

Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies^{a)}

Kyoung-Ho Kang^{b)} and Susan G. Guion

Department of Linguistics, University of Oregon, Eugene, Oregon 97403-1290

(Received 28 October 2007; revised 25 August 2008; accepted 26 August 2008)

The proposal that phonological contrast is enhanced through greater approximation of phonetic targets was investigated by comparing clear speech to conversational and citation-form speech produced by Korean speakers. The stop system of Korean is undergoing a sound change in which younger speakers produce the aspirated-lenis stop contrast differently from previous generations. Older speakers differentiate this contrast primarily with the acoustic correlate of voice onset time (VOT) and secondarily with *F0*. Younger speakers are merging VOT values for this contrast. As a result, the primary acoustic correlate is now *F0* for younger speakers. These production differences likely indicate that younger speakers have developed different phonetic targets for stop production. These different phonetic targets were predicted to result in different enhancement patterns in clear speech: Younger speakers were predicted to enhance *F0* differences, whereas older speakers were predicted to enhance VOT differences in clear speech. Results indicated that the older group solely used VOT to enhance the contrast in clear speech, whereas the younger group primarily used *F0* but also demonstrated small VOT enhancement. These results indicate that clear speech enhancement strategies reflect phonetic targets. Older and younger speakers have different *F0* and VOT targets and these different targets conditioned different enhancement strategies.

© 2008 Acoustical Society of America. [DOI: 10.1121/1.2988292]

PACS number(s): 43.70.Fq, 43.70.Bk [BHS]

Pages: 3909–3917

I. INTRODUCTION

The purpose of this study was to investigate the proposal that distinctiveness of phonological categories is enhanced in clear speech through the greater approximation of phonetic targets (Johnson *et al.*, 1993; Lindblom, 1990). For this purpose, the enhancement of acoustic correlates to Korean stop contrasts was examined in hyperarticulated clear speech compared with production in more casual speaking styles. We investigated whether there was a difference in the clear speech enhancement strategies between two groups of Korean speakers, one of which had undergone a diachronic change in the Korean stop system (Silva *et al.*, 2004; Silva, 2006; Wright, 2007). As a result of the sound change, younger speakers are thought to have different phonetic targets than older speakers for Korean aspirated and lenis stops. These different phonetic targets are predicted to result in different enhancement patterns in clear speech production between the two groups.

Talkers tend to adopt a speaking style, termed clear speech, to minimize perceptual confusions when they estimate potential speech perception difficulty on the listener's part. Clear speech is an intelligibility-enhancing mode of speech production. The intelligibility improvement effect of clear speech has been found with normal-hearing listeners (Ferguson, 2004; Krause and Braid, 2004), hearing-impaired listeners, and listeners with limited experience in

the test language (Bradlow and Alexander, 2007; Bradlow and Bent, 2002; Ferguson and Kewley-Port, 2002). In addition to this perceptual "clear speech effect," clear speech production has the effect of acoustic-phonetic modification of speech signals. Distinctiveness between phonological categories is enhanced in clear speech compared with conversational speech. This phonological contrast enhancement in clear speech production has been found in single language studies (Ferguson and Kewley-Port, 2002; Picheny *et al.*, 1986) as well as in cross-linguistic studies (Bradlow, 2002; Smiljanic and Bradlow, 2005).

The H & H theory (Lindblom, 1990) provides an explanation for the acoustic-phonetic modifications in clear speech production. The theory hypothesizes that talkers actively adjust articulatory effort according to the perceived difficulty of intelligibility for the listener. When intelligibility could be degraded, talkers will adopt a speech mode called hyperspeech. By hypothesis, talkers use greater articulatory energy in an adaptive effort to increase intelligibility for listeners, and this includes the enhancement of distinctiveness between phonological contrasts. When talkers do not estimate a need to improve intelligibility, they will default to a more economical system-oriented production, which is termed hypspeech. Thus, speech signals vary along the continuum between the two extremes of hypo- and hyperspeech. Clear speech can be viewed as production along the hyperarticulated end of the continuum.

Hyperspeech production may be investigated by the examination of specific phonetic targets associated with phonological categories. These phonetic targets are envisioned as stable, but their attainment variable, depending on produc-

^{a)} Portions of this work were presented at the 153rd meeting of the Acoustical Society of America, Salt Lake City, UT, June 2007 and the 17th Japanese-Korean Linguistics Conference, Los Angeles, CA, November 2007.

^{b)} Electronic mail: kkang@uoregon.edu

tion constraints, such as articulatory effort, coarticulatory force, and speech rate (Johnson *et al.*, 1993; Moon and Lindblom, 1994). In this view, hyperarticulated clear speech production would approximate phonetic targets most closely. In contrast, hypoarticulation is a phonetic process in which the phonetic targets are underrealized. The implication of this theoretical view is that contrastive phonological categories would be maximally distinctive in hyperarticulated speech due to the greater approximation of phonetic targets.

The contrast enhancement effect of clear speech is supported by studies on various acoustic-phonetic features serving to differentiate phonological categories. First, acoustic correlates to vowel quality were found to be enhanced in clear speech. Vowel space expansion in clear speech has been reported in various languages: Spanish (Bradlow, 2002), English (Bradlow, 2002; Ferguson and Kewley-Port, 2002; Picheny *et al.*, 1986; Smiljanic and Bradlow, 2005), and Croatian (Smiljanic and Bradlow, 2005). Ferguson and Kewley-Port (2002) reported that formant movement over the vowel nucleus was significantly larger in clear speech than in conversational speech. In addition, Moon and Lindblom (1994) found that English front vowels were less coarticulated in clear speech compared to citation-form speech. These results suggest that clear speech production aims at greater approximation of vowel formant target frequencies.

Secondly, vowel duration differences have been found to be enhanced in clear speech in a way related to language-specific phonological structure. Smiljanic and Bradlow (2008) examined the effect of clear speech production on the duration of English tense and lax vowels, English vowels before voiced and voiceless coda stops, and Croatian short and long vowels. Durational analyses indicated a greater durational difference in clear speech for the Croatian short and long vowels and the English vowels before voiced and voiceless coda stops, but not for the English tense and lax vowels, suggesting an effect of language-specific phonological structure on durational enhancement. Additionally, when duration ratio (clear minus conversational divided by conversational) was considered, the proportional duration contrast for both the Croatian vowels and the English vowels before coda stops remained stable.

Finally, enhancement of stop contrasts has been investigated in different languages. However, the results provided so far are inconclusive. Languages differ in the acoustic-phonetic modification of voice onset time (VOT) for stops. In a comprehensive acoustic study of conversationally and clearly produced English sentences, Picheny *et al.* (1986) found that the voicing contrast for English stops was enhanced in clear speech. Significant VOT lengthening for word-initial voiceless stops, but not for voiced stops, was found in clear speech. In contrast, Ohala (1994) did not find an enhancement of the VOT difference between English voiceless and voiced stops in clear speech. Words produced in response to perceptual clarification requests did not show significant VOT enhancement. Recently, Smiljanic and Bradlow (2008) examined the voicing contrast between voiced and voiceless stops in English and Croatian that were produced in conversational and clear speech. For English, VOT for voiceless stops lengthened in clear speech while the VOT

for voiced stops remained unchanged. This voicing contrast enhancement occurred only in word-initial position, not in word-medial position. As for Croatian, in clear speech, VOT for prevoiced stops showed a nonsignificant trend toward greater negative values, while the VOT of voiceless stops remained unchanged. When a proportional measurement of VOT as a percentage of stop duration (i.e., closure + aspiration) was considered, the proportional VOT remained stable across the speaking styles.

The combined results from studies on segmental contrast enhancement in clear speech suggest that hyperarticulated clear speech production tends to enhance acoustic distance between phonological categories in a way responsive to the language-specific phonological structure. In the current study, the language-specific influence of phonological structure on clear speech production is investigated in Korean. The Korean stop system manifests a three-way manner contrast termed aspirated, lenis (or lax), and fortis (or tense). In utterance-initial position, the three stop types are all voiceless. Previous studies have shown that aspirated stops have the longest VOT values, fortis stops have the shortest values, while lenis stops have intermediate values, closer or farther from the aspirated stops depending on the study. Vowels following aspirated and lenis stops are breathier than the vowels following fortis stops (indicated by greater H1-H2 values). In addition, fundamental frequency (F_0) on the vowels following aspirated and fortis stops is higher than that of vowels following lenis stops (for previous studies describing VOT, H1-H2, and F_0 in Korean stops, see Ahn, 1999; Cho *et al.*, 2002; Hardcastle, 1973; Han and Weitzman, 1970; Kagaya, 1974; Kang and Guion, 2006; Kim, 1965; Kim, 1994; Lisker and Abramson, 1964; and Shimizu, 1996).

However, the acoustic correlates related to the aspirated and lenis contrast have undergone a sound change in speakers born after roughly 1965 (Silva, 2006). In early studies on Korean stops (Han and Weitzman, 1970; Kim, 1965; Lisker and Abramson, 1964), VOT values reported for aspirated stops were high (93–108 ms on average) and values for reported lenis stops were low (30–35 ms on average), so that VOT values for these two stop types did not overlap. As such, VOT was considered to be the primary acoustic correlate to the lenis-aspirated stop distinction and F_0 was considered a redundant correlate (Silva, 2006, p. 298). In contrast, recent studies, examining speakers born after about 1965, have reported nondistinctive VOT values between lenis and aspirated stops (Kang and Guion, 2006; Silva *et al.*, 2004; Silva, 2006; Wright, 2007). Silva *et al.* (2004, 2006) have proposed that Korean stops are undergoing a diachronic change and that VOT values for Korean aspirated stops have shortened over the past two generations. As a result, the distinction between lenis and aspirated stops, which was previously coded primarily by a VOT difference, has come to be coded primarily by a F_0 difference on the vowels following the stops. Thus, older (born before 1965) and younger speakers may have different phonetic targets associated with the aspirated and lenis stop contrast.

If younger and older speakers have different VOT and F_0 targets for the production of Korean aspirated and lenis stops, the H & H theory (Lindblom, 1990) predicts that this

difference in phonetic targets will lead to different enhancement strategies. As phonetic targets are thought to be more greatly approximated in clear speech, different phonetic targets are predicted to result in difference in the usage of VOT and F_0 between younger and older speakers in clear speech. Specifically, the question addressed here is whether younger and older speakers use different strategies to produce Korean aspirated and lenis stops in clear speech. Younger speakers are predicted to enhance F_0 differences in clear speech, whereas older speakers are predicted to enhance VOT differences in clear speech.

II. EXPERIMENT 1

A. Participants

The data were collected from two groups of 11 native speakers of Korean. Each group had five male and six female speakers. The age in the “younger” group ranged from 20 to 29 years old and the participants were all born after 1977 (mean=25.8 years). Age ranged from 40 to 60 years old for the “older” group and the participants were all born before 1966 (mean=46.7 years). The speakers were affiliates of the University of Oregon, either foreign students, visiting scholars, or their family members. Self-reported daily use of English was low, especially with older speakers (15% for the older group and 34% for the younger group). None of the participants reported any hearing or speech disorders. All were speakers of the Seoul dialect of Korean.

B. Speech stimuli

The target words used for recordings were prepared to elicit the Korean stops in different speaking styles. For each of the nine Korean stops (aspirated, lenis, and fortis stops at three places of articulation), a pair of target words was selected for a total of 18 target words. Each of the 18 words had an initial CVC syllable, where the onset C was one of the nine Korean stops and V was the low vowel /a/ (see Appendix).

C. Procedure

To investigate potential contrast enhancement in clear speech production among the stop types, recordings were made in three different speaking styles: conversational, citation-form, and clear speech. Unlike previous studies on clear speech, which employed citation-form and clear speech only, the present study included a production task that aimed to elicit more natural “conversational” speech signals as well. Since eliciting true conversational speech in an experimental situation is difficult, this study assumes that the elicited conversational speech should be “more conversational” than citation-form speech even though it is not literally conversational. It is expected that the conversational speech will differ more when compared to clear speech than citation-form speech.

As the target words were repeated using different speaking styles, it was crucial to keep the target words in comparable prosodic and discourse environments. The target words were produced in utterance-initial position in all of the

speaking styles. Additionally, all words were in a focused context. In conversational speech, they were focused as the topic of conversation; in the citation-form condition, they were focused as new information in a carrier phrase that was repeated across the trials; and in the clear speech condition, they were focused as new or contrastive information in an isolated context. Thus, prosodic and discourse environments were controlled, allowing for the unconfounded manipulation of the instruction to speak clearly.

Overall, a total of 2376 tokens were obtained, 792 tokens (18 target words \times 2 repetitions \times 22 speakers) for each of the three speech styles. Interaction between the experimenter (the first author of this paper) and the speakers was conducted exclusively in Korean. The recordings were made separately for each speaker in a sound-attenuated booth using a high quality head-mounted microphone (Shure SM 10A) and a Marantz digital recorder (PMD 670). The utterances were digitally recorded at a sampling rate of 22 500 Hz and saved as wave files in a personal computer.

1. Conversational speech

Speech produced in a conversational style was elicited by asking speakers to explain the meaning of the target words or elaborate on the contexts where the target words are commonly used. The target words were real words whose meaning could not be easily explained in one or two words. The aim of choosing real word stimuli with complex definitions was to orient the speakers’ attention toward exploring the word meaning. This task was performed first, and speakers were not told about other tasks before beginning. This was to prevent the speakers from noticing the links between the production tasks in different speaking styles.

Each of the 18 target words was presented individually on flash cards using Korean orthography. The target words were presented to each speaker in a random order. The speakers were allowed enough time to explore the meaning of the words. While speakers were giving explanations, the experimenter responded to the speaker by agreeing, disagreeing, or giving his own opinion to the speaker’s explanations. This was to make the interaction more conversational.

To make sure that the target words were included in the explanations speakers gave, the speakers were asked to indicate which word they were explaining. The explanations they gave were typically constructed as Korean sentences beginning /__ +nin.../, or /__, ig Δ +nin.../, meaning “__ +TOPIC MARKER...,” or “__, this +TOPIC MARKER...” The first two productions of a given target word that occurred in the utterance-initial position were submitted to analysis.

2. Citation-form speech

After the task for conversational speech ended, citation-form speech was recorded. The 18 target words were presented in a random order on 18 flash cards two times each. The speakers produced each target word in a frame sentence, /__ ha.se.yo/. In English this translates to “Say__.” Speakers were instructed to produce the sentences at a comfortable speech rate and loudness level. When producing the words, the speakers were asked to read one card at a time and hand

it back to the experimenter. This was to prevent the speakers from reading too fast as is commonly observed when words are read in a list. Thus, they had to pause before reading the next flash card.

3. Clear speech

In order to elicit clearly produced speech, a situation in which listener-oriented careful speech is required was set up. The speakers were asked to imagine that they were teaching Korean to second language learners of Korean and reading out Korean words to their students. In order to assist the speaker reading the different target words clearly, each of the 18 target words was presented with another target word that shared a place of articulation but differed in the type of the initial stop (e.g., 탄탄하다 /^han.t^han.ha.ta/ and 단단하다 /tan.tan.ha.ta/; 단단하다 /tan.tan.ha.ta/ and 똥똥하다 /^t*an.t^{*}an.ha.ta/; 똥똥하다 /^t*an.t^{*}an.ha.ta/ and 탄탄하다 /^han.t^han.ha.ta/). So, each flash card had two of the 18 target words, but only the production of the first word of each pair was submitted to the acoustic measurements. Speakers were instructed to read the target words as clearly as possible and no other instructions, such as to read loudly or slowly, were given to the speakers. In this manner, two productions of each target word were elicited.

D. Measurements

For each of the tokens obtained from the production tasks, three acoustic properties encoding Korean stop contrasts were measured using PRAAT (Boersma and Weenink, 2005). VOT was measured as the time duration between the point of stop burst release and the onset of the periodic portion of the waveform for the initial vowel of the target words. In other words, the VOT measure used here was a true measure of voice onset and did not include a period of nonmodal phonation (if present), following the method described by Lisker and Abramson (1964). Other approaches to measuring VOT have been used for Korean stops. See Wright, 2007, pp. 14–30 for a discussion. H1-H2 and F_0 were measured on the following vowel. H1-H2, which is the amplitude (dB) difference between the first (H1) and the second (H2) harmonic, was measured just after the first full glottal pulse of the waveform. The onset of the vowel in the waveform was determined by the onset of the first full glottal pulse of the vowel in conjunction with the second formant. The amplitude values were calculated using a narrowband fast Fourier transform spectrum (window length of 30 ms). F_0 was measured at the temporal midpoint of the vowel using the pitch tracking function in PRAAT. When the pitch line abruptly moved or was discontinued, F_0 was recorded by measuring the duration of the relevant period in milliseconds and dividing it by 1000.

E. Results

Several statistical tests were used to assess the predictions that the acoustic correlates encoding Korean stop contrasts would be enhanced in clear speech production and that the enhancement would be made differently by younger and older speakers. First, a multivariate mixed analysis of vari-

ance (MANOVA) was performed with the following factors: group (younger and older), speaking style (conversational, citation-form, and clear), and stop type (aspirated, lenis, and fortis), with the last two factors treated as repeated measures. Given that Korean stops contrast with one another in VOT, H1-H2, and F_0 , these three acoustic measures were all entered as dependent variables. The individual data points submitted to analysis were an average of the two repetitions of each test item for each speaking condition for each measure. The MANOVA returned significant main effects for group [$F(3,128)=5.34$, $p<0.001$], speaking style [$F(6,125)=54.357$, $p<0.001$], and stop type [$F(6,125)=893.413$, $p<0.001$]. Significant two-way and three-way interactions were found for all combinations of these three factors (F -values ranged from 6.21 to 19.207, $p<0.001$). These results indicate that the Korean stops were produced differently by the different groups, in different speaking styles. In order to further explore the interactions, the data were split by age group, and then the effects of speaking style and stop type on each of the three acoustic properties were investigated in separate repeated measure ANOVAs. Both subject (F_1) and item (F_2) analyses were run. Both factors were treated as repeated measures in the subject analysis, but only speaking style was treated as repeated in the item analysis. Due to the fairly large degrees of freedom, the most conservative statistic, the lower bound, is reported here, and the focus of the analysis is on the interaction of speaking style and stop type. Main effects are only reported in the absence of significant interactions. Our hypotheses revolved around the interactions, as we sought to determine the nature of the speaking style effect as modulated by stop type for each dependent measure.

1. Younger group

For the younger group, univariate tests for each of the dependent measures VOT, H1-H2, and F_0 revealed significant interactions for the factors speaking style and stop type for both subject and item analyses (VOT [$F_1(1,65)=10.220$, $p=0.002$, $F_2(2,15)=12.223$, $p=0.001$], H1-H2 [$F_1(1,65)=11.730$, $p=0.001$, $F_2(2,15)=10.091$, $p=0.002$], and F_0 [$F_1(1,65)=12.281$, $p=0.001$, $F_2(2,15)=26.496$, $p<0.001$]). Thus, the patterns found in the subject analyses are consistent across different test items. These interactions are seen in Fig. 1 and interpreted in the paragraphs below.

As shown in Fig. 1(a), the interaction for VOT is due to differences in the patterning for the stop types across the speech styles. VOT for aspirated stops slightly increased in citation-form and clear speech while that for lenis stops slightly decreased. As a result of this, the VOT distance for the aspirated and lenis stop contrast expanded in clear speech.

As for the H1-H2 measure, the interaction is due to the enhancement pattern of fortis stops [see Fig. 1(b)]. Vowels following fortis stops had a more negative H1-H2 value in clear speech, indicating greater creaky voice quality in this speech style. On the other hand, the other stop types showed minimal effects of speech style.

The interaction for F_0 is seen in Fig. 1(c). F_0 increased in clear speech production for all of the three stop types.

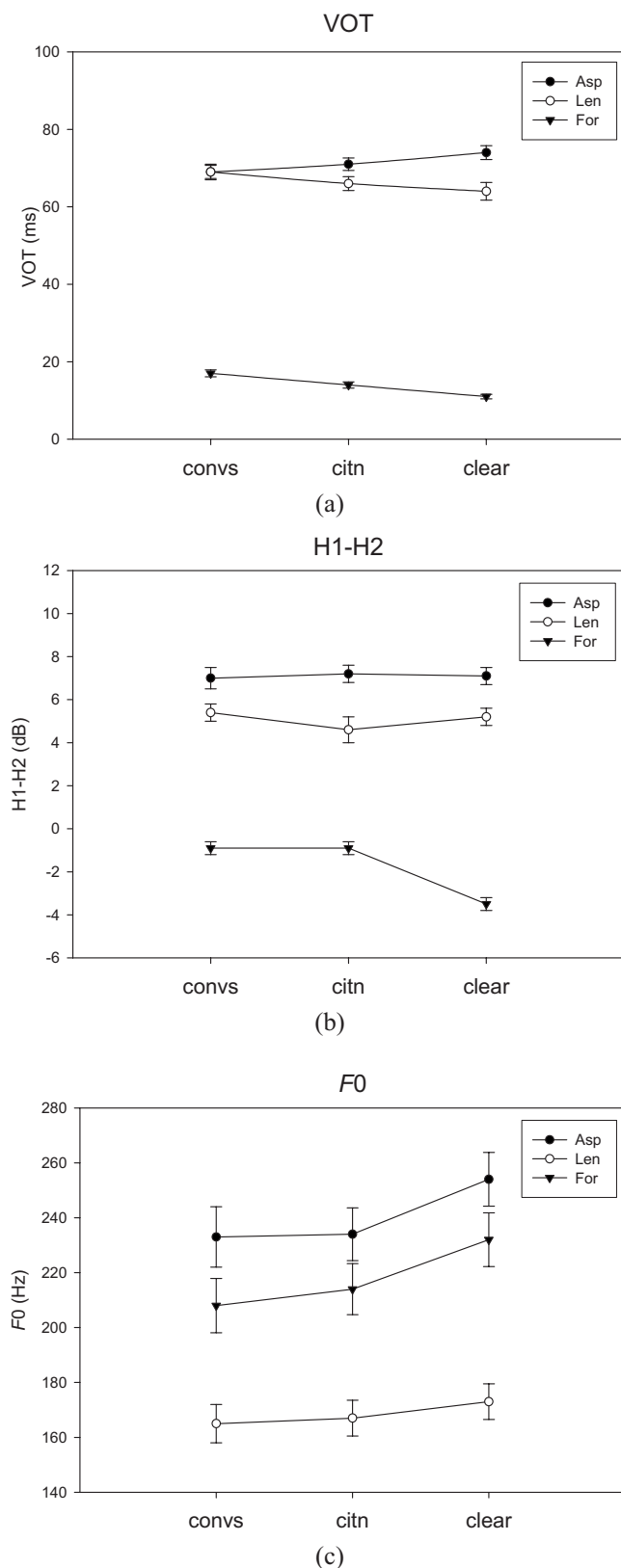


FIG. 1. Mean values with standard errors for the production of Korean stops (aspirated, lenis, and fortis) in conversation, citation-form, and clear speech styles by the younger group ($n=11$) for three acoustic correlates [(a) VOT, (b) H1-H2, and (c) F_0].

However, because the F_0 increase in clear speech for vowels following aspirated and fortis stops was relatively larger than that of lenis stops, the F_0 distance between lenis stops and the other two stop types was expanded in clear speech.

2. Older group

For the older group, univariate tests returned significant interactions of speaking style by stop type for each of VOT, H1-H2, and F_0 (VOT [$F_1(1,65)=14.717$, $p<0.001$, $F_2(2,15)=7.867$, $p=0.005$], H1-H2 [$F_1(1,65)=5.826$, $p=0.019$, $F_2(2,15)=5.850$, $p=0.013$], and F_0 [$F_1(1,65)=12.208$, $p=0.001$, $F_2(2,15)=19.223$, $p<0.001$]). Interactions between speaking style and stop type are displayed in Fig. 2.

Figure 2(a) shows that the VOT distance between aspirated and lenis stops was expanded in clear speech. Mean VOT for lenis stops decreased substantially, whereas that for aspirated stops was roughly the same between conversational and clear speech.

As for the H1-H2 measure [see Fig. 2(b)], the interaction was due to the substantial reductions found in citation-form and clear speech for fortis stops. These results indicate that fortis stops were produced with creakier voicing on the following vowels in these speaking styles, whereas the other two stop types showed little effect of speech style.

Finally, for the F_0 measure, the interaction was primarily reflected in the patterning in citation-form speech [see Fig. 2(c)]. The mean F_0 for aspirated and fortis stops decreased vis-a-vis the other two speaking styles. Also, in clear speech, mean F_0 for aspirated stops slightly decreased compared to conversational speech, whereas mean F_0 's for lenis and fortis stops increased. Consequently, the overall F_0 distance among the stop categories was slightly reduced in clear speech.

F. Discussion

The results of the analyses suggest that the phonological contrasts between the Korean stops are enhanced in clear speech production for both the younger and older speakers, but in a different manner. The younger speakers produced Korean aspirated and lenis stops with merged VOT values in the conversational condition. However, in clear speech, the two stops were differentiated with a small VOT difference of 10 ms. In addition, the H1-H2 difference between the fortis and the other two stop types was more exaggerated in clear speech, as the following vowels became even creakier in the context of fortis stops, but not in other stop contexts (the difference between aspirated and fortis stops in conversational speech was 7.9 dB, and 10.7 dB in clear speech). The F_0 difference between the lenis and the other two stop types was also enhanced in clear speech by a greater F_0 increase in the vowels following the aspirated and fortis stops than the lenis stops (the difference between aspirated and lenis stops in conversational speech was 68 Hz, and 81 Hz in clear speech).

For the older speakers, a VOT difference between aspirated and lenis stops was found in conversational speech, but the difference became even greater in clear speech (18 ms difference in conversational speech and 31 ms difference in clear speech). As for the H1-H2 measure, fortis stops and the other two stop types, aspirated and lenis, became more distinctive from each other in clear speech due to enhanced creaky voicing on the vowels following fortis stops (the dif-

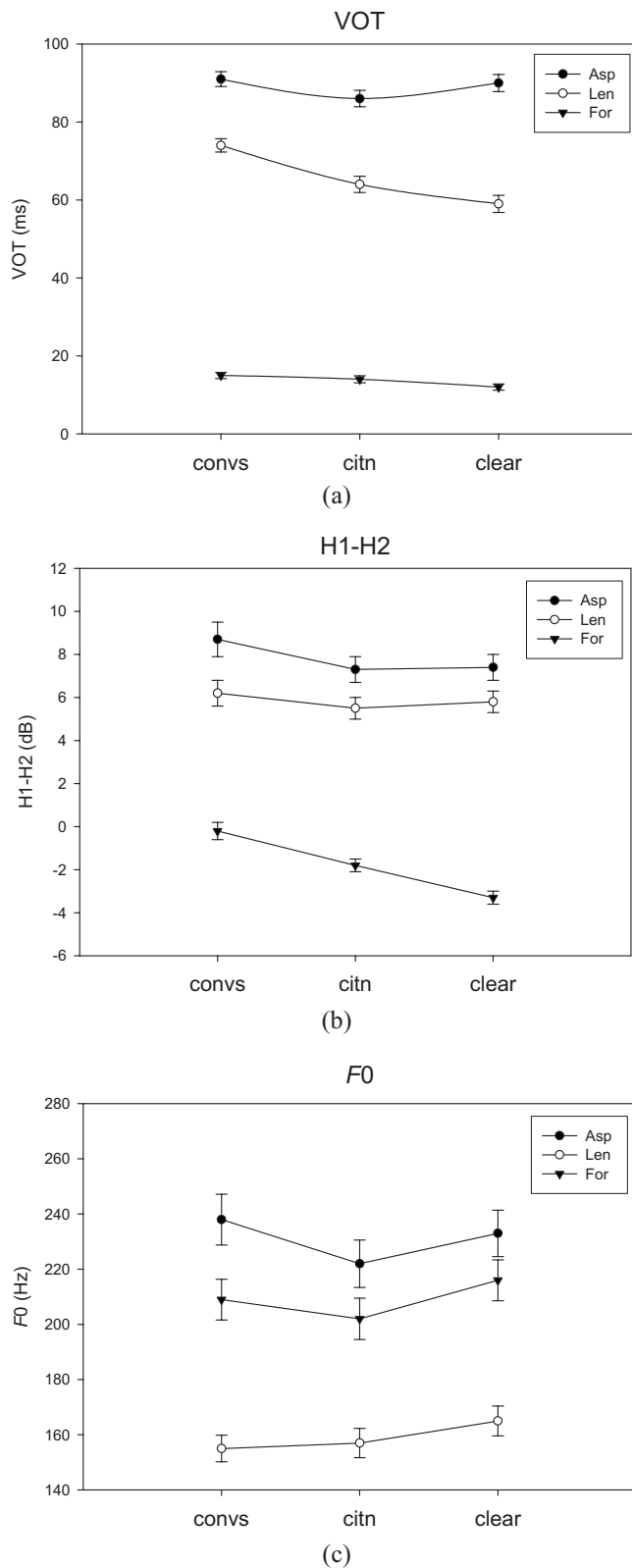


FIG. 2. Mean values with standard errors for the production of Korean stops (aspirated, lenis, and fortis) in conversation, citation-form, and clear speech styles by the older group ($n=11$) for three acoustic correlates [(a) VOT, (b) H1-H2, and (c) F_0].

ference between aspirated and fortis stops in conversational speech was 9 dB, and 10.7 dB in clear speech). Finally, in contrast to the case of the younger speakers, the F_0 difference between the stop categories reduced in clear speech

production (the difference between aspirated and lenis stops in conversational speech was 84 Hz, and 68 Hz in clear speech). As a result, F_0 did not enhance any phonological distinctions in clear speech production in the case of the older speakers.

To sum up, these results indicate that there is a difference in the use of VOT and F_0 to enhance the aspirated and lenis stop contrast between the two age groups. The older group used VOT, whereas the younger group used both VOT and F_0 , with F_0 being a stronger correlate. However, interpreting the difference in F_0 enhancement in clear speech between the two groups requires closer examination. One question that arises is whether the greater F_0 difference in clear speech for the younger speakers was induced by the enhancement of a phonetic target used to realize the phonological contrast of aspirated and lenis stops, or whether it was due to an overall greater F_0 use by the younger speakers in the clear speech style. That is, an alternative interpretation of the results may be that the younger speakers generally use F_0 more robustly than older speakers, and that the F_0 enhancement in clear speech is simply attributable to a larger F_0 range in clear speech.¹ In order to test this alternative possibility, another experiment was conducted with a separate set of test items containing nonstop consonants.

III. EXPERIMENT 2

Experiment 2 examines F_0 variation of aspirated consonants as a function of speaking style in comparison with lenis consonants. The purpose was to investigate the question of whether the F_0 enhancement of the younger speakers in experiment 1 was due to a phonological contrast enhancement between the aspirated and lenis stops or an expanded F_0 range in clear speech. Specifically, the phrase initial tonal contrast between the high-tone associated Korean aspirated and fortis stops and the low-tone associated lenis stops (Cho *et al.*, 2002; Jun, 1993) was investigated in comparison with other nonstop consonants showing a similar tonal specification phrase initially. Thus, a high-tone triggering voiceless fricative /h/ and a low-tone triggering voiced nasal /n/ were examined for their F_0 realization in citation-form and clear speech. If the larger F_0 difference in clear speech for the younger speakers were induced by a generally greater pitch enhancement regardless of segment type, the F_0 difference for the vowels following /h/ and /n/ would be greater in clear speech than in citation-form speech, similar to the pattern found for the aspirated and lenis stops. On the contrary, if the F_0 differentiation was part of the enhancement strategy for the phonological contrast of the aspirated and lenis stops, the F_0 difference between the vowels following /h/ and /n/ would not be enhanced in clear speech for either the younger or the older speakers.

A. Participants

Twenty native Korean speakers who did not participate in experiment 1 were recruited. The participants were grouped separately based on their age. The ten younger speakers were 19–29 years old (mean of 23 years) and the ten older ones were 40–58 years old (mean of 47 years). All

of the speakers reported that they were born in Seoul or Kyunggi province (i.e., the greater Seoul area) and used the Seoul or Kyunggi dialect of Korean.

B. Procedure

Each of the 20 speakers read 35 Korean words in two speaking styles: citation-form and clear speech. For the citation-form style, the speakers used their own comfortable speech rate and loudness level. For the clear style, they were instructed to read in a “clear” way, as if speaking to a “foreigner” audience who needs greater linguistic-phonetic resources to have full access to the linguistic information.

The carrier phrase, /__la.ko mal.ha.se.yo/, meaning “Say__+QUOTATIVE PARTICLE,” was used to ensure that the target words were produced in the same prosodic context in both speaking styles. Each word began with an aspirated or lenis consonant. The initial consonants examined in the analysis were /t^h, h/ for the aspirated series and /t, n/ for the lenis series. The words were 타다 /t^ha.ta/ ‘to ride’, 하다 /ha.ta/ ‘to do’, 다도 /ta.to/ “tea ceremony,” and 나비 /na.pi/ “butterfly.” Each word was produced three times in three separate blocks for each speaking style. The reading was recorded at a sampling rate of 22 050 Hz using a portable Marantz digital recorder (PMD 670) and a high quality headset microphone (Shure SM10A). The recording was made in a sound-attenuated booth or in a quiet room at the participants’ home. The recorded utterances were saved to a computer in a wave sound format. *F0* was measured on the first syllable of the utterance-initial target words. *F0* was obtained at the temporal midpoint of the vowel using the pitch analysis function of PRAAT (Boersma and Weenink, 2005).

C. Results

The *F0* measurements for the aspirated /t^h/ and /h/ and the lenis /t/ and /n/ consonants were submitted to a three-way repeated measures ANOVA. Speaking style (citation-form and clear speech), condition (stop /t^h, t/ and continuant /h, n/), and consonant type (aspirated /t^h, h/ and lenis/t, n/) were entered as factors. Mean *F0* values across three repetitions of each test item were used as the dependent measure.

Analyses were conducted separately for each of the younger and older speaker groups. The test for the younger speakers returned significant main effects for speaking style [$F(1, 9)=9.284, p=0.014$], condition [$F(1, 9)=29.269, p < 0.001$], and consonant type [$F(1, 9)=85.582, p < 0.001$]. A significant interaction was found for speaking style by consonant type [$F(1, 9)=5.471, p=0.044$] and for speaking style by condition by consonant type [$F(1, 9)=25.081, p=0.001$]. These results indicate that the effect of speaking style on the *F0* difference between aspirated and lenis stop stops, /t^h/ and /t/, was different from that found for the fricative and the nasal, /h/ and /n/. As shown in Table I, the *F0* difference for the aspirated and lenis stop contrast increased in clear speech (63 Hz versus 76 Hz) due to a higher *F0* for aspirated stops, whereas the *F0* difference for the fricative and nasal contrast largely stayed the same (69 Hz versus 72 Hz) between citation-form and clear speech.

TABLE I. Mean *F0* with standard error in parentheses.

	Younger (<i>n</i> =10)		Older (<i>n</i> =10)	
	Citation form	Clear	Citation form	Clear
/t ^h /	231 (20.2)	244 (20.7)	227 (20.8)	235 (22.7)
/t/	168 (14.8)	168 (13.7)	162 (13.9)	168 (14.8)
/h/	228 (20.2)	236 (20.6)	224 (20.0)	234 (22.3)
/n/	159 (14.0)	164 (13.7)	156 (12.5)	165 (13.9)

In contrast, in the case of the older speakers, the main effect of consonant type (aspirated /t^h, h/ and lenis /t, n/) was the only significant finding [$F(1, 9)=59.077, p < 0.001$]. None of the other main effects or interactions were significant ($p > 0.05$). The significant effect of consonant type is due to the large *F0* difference between aspirated (/t^h, h/) and lenis (/t, n/) consonant types. The results indicate that the older speakers did not differentially enhance *F0* for the aspirated and lenis stop contrast vis-a-vis that of the continuant contrast (/h/ and /n/) between the speaking styles. As shown in Table I, the mean *F0* differences for the stop contrast, as well as the continuant contrast, marginally increased in clear speech (stops: 65 Hz versus 67 Hz; continuant: 68 Hz versus 69 Hz).

D. Discussion

The results indicate that the greater *F0* difference in clear speech for the younger speakers is specific to the aspirated and lenis stops: *F0* difference substantially increased in clear speech for the stop contrast (/t^h/ versus /t/), but increased only slightly for the continuant contrast (/h/ versus /n/). Older speakers showed a similar difference in the two speaking styles for both contrast types. Combined with the results from experiment 1, these results suggest that the greater *F0* difference for the aspirated and lenis stops found for the younger speakers can be attributed to the greater approximation of the high *F0* target associated with aspirated stops in clear speech. The results were not consistent with the alternative hypothesis of an overall expanded *F0* range for the younger speakers.

IV. GENERAL DISCUSSION

Based on the H & H theory (Lindblom, 1990), we predicted that phonological contrasts would be enhanced in clear speech by more fully realizing phonetic targets in an effort to aid listeners in speech perception. Given the sound change in the Korean stop system (Silva *et al.*, 2004; Silva, 2006; Wright, 2007), younger and older Korean speakers were predicted to have different phonetic targets for stop production and, in turn, to display different enhancement patterns in clear speech production. Specifically, for the aspirated and lenis stop contrast, younger speakers were predicted to enhance *F0* differences, whereas older speakers were predicted to enhance VOT differences.

The results of the two experiments upheld these predictions. The three-way manner contrast of Korean stops was found to be enhanced in clear speech compared with conversational or citation-form speech. Furthermore, there was a

difference in the use of VOT and *F0* between the younger and older speakers. The older group used VOT in order to enhance the aspirated and lenis contrast in clear speech. The younger group enhanced the difference between aspirated and lenis stops by increasing the relative *F0* of aspirated stops in clear speech. VOT also slightly diverged for the two stop types in clear speech but to a much lesser degree than the older speakers.

This small VOT usage by younger speakers merits further discussion. One possibility is that the small VOT usage may reflect a small difference in VOT targets between aspirated and lenis stops that is only revealed in clear speech. Phonetic targets may have shifted such that *F0* is used as a primary acoustic correlate and VOT as a secondary one by younger speakers. In other words, the small VOT usage for the younger speakers may reflect a decreased, yet residual, role of VOT in the distinction of aspirated and lenis stops. Another possibility is that the small VOT usage may not reflect differences in VOT targets between the aspirated and lenis stops. In other words, the stops have the same or quite similar VOT targets. The small VOT dispersion in clear speech may rather reflect an attempt to use acoustic correlates heard in the speech of others. That is, exposure to older speakers, who maintain VOT differences, has resulted in the partial ability to manipulate VOT, as a type of style shifting or mimicking of a more traditional speaking style that employs VOT.

Another finding from this study was that the older group enhanced the VOT difference by lowering the VOT for lenis stops and keeping the VOT for aspirated stops relatively stable across the speaking styles. One may wonder why the VOT for the aspirated stops did not also increase. We believe an explanation may lie in the prosodic environment in which the stops were produced. In studies investigating varying realizations of VOT of Korean stops as a function of different prosodic contexts, Cho and Jun (2000) and Cho and Keating (2001) reported that the VOT values for aspirated and lenis stops in higher domain-initial positions (such as intonational phrase or utterance initial) was greater than in lower domain-initial positions (such as accentual phrase initial). As stimuli were elicited in the utterance-initial position in the current study, we would expect high VOT values for both aspirated and lenis stops due to the domain-initial strengthening effect.² In the case of aspirated stops, these high VOT values may be close to the phonetic target, but in the case of lenis stops, these high VOT values may overshoot the phonetic target. Thus, in clear speech production, the lower VOT target for lenis stops may be more closely approximated, resulting in lower VOT values for clear speech. On the other hand, if the utterance-initial position had already conditioned targetlike values for the aspirated stops, an increased VOT in clear speech would not be expected.

The idea that lenis stops have a relatively low VOT target is supported by previous investigations. Early studies report VOT values for lenis stops in the range 21–41 ms, sometimes even overlapping with fortis stops (Han and Weitzman, 1970; Kim, 1965; and Lisker and Abramson, 1964). More recent studies with speakers born in the 1960s

or earlier report VOT values from approximately 5 to 40 ms for lenis stops (Cho and Keating, 2001; Wright, 2007).

Now let us address questions about the nature of the Korean stop sound change. One possibility is that the sound change is incremental and, thus, related to speaker's age in a continuous fashion. Another possibility is that the sound change has occurred in the younger speakers but not in the older speakers and, thus, evidence for the change will only be found in the younger group. An examination of individual production revealed no appreciable *F0* enhancement by any speakers in the older group, suggesting that they have not been affected by the sound change. In the case of the younger group, there was no apparent relationship between age (speakers ranged from 20 to 30 years old) and the amount of *F0* or VOT enhancement. However, a sample larger than 21 younger speakers in the two experiments may be needed to detect such a gradient relationship. Additionally, data from speakers in the 30–40 year age range would prove probative to the question. Given these findings, we tentatively propose that speakers of the Seoul dialect born in the 1970s or later have, as a whole, undergone the sound change and that speakers before this have not been affected by it.

The H & H theory (Lindblom, 1990) hypothesizes that talkers make articulatory modifications in response to the estimated perceptual needs of the listeners, and such modifications are predicted to result in enhancement of phonological contrasts. The current study presented an instance of stop contrast enhancement in clear speech in Korean, lending typological support to the language universal predictions of the H & H theory. What remains for further research is to experimentally test whether such articulatory modifications result in the improvement of intelligibility for listeners. To this end, a perception study testing the intelligibility improvement effect of clear speech is required. As the Korean stop system is experiencing changes in the acoustic correlates to the stops, differences are predicted between older and younger listeners in the extent to which each acoustic correlate contributes to intelligibility improvement. A perception experiment that individually manipulates the acoustic correlates is in preparation for a subsequent study.

ACKNOWLEDGMENTS

The authors thank Melissa Redford, Mieko Ueno, and three anonymous reviewers for their helpful comments on earlier versions of this manuscript.

APPENDIX

Korean words used in the production tasks for conversational, citation-form, and clear speech. The consonant-vowel sequences submitted to acoustic measurement are bolded.

(1) Aspirated stop series

판판하다	/p ^h an.p ^h an.ha.ta/,	“to be even/to be flat”
팔팔하다	/p ^h al.p ^h al.ha.ta/,	“to be lively”
탄탄하다	/t ^h an.t ^h an.ha.ta/,	“to be solid/to be firm”
탕감하다	/t ^h ag.kam.ha.ta/,	“to write off a debt”

칼칼하다	/k ^h al.k ^h al.ha.ta,	“to be spicy (food)/ to have a scratchy throat”
깜깜하다	/k ^h am.k ^h am.ha.ta/,	“to be dark/to be gloomy”
(2) Lenis stop series		
반반하다	/pan.pan.ha.ta/,	“to be comely/to be pretty”
발발하다	/pal.pal.ha.ta/,	“to break out”
단단하다	/tan.tan.ha.ta/,	“to be hard/to be strong”
당당하다	/taŋ.taŋ.ha.ta/,	“to be dignified/to be imposing”
간간하다	/kan.kan.ha.ta/,	“to be nicely salted”
깜깜하다	/kam.kam.ha.ta/,	“to forget entirely/to have no news”
(3) Fortis stop series		
반짝하다	/p [*] an.c [*] ak.ha.ta/,	“to shine/to have ephemeral fame”
뽕뽕하다	/p [*] aŋ.p [*] aŋ.ha.ta/,	“to be packed/to have good condition (per- son)”
똥똥하다	/t [*] an.t [*] an.ha.ta/,	“to be strong/to be hard”
땡땡하다	/t [*] aŋ.t [*] aŋ.ha.ta/,	“to be short and chubby”
깐깐하다	/k [*] an.k [*] an.ha.ta/,	“to be fastidious/to be strict”
깜깜하다	/k [*] am.k [*] am.ha.ta/,	“to be pitch dark/to be ignorant”

¹We thank an anonymous reviewer for bringing this possibility to our attention.

²The VOT strengthening effect in higher domain-initial positions has been found in several languages (Hsu and Jun, 1998, in Taiwanese; Jun, 1993, in Korean; and Pierrehumbert and Talkin, 1992, in English), suggesting that it may be a general articulatory process.

Ahn, H. (1999). “Post-release phonatory processes in English and Korean: Acoustic correlates and implications for Korean phonology,” Ph.D. thesis, University of Texas at Austin, Austin, TX.

Boersma, P., and Weenink, P. (2005). “PRAAT: Doing phonetics by computer (Version 4.4) [computer program],” from <http://www.Praat.org/> (Last viewed December, 2005).

Bradlow, A. R. (2002). “Confluent talker- and listener-related forces in clear speech production,” in *Laboratory Phonology 7*, edited by C. Gussenhoven and N. Warner (Mouton de Gruyter, Berlin), pp. 241–272.

Bradlow, A. R., and Alexander, J. A. (2007). “Semantic and phonetic enhancement for speech-in-noise recognition by native and non-native listeners,” *J. Acoust. Soc. Am.* **112**, 2339–2349.

Bradlow, A. R., and Bent, T. (2002). “The clear speech effect for non-native listeners,” *J. Acoust. Soc. Am.* **112**, 272–284.

Cho, T., and Jun, S.-A. (2000). “Domain-initial strengthening as enhancement of laryngeal features: Aerodynamic evidence from Korean,” in *Chicago Linguistic Society 36*, edited by J. Boyle, J.-H. Lee, and A. Okrent

(Department of Linguistics, University of Chicago, Chicago, IL), pp. 31–44.

Cho, T., and Keating, P. A. (2001). “Articulatory and acoustic studies on domain-initial strengthening in Korean,” *J. Phonetics* **29**, 155–190.

Cho, T., Jun, S.-A., and Ladefoged, P. (2002). “Acoustic and aerodynamic correlates of Korean stops and fricatives,” *J. Phonetics* **30**, 193–228.

Ferguson, S. (2004). “Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners,” *J. Acoust. Soc. Am.* **116**, 2365–2373.

Ferguson, S., and Kewley-Port, D. (2002). “Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners,” *J. Acoust. Soc. Am.* **112**, 259–271.

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

Hardcastle, W. J. (1973). “Some observations on the tense-lax distinction in initial stops in Korean,” *J. Phonetics* **1**, 263–272.

Hsu, C.-S., and Jun, S.-A. (1998). “Prosodic strengthening in Taiwanese: Syntagmatic or paradigmatic?,” *UCLA working papers in phonetics* **96**, 69–89.

Johnson, K., Flemming, E., and Wright, R. (1993). “The hyperspace effect: Phonetic targets are hyperarticulated,” *Language* **69**, 505–528.

Jun, S.-A. (1993). “The phonetics and phonology of Korean prosody,” Ph.D. thesis, Ohio State University, Columbus, OH.

Kagaya, R. (1974). “A fiberoptic and acoustic study of Korean stops, affricates, and fricatives,” *J. Phonetics* **2**, 161–180.

Kang, K.-H., and Guion, S. G. (2006). “Phonological systems in bilinguals: Age of learning effects on the stop consonant systems of Korean-English bilinguals,” *J. Acoust. Soc. Am.* **119**, 1672–1683.

Kim, C.-W. (1965). “On the autonomy of the tensify feature in stop classification (with special reference to Korean stops),” *Word* **21**, 339–359.

Kim, M.-R. C. (1994). “Acoustic characteristics of Korean stops and perception of English stop consonants,” Ph.D. thesis, University of Wisconsin at Madison, Madison, WI.

Krause, J., and Braida, L. (2004). “Acoustic properties of naturally produced clear speech at normal speaking rates,” *J. Acoust. Soc. Am.* **115**, 362–378.

Lindblom, B. (1990). “Explaining phonetic variation: A sketch of the H and H theory,” in *Speech Production and Speech Modeling*, edited by W. J. Hard-Castle and A. Marchal (Kluwer, Dordrecht), pp. 403–439.

Lisker, L., and Abramson, A. S. (1964). “A cross-language study of voicing in initial stops: Acoustical measurement,” *Word* **20**, 384–422.

Moon, S.-J., and Lindblom, B. (1994). “Interaction between duration, context, and speaking style in English stressed vowels,” *J. Acoust. Soc. Am.* **96**, 40–55.

Ohala, J. J. (1994). “Acoustic study of clear speech: A test of the contrastive hypothesis,” *International Symposium on Prosody*, Vol. **18**, pp. 75–89.

Picheny, M., Durlach, N., and Braida, L. (1986). “Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech,” *J. Speech Hear. Res.* **29**, 434–446.

Pierrehumbert, J., and Talkin, D. (1992). “Lention of /h/ and glottal stop,” in *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, edited by G. Docherty and D. R. Ladd (Cambridge University Press, Cambridge, UK), pp. 90–117.

Shimizu, K. (1996). *A Cross-Language Study of Voicing Contrasts of Stop Consonants in Asian Languages* (Seibido, Tokyo, Japan).

Silva, D. J. (2006). “Acoustic evidence for the emergence of tonal contrast in contemporary Korean,” *Phonology* **23**, 287–308.

Silva, D. J., Choi, Y.-J., and Kim, J. E. (2004). “Diachronic shift in VOT values for Korean stop consonants,” in *Harvard Studies in Korean Linguistics X.S*, edited by Susumo Kuno, Ik-Hwan Lee, John Whitman, Joan Maling, Young-se Kang, and Young-joo Kim (Harvard University, Department of Linguistics, Cambridge, MA), pp. 173–185.

Smiljanic, R., and Bradlow, A. R. (2005). “Production and perception of clear speech in Croatian and English,” *J. Acoust. Soc. Am.* **118**, 1677–1688.

Smiljanic, R., and Bradlow, A. R. (2008). “Stability of temporal contrasts across speaking styles in English and Croatian,” *J. Phonetics* **36**, 91–113.

Wright, J. D. (2007). “Laryngeal contrast in Seoul Korean,” Ph.D. thesis, University of Pennsylvania, Philadelphia, PA.