a non-native language retains a bias toward forms which have close analogies in their native languages. In this case, Japanese-speaking listeners would be more likely to perceive VC forms as resyllabified into

CV forms because CV forms occur in Japanese native vocabulary. However, such an explanation does not fit well with the Korean listeners, for whom VC forms are native. The similarity of the Japanese and Korean listeners with regard to resyllabification suggests a somewhat different explanation. Rather than non-native listeners exhibiting a bias toward native forms, the English listeners are exhibiting resistance to labeling the VC forms as something other than what the speaker intended. That is, English listeners may simply be better at detecting the traces of the fast rate stimuli's VC origin than the non-native listeners. Previous acoustic analyses of the stimuli that the current study used show differences between resyllabified codas and onsets both in the temporal characteristics of the closure and in the spectral quality of the vowel (de Jong 2001). Apparently, native listeners are sensitive to these differences.

3.2. Effects of Phonemic Contrasts on Voicing Perception

As previously mentioned, both Japanese and Korean have a contrast in onset voicing roughly analogous to the English contrast. Thus, it is not surprising that both non-native listener groups are generally good at identifying voicing in the current data. Further, the voicing categories in the three languages are different in detail. For instance, there are typically different boundaries on a VOT continuum for each language. This difference in voicing categorization seems to be reflected in subtle biases in the non-native responses; English /b/ tended to be heard as [p] by Japanese listeners while English /p/ tended to be heard as [b] by Korean listeners.

At first blush, this pattern of results seems to match that for syllabification. Non-native speakers of English generally match native speakers with a small bias toward responding in terms of a native category. However, there is a difference between the two sets of results with respect to rate. Specifically, syllabification biases showed up at fast rates, where there is some uncertainty in the native speaker responses. Voicing biases are detectable at all rates, particularly the Japanese bias toward voiceless responses. Hence, it seems that the existence of the non-native category is detectable even in cases in which native listeners are consistent in their responses.

Considering the voicing contrast in codas, which we expect to be non-native, we find yet another pattern. Non-native listeners are simply less consistent at identifying the voicing of coda stops. This is particularly evident in the Korean listeners, whose identification functions were both shifted toward chance from the native identification functions. This, again, makes intuitive sense if we consider that not having a native contrast will result in less developed identification skills. It is surprising that the Korean listeners had problems with voiceless stops, even though they are said to produce the voiceless variants. We suspect this indicates that the neutralization of voicing in coda position creates in perception a lack of attention to voicing cues such as vowel duration. If this is the best explanation for the current results, it underscores the point made in the previous section concerning syllabification. Despite not having experience with determining the syllabic

affiliation of stops between vowels, both non-native listener groups performed very much like natives of English. Syllabification of obstruents does seem to be remarkably accessible to non-native listeners.

There are a number of additional issues which the current results raise, most particularly concerning the general pattern of voicing identification in onsets and codas. For example, voiceless onsets and voiced codas both are well identified, while voiced onsets and voiceless codas are relatively poorly identified. In both of the latter cases, in addition, the poor identification gets worse as speech rate increases. What is particularly remarkable about this pattern is how similar it is for native and non-native listeners. This leads us to believe that the explanation for these overall patterns of responses will be found in how voicing is expressed in the repetitive speech, and not in peculiarities of the listeners' perceptual categories. We are currently in the process of sorting out the relationship between the numerous cues to voicing and the identification patterns of the three groups.

References

- de Jong, K. J. 2001. Rate-Induced Resyllabification Revisited. *Language and Speech* 44(2):197-216.
- Kessinger, R. H., and S. E. Blumstein. 1997. Effects of Speaking Rate on Voice-Onset Time in Thai, French, and English. *Journal of Phonetics* 25:143-68.
- Kessinger, R. H., and S. E. Blumstein. 1998. Effects of Speaking Rate on Voice-Onset and Vowel Production: Some Implications for Perception Studies. *Journal of Phonetics* 26:117-28.
- Lisker, L., and A. S. Abramson. 1964. A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. *Word* 20:384-422.
- Miller, J. L., K. P. Green, and A. Reeves. 1986. Speaking Rate and Segments: A Look at the Relation Between Speech Production and Speech Perception for the Voicing Contrast. *Phonetica* 43:104-15.
- Miller, J. L., and L. E. Volaitis. 1989. Effects of Speaking Rate on the Perceptual Structure of a Phonetic Category. *Perception and Psychophysics* 46:505-12.
- Pind, J. 1995. Speaking Rate, Voice-Onset Time, and Quantity: The Search for Higher Order Invariants for Two Icelandic Speech Cues. *Perception and Psychophysics* 57:291-304.
- Shimizu, K. 1996. *A Cross-Language Study of Voicing Contrasts of Stop Consonants in Six Asian Languages*. Tokyo: Seibido.
- Stetson, R. H. 1951. *Motor Phonetics: A Study of Speech Movements in Action*. Amsterdam: Noth-Holland.
- Summerfield, A. Q. 1981. Articulatory Rate and Perceptual Constancy in Phonetic Perception. *Journal of Experimental Psychology: Human Perception and Performance* 7:1074-95.
- Tuller, B., and J. A. S. Kelso. 1991. The Production and Perception of Syllable Structure. *Journal of Speech and Hearing Science* 34:501-8.

Cross-Language Perception of Syllable Affiliation

Volaitis, L. E., and J. L. Miller. 1992. Phonetic Prototypes: Influence of Place of Articulation and Speaking Rate on the Internal Structure of Voicing Categories. *Journal of the Acoustical Society of America* 92:723-735.

Byung-Jin Lim
Van Hise Hall 1214
East Asian Languages and Literature Department
The University of Wisconsin-Madison
1220 Linden Drive
Madison, WI 53706

blim3@wisc.edu

Ken de Jong
Memorial Hall 322
1021 East 3rd Street
Linguistics Department
Indiana University
Bloomington, IN 47405

kdejong@indiana.edu

Kyoko Nagao
Memorial Hall 322
1021 East 3rd Street
Linguistics Department
Indiana University
Bloomington, IN 47405

knagao@indiana.edu