# Learning to produce a multidimensional laryngeal contrast

*Charles B. Chang*[*]

University of California, Berkeley
cbchang@cal.berkeley.edu

**ABSTRACT**

Research on how second-language (L2) learners acquire L2 laryngeal categories has focused on languages with "voiced" and "voiceless" categories that differ in terms of one main cue: voice onset time. The present study examines how L2 learners come to produce a laryngeal contrast that requires the use of a second phonetic dimension—namely, the three-way Korean laryngeal contrast among lenis, fortis, and aspirated stops. In a five-week longitudinal study, 26 adult native English speakers learning Korean completed a reading task in which they pronounced Korean stops in a low vowel context. Results of acoustic analyses show that while the majority of learners are eventually successful at producing a full three-way contrast, there is wide variation in the way in which they produce it. This paper describes the range of variation in phonetic spaces that learners produce, shows how these differ from the findings of cross-linguistic perception studies on English speakers hearing Korean, and concludes that a perseverative kind of "equivalence classification" plays a large role in how learners link L2 laryngeal categories to L1 laryngeal categories.

**Keywords:** laryngeal contrast, voice onset time, fundamental frequency, Korean, equivalence classification.

## 1. INTRODUCTION

Research on how second-language (L2) learners acquire laryngeal categories that differ from the laryngeal categories of their first language (L1) has generally concentrated on languages with two laryngeal categories differing between L1 and L2 in terms of the same primary cue: voice onset time, or VOT (e.g. French and English: Caramazza *et al*. 1973, Flege 1987; Spanish and English: Flege and Eefting 1988; Italian and English: Flege *et al*. 1995; Portuguese and English: Major 1996). In the present study, I examine how L2 learners come to produce a laryngeal contrast that requires the use of a second phonetic dimension in addition to VOT—namely, the three-way Korean laryngeal contrast among lenis, fortis, and aspirated stops, which in initial position differ primarily in terms of VOT and fundamental frequency ($f_0$) onset (cf. Han and Weitzman 1970, Kim 2004, *inter alia*). How do learners use (or not use) $f_0$ onset in conjunction with VOT to realize this three-way contrast?

Relatively little work on L2 speech has examined Korean as L2, rather than L1. With regard to L2 perception of Korean, two studies have examined how L1 English speakers interpret Korean word-initial stop consonants. Francis and Nusbaum (2002) found that before training, L1 English speakers (naïve listeners who were not learning Korean) mostly relied on differences in VOT (and co-varying differences in rate of amplitude change) to distinguish the three laryngeal categories, but after training, seemed to use both VOT and $f_0$ onset differences (along with co-varying differences in the clarity of formant structure at vowel onset) to distinguish them (however, see Shin 2007 for differing results with trained learners of Korean). The perceptual data show, moreover, that after training, English speakers' perception approximates that of native Korean speakers, who break up the [VOT x $f_0$] phonetic space in the manner shown in Kim (2004), where tokens with short-lag VOT are consistently perceived as fortis and tokens with long-lag VOT are perceived as either lenis or aspirated depending on the VOT and on the $f_0$ onset.

While Francis and Nusbaum's (2002) perception study utilized identification and difference rating tasks, Schmidt's (2007) cross-linguistic perception study instead had subjects—also L1 English speakers with no knowledge of Korean—label Korean sounds as the perceptually closest English sound and rate the similarity of the English sound to the Korean sound. Her results show that subjects overwhelmingly labeled Korean lenis stops and aspirated stops as English voiceless stops and Korean fortis stops as English voiced stops. However, the Korean categories differed in terms of how similar to English categories they were perceived as being: aspirated stops were rated as more similar to English stops than lenis or fortis stops were. This suggests that for L1 English learners of Korean, the default "equivalence classifications" (Flege 1987) of Korean and English stops are aspirated-voiceless, lenis-voiceless, and fortis-voiced, but that the strength of the cross-language category identification varies across category pairings.

Whereas research on L2 perception of Korean has often focused on L2-naïve subjects, studies that have looked at L2 production of Korean have generally examined people actively learning the language. In one such study, Kim and Lotto (2002) found that intermediate Korean learners (most of whom were L1 English speakers) produced distinctions between the three stop types using VOT, but not closure duration or $f_0$ onset. Shin's (2007) study of elementary Korean learners resulted in similar findings with L1 English learners, who tended to rely just on VOT to produce the contrast. On the other hand, learners whose L1 was a tone language (e.g. Mandarin, Cantonese) were found to use $f_0$ as a cue more often than the L1 English learners.

Taken together, the results of studies of L2 perception and production of Korean suggest that L1 English speakers, and perhaps speakers of non-tone languages more generally, can be trained to use $f_0$ in perception, but nevertheless tend to utilize VOT rather than $f_0$ to distinguish the Korean laryngeal categories in production. This pattern of production contrasts with that of mature native speakers, who use both dimensions, as well as with that of children acquiring Korean as L1, who separate the lenis and aspirated categories in $f_0$ well before they separate them in VOT (cf. Jun 2006), although by the age of five years they use both cues reliably (Lee and Iverson 2008).

Although Schmidt's (2007) cross-linguistic perceptual findings show consistency in the way learners assimilate Korean categories to English categories, they make no predictions regarding how learners will distinguish the lenis and aspirated categories that are both assimilated to the voiceless category of English. Kim and Lotto (2002), as well as Shin (2007), suggest that learners mainly use VOT to distinguish these categories in production; however, the amount of VOT overlap between learners' lenis and aspirated stop productions—even within one place of articulation—is so large that it is unclear whether learners are actually producing a reliable three-way contrast in VOT.

A second reason to re-examine L2 learners' production of this contrast is the existence of a conflict in cues contributing to cross-language equivalence classification. If we were to pair the Korean and English laryngeal categories on the basis of phonetic similarity (specifically, in terms of similarity in VOT and $f_0$), aspirated stops would be paired with voiceless stops, since these categories are both long in VOT and high in $f_0$ onset. However, it is unclear how lenis stops and fortis stops should be classified, since each of these categories resembles voiced stops in one way and voiceless stops in another way. Lenis stops are relatively long in VOT like voiceless stops, but low in $f_0$ like voiced stops; fortis stops, on the other hand, are short in VOT like voiced stops, but high in $f_0$ like voiceless stops. Thus, linking lenis and fortis stops to English categories is not straightforward, given that most English speakers show some degree of sensitivity to the $f_0$ difference between voiced and voiceless stops (cf. Haggard *et al*. 1970).

In the present study, I re-examine how L1 English late learners of Korean produce the Korean laryngeal contrast, focusing on an L1 and L2 that do not share the same orthography to avoid the confound of orthographic equivalence present in the majority of studies on L2 voicing categories. The main research question is the following: given little to no explicit phonetic instruction, how successful are late learners of Korean at producing Korean laryngeal categories like native speakers? We will see if, using VOT and $f_0$ onset, learners manage to produce a full three-way contrast, as well as if they are consistent in their L2 phonetic spaces. Finally, we will make some generalizations about the nature of learners' deviation from the native Korean phonetic space.

## 2. METHODS

A production experiment was conducted weekly starting from one week into the language class that study participants were taking. Every week participants completed a reading task in which they saw a Korean stimulus (spelled in Korean orthography) and read it aloud. Stimuli were presented a total of four times, once each in four randomized blocks following a practice session of five items. Each item was presented on screen for 1.5 seconds and then replaced by a picture of a green traffic light to cue the participant to produce the item. Audio was recorded via a head-mounted condenser microphone for two seconds starting at the time point at which the green light appeared on screen, and the inter-stimulus interval from the end of this recording to the presentation of the following item was one second. All stimuli presentation and audio recording was done in DMDX 3.2.6.3 (Forster 2008) on a laptop computer.

The set of Korean stimuli consisted of 22 Korean monosyllables representing most of the phonemic contrasts in the language. The stimuli were generally of the form CV to make them as easy as possible for novice learners to read, with the vowels in the nine critical items (3 laryngeal categories x 3 stop places of articulation) being uniformly /a/. The same set of stimuli was used in every week of the study.

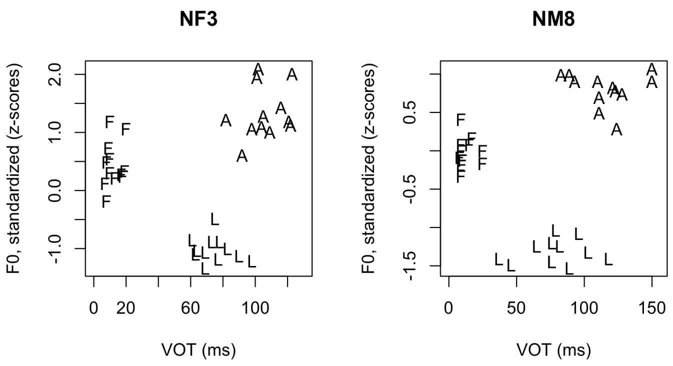
Participants were 26 late learners of Korean (4 males, 22 females; 21–26 years old), native speakers of American English with no prior exposure to Korean taking a six-week course of intensive Korean immersion instruction at the time of the study. On average these learners received four hours of instruction a day, for a total of approximately 82 hours of instruction by the end of the program (roughly equivalent to one semester of college-level Korean). In exit questionnaires, participants reported that class time constituted the majority of their experience with Korean, both in terms of listening and speaking.

Acoustic analysis of recordings was conducted using Praat 5.0.26 (Boersma and Weenink 2008). Manual measurements of VOT and $f_0$ onset were taken on learners' productions of critical items. VOT was measured off a wide-band Fourier spectrogram with a Gaussian window shape (window length: 5 ms; dynamic range: 50 dB; pre-emphasis: 6.0 dB/oct) as the time at voicing onset minus the time at the stop burst. To obtain stable measurements of $f_0$ onset, the average wavelength of the first three regular glottal periods in the vowel was calculated from the waveform and converted into a frequency value. Initial periods were skipped if they were irregular (e.g. more than 33% longer or shorter than the following period); however, tokens requiring more than five periods of the vowel onset to be skipped were discarded. In order to put male and female learners on the same $f_0$ scale, raw $f_0$ measurements were furthermore standardized to *z*-scores by learner (by subtracting the learner's mean $f_0$ over the duration of the study and dividing by the square root of the learner's variance in $f_0$ over the duration of the study).

## 3. RESULTS

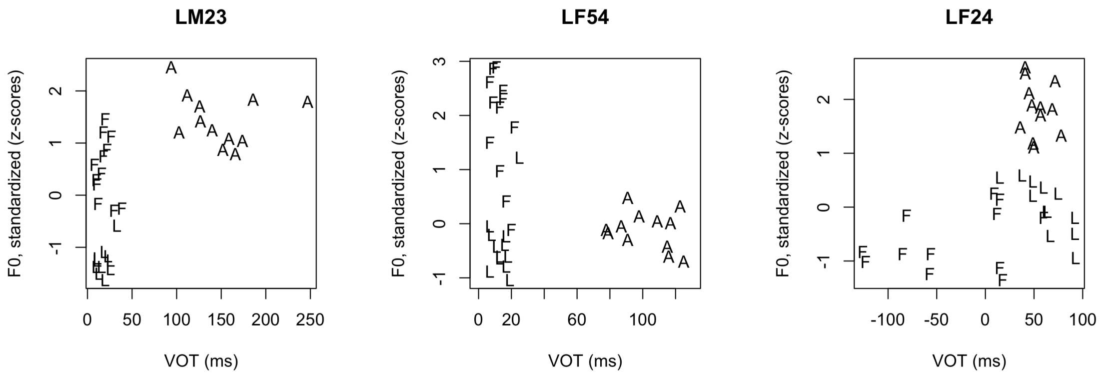
The phonetic spaces of native Korean speakers are generally consistent with Kim (2004) in terms of how the Korean laryngeal categories are realized with respect to VOT and $f_0$ onset. Fortis stops are produced with short VOT and an elevated $f_0$ onset; lenis stops are produced with longer VOT and a low $f_0$ onset; and aspirated stops are produced with the longest VOT and the highest $f_0$ onset (cf. Figure 1). For most native speakers, lenis and aspirated stops overlap considerably in VOT, and fortis and aspirated stops overlap considerably in $f_0$, but none of these categories overlap in both dimensions. Thus, VOT and $f_0$ are necessary and sufficient cues for distinguishing the three laryngeal types.

**Figure 1:** Representative scatter plots of native Korean speakers' productions (L = lenis, F = fortis, A = aspirated).



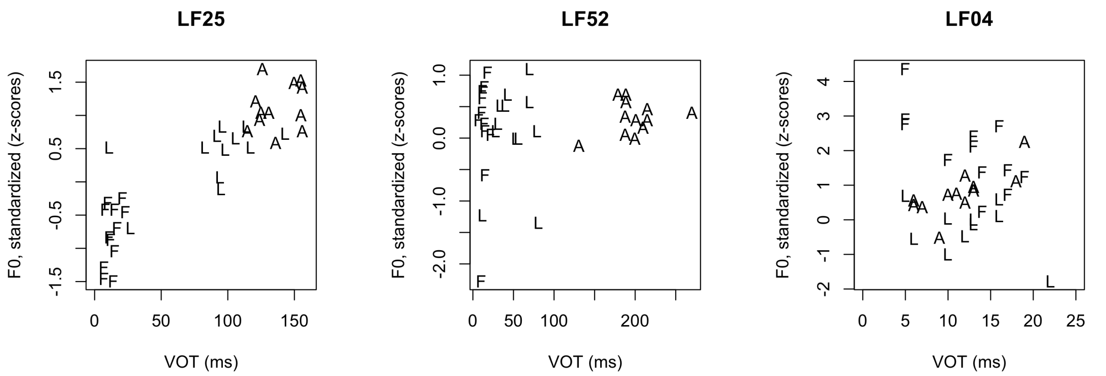
The phonetic spaces of L2 learners look markedly different. One of the most common patterns is found in Groups A ($n$=7) and B ($n$=2), where learners essentially produce two two-way contrasts, each in one dimension. In subgroup A1, lenis and fortis stops are both produced with short VOT and are contrasted on $f_0$, while fortis and aspirated stops tend to be produced with similar $f_0$ and are contrasted on VOT (cf. Figure 2, LM23). Subgroup A2 is similar, except that aspirated stops are produced with a relatively low $f_0$ onset in the range of the lenis stops rather than an elevated $f_0$ onset in the range of the fortis stops (cf. Figure 2, LF54). Subgroup A3 resembles subgroup A2, except lenis and fortis stops are reversed in the $f_0$ dimension: lenis stops are produced with higher $f_0$ than fortis stops, though lenis and aspirated stops are still produced in the same $f_0$ range. In Group B, fortis and lenis stops are produced in the same $f_0$ range and are distinguished on the basis of VOT, while lenis and aspirated stops are produced in the same VOT range and are distinguished on the basis of $f_0$ (cf. Figure 2, LF24).

**Figure 2:** Representative scatter plots of Week 5 productions in learner groups A and B (L = lenis, F = fortis, A = aspirated).




In Group C ($n$=7), learners produce a three-way contrast using either VOT, $f_0$, or both dimensions. The learners in subgroup C1 (e.g. LF25, cf. Figure 3) make use of both VOT and $f_0$ to make the contrast, producing fortis stops with short VOT and low $f_0$, lenis stops with longer VOT and higher $f_0$, and aspirated stops with the longest VOT and highest $f_0$. However, the learners in subgroup C2 (e.g. LF52, cf. Figure 3)—much like the learners described in Kim and Lotto (2002) and Shin (2007)—rely just on VOT to make a three-way contrast. In contrast, the learner in subgroup C3 (LF04, cf. Figure 3) relies almost entirely on $f_0$ to make the contrast, producing all three categories in the short-lag VOT range and distinguishing between them by producing lenis stops with the lowest $f_0$, aspirated stops with intermediate $f_0$, and fortis stops with the highest $f_0$.

**Figure 3:** Representative scatter plots of Week 5 productions in learner group C (L = lenis, F = fortis, A = aspirated).




Finally, a minority of learners fail to keep the three categories apart with these cues. In Group D ($n$=8), learners just produce a two-way contrast, showing nearly all possible types of merger, while in Group E

($n$=2), learners do not keep any of these categories distinct from the others in terms of VOT and/or $f_0$ onset, producing all of them over the same wide phonetic space.

## 4. DISCUSSION AND CONCLUSIONS

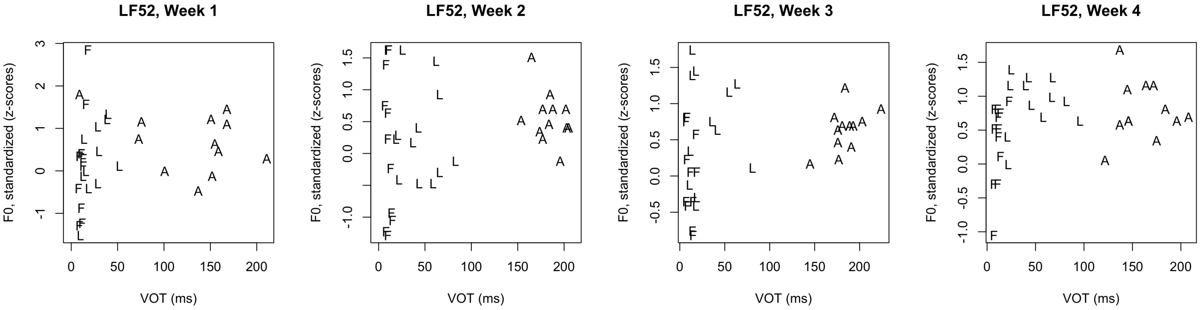
As seen above, there is wide variation in learners' success at restructuring the L1 phonetic space of two laryngeal categories into an L2 phonetic space of three laryngeal categories resembling native Korean. Some learners fail to produce a three-way contrast, merging two or more categories with different degrees of overlap, but the majority of learners do manage to produce three distinct categories. In addition, there is a dichotomy in the phonetic spaces of learners who produce a three-way contrast (the "full distinguishers") and those who only produce a two-way contrast (the "partial distinguishers"). In both groups, some learners appear to identify lenis stops as a category similar to voiced stops—full distinguishers separating fortis stops from lenis stops on the basis of $f_0$ onset, and partial distinguishers combining fortis and aspirated stops into a category similar to voiceless stops. However, in both groups there are other learners who identify fortis stops as the voiced-like category. Here the full distinguishers separate lenis stops from fortis stops on the basis of VOT and/or $f_0$, while the partial distinguishers combine lenis and aspirated stops into a voiceless-like category. These findings are consistent with the predictions of the ambiguous cross-linguistic category correspondences described above. Despite having the same L1 background, learners interpret this L2 contrast in multiple ways, resulting in disparate phonetic spaces of the L2 contrast that all depart in one or more ways from the native phonetic space.

We are left then to wonder: why is there so much variation? If we ignore the influence of affective variables, which, as suggested by background questionnaires, do not differ across the groups delineated above in any clear way, we are left with three possible explanations for the variation in learner production.

First, variation in production may be attributable to variation in input. After all, learners had different teachers, and there are some differences among the teachers in production, though the general pattern is the same (cf. Figure 1). Inspection of differences among learners along with their class affiliations does not support this hypothesis, however. For example, learners LM23 and LF54 were in the same class, yet still differ from each other: LM23 produces aspirated stops with high $f_0$, while LF54 produces them with low $f_0$ (cf. Figure 2), even though their teachers both produce them with high $f_0$. These facts indicate that even if some inter-learner variation is rooted in input disparities, input cannot be the whole story.

Second, there could be differences across participants with respect to how VOT and $f_0$ are weighted in distinguishing English voiced and voiceless stops. This variability in cue weighting could lead to variation in L2 production, in that learners would not necessarily be biased towards the same schemas of L1-L2 equivalence classifications. The fact that there is some variability among English speakers with respect to how sensitive they are to $f_0$ as a cue to the English voicing contrast (cf. Haggard *et al.* 1970) is consistent with this explanation—an interesting possibility that should be tested more thoroughly.

**Figure 4:** Scatter plots of learner LF52's productions in Weeks 1–4 (L = lenis, F = fortis, A = aspirated).




Third, learners might utilize explicit strategies to achieve L2 contrast that may or may not be based on actual L2 input patterns (such strategies being likely to differ between individuals). In fact, strategy does seem to account for what at least some learners do. For instance, learner LF52 (who produces a three-way contrast in the VOT dimension only, cf. Figure 3) expressed in study debriefings that she thought the contrast just had to do with aspiration, and so she ignored pitch. This sort of strategic bias largely accounts for why she started producing a three-way VOT contrast in Week 1 of the language program and continued to do so through Week 5, failing to make significant use of $f_0$ at all time points in this study (cf. Figures 3–4).

One noteworthy aspect of this production study is that the results differ substantially from those of the perception studies described above. Relatively few learners produce the L2 laryngeal categories with a phonetic space that might be predicted from cross-linguistic perception results or with one resembling that of native speakers. Moreover, there is a large amount of variation in learners' phonetic spaces, in contrast to the high degree of consistency seen in the perceptual performance of listeners in Schmidt (2007). This variation in L2 production spaces suggests that a number of factors are at work in the acquisition of L2 speech that are not necessarily seen in naïve non-native perception of an L2. Some possible sources of this variation have been discussed here, though much more work is needed to tease apart their effects.

While learners show a high degree of variation in the organization of their L2 phonetic spaces, what is consistent among them is that, with few exceptions, the production pattern they show in Week 5 is largely the same as the one they show in Week 1, suggesting that initial L1-L2 equivalence classifications tend to persevere, rather than change over the course of acquisition. The implication for L2 learning is clear: building an accurate representation of an L2 sound early in acquisition is crucial, since changing this representation significantly may become increasingly difficult later on.

## 5. REFERENCES


Boersma, P., Weenink, D. 2008. Praat: Doing phonetics by computer. <http://www.praat.org>. Last access 15 Jun 2008.

Caramazza, A., Yeni-Komshian, G., Zurif, E., Carbone, E. 1973. The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *Journal of the Acoustical Society of America* 54(2). 421–428.

Chang, C. In press. The implementation of laryngeal contrast in Korean as a Second Language. In: Kuno, S. et al. (eds.), *Harvard Studies in Korean Linguistics XIII*. Seoul: Hanshin Publishing Company.

Flege, J. 1987. The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15(1). 47–65.

Flege, J., Eefting, W. 1988. Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation. *Journal of the Acoustical Society of America* 83(2). 729–740.

Flege, J., Munro, M., MacKay, I. 1995. Effects of age of second-language learning on the production of English consonants. *Speech Communication* 16(1). 1–26.

Forster, J. 2008. DMDX. <http://www.u.arizona.edu/~kforster/dmdx/dmdx.htm>. Last access 15 Jun 2008.

Francis, A., Nusbaum, H. 2002. Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance* 28(2). 349–366.

Haggard, M., Ambler, S., Callow, M. 1970. Pitch as a voicing cue. *Journal of the Acoustical Society of America* 47(2B). 613–617.

Han, M., Weitzman, R. 1970. Acoustic features of Korean /P, T, K/, /p, t, k/, and /$p^h$, $t^h$, $k^h$/. *Phonetica* 22(2). 112–128.

Jun, S.-A. 2006. Phonological development of Korean: A case study. In: Kuno, S., Lee, I.-H., Whitman, J., Maling, J., Kang, Y.-S., Sells, P., Sohn, H.-S. (eds.), *Harvard Studies in Korean Linguistics XI*. Seoul: Hanshin Publishing Company, 17–33.

Kim, M. 2004. Correlation between VOT and F0 in the perception of Korean stops and affricates. *INTERSPEECH-2004*. 49–52.

Kim, M.-R., Lotto, A. 2002. An investigation of acoustic characteristics of Korean stops produced by non-heritage learners. In: Ree, J. (ed.), *The Korean Language in America*, Vol. 7. American Association of Teachers of Korean, 177–188.

Lee, S., Iverson, G. 2008. Development of stop consonants in Korean. *Korean Linguistics* 14. 21–39.

Major, R. 1996. L2 acquisition, L1 loss, and the critical period hypothesis. In: James, A., Leather, J. (eds.), *Second-Language Speech: Structure and Process*. Berlin: Mouton de Gruyter, 147–159.

Schmidt, A. 2007. Cross-language consonant identification: English and Korean. In: Bohn, O.-S., Munro, M. (eds.), *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*. Amsterdam: John Benjamins, 185–200.

Shin, E. 2007. How do non-heritage students learn to make the three-way contrast of Korean stops? In: Wang, H.-S. (ed.), *The Korean Language in America*, Vol. 12. American Association of Teachers of Korean, 85–105.


NOTES


* This research is reported at greater length in Chang (in press). The author gratefully acknowledges the financial support of the National Science Foundation, the UC Berkeley Center for Korean Studies, and the UC Berkeley Department of Linguistics; the helpful comments of Keith Johnson, the UC Berkeley Phonology Lab, and the audience at HISOKL 2009; the logistical assistance of the Fulbright Korean-American Educational Commission; and the dedication of all participants who completed the study.