A CROSS-LINGUISTIC STUDY OF KOREAN LARYNGEAL STOPS BY THE NATIVE SPEAKERS OF CHINESE, ENGLISH, KOREAN, AND SPANISH

Seung-Eun Chang, Marjorie Burge & Younghun Choi

University of California, Berkeley, USA
sechang71@berkeley.edu; margiburge@berkeley.edu; younghun_choi@mfe.berkeley.edu

ABSTRACT

This paper provides the results of an acoustic study of Korean laryngeal stop production by eight adult native speakers each of Mandarin Chinese, American English, Seoul Korean, and Latin American Spanish. Three measurements were taken: *f0* (fundamental frequency), VOT (Voice Onset Time), and *H*1-*H*2 (the relative amplitude of the first versus the second harmonic of the voice source). The data indicate that while Chinese speakers are able to make use of *f0* and *H*1-*H*2 cues to create the three-way stop contrast in a similar way to Korean native speakers, native speakers of English and Spanish are primarily dependent on a VOT cue. The results suggest that L2 (second language acquisition) production is largely based on the phonetic structure of L1 (native language), and this confirms, quantitatively, what has been long claimed in the second language literature but previously usually by conjectures.

Keywords: Korean laryngeal stops, VOT, F0, H1-H2, a cross-linguistic study

1. INTRODUCTION

Korean is unusual among the major languages of East Asia in its consonant system. It includes a typologically rare distinction among three types of voiceless obstruents: plain (or lax), tense (or fortis), and aspirated. Although it has been a long-standing issue what phonetic dimension makes them distinctive, much research has focused on the same primary cue: Voice Onset Time (VOT).

It has recently been reported, however, that as a result of sound change, younger speakers may in fact have different phonetic targets than older speakers for aspirated and plain stops: the contrast between plain and aspirated stops is phonetically manifested in terms of differentiated tonal melodies [5, 13, 15]. Differences in VOT values most prominently differentiate aspirated stops from tense and plain stops, while differences in fundamental frequency (*f0*) serve to differentiate plain stops from aspirated and tense stops. On average, VOT is shortest for Korean tense stops, longer for plain, and longest for aspirated. By contrast, *f0* is lowest for vowels following plain stops, higher following tensed stops, and highest following aspirated [5, 6, 10, 13]. A spectral tilt has also been reported to function to differentiate the three types of Korean stops. For this cue, usually the relative amplitude of the first versus the second harmonic of the voice source (*H*1-*H*2) is measured. *H*1-*H*2 differences are what distinguish Korean tense stops from aspirated and plain stops [5, 6]: *H*1-*H*2 difference at vowel onset was larger for aspirated and plain stops than for tense stops.

This paper examines how L2 learners use (or not use) *f0* and *H*1-*H*2 cues in conjunction with VOT to implement this three-way contrast? Can we predict that native speakers of languages that manipulate an *f0* cue in their native languages (such as Chinese) would be better equipped to produce these three contrasts? Further, can we predict that such language speakers' L2 acquisition would be significantly different from that of native speakers of languages which lack phonemic voice quality and pitch contrasts (such as English or Spanish)? Although there have been pioneering attempts at cross-language investigation of the Korean three-way stop contrast [2, 12], at present still very little is known about how native speakers of Chinese, English, and Spanish learn to produce the multiple phonetic features of this stop system, and how their range variations in phonetic spaces differ from native speakers of Korean.

Therefore, the current study aims to develop this inquiry by examining three acoustic cues - VOT, *f0,* and *H*1-*H*2 - for the production of the three-way contrast of Korean laryngeal stops by eight native speakers each of Chinese, English, Korean, and Spanish. It is predicted that speakers in each language group will actively use the acoustic dimension(s) known to them from their own native languages to differentiate the three types of stops. Accordingly, it is hypothesized that while Chinese speakers will rely on the *f0* and *H*1-*H*2 factors to produce contrast, English and Spanish speakers will likely make use of VOT more than the other acoustic factors.

2. METHODOLOGY

2.1. Subjects

Eight adult native speakers each of Mandarin Chinese, American English, Seoul Korean, and Latin American Spanish participated. The native speakers of Chinese and Korean spoke only Mandarin Chinese for at least 19 years and Seoul Korean for at least 24 years, respectively, before coming to the U.S., and they still speak Chinese and Korean in their family and daily community. The native speakers of American English were born in the U.S. and raised in English speaking households. The Spanish native speakers were raised in Spanish speaking households for at least 17 years before coming to the U.S. Detailed information about the participants, including age range, number of participant per gender, and years of Korean study is given in Table 1.

Table 1: (1) Age range, (2) number by gender, and (3) years of Korean study (range and mean) of participants

Language	Age	Number	Years of Korean study
Korean	23-41	4m, 4f	Native speaker
Chinese	21-24	4m, 4f	0.5-1.5, mean=0.6
English	21-27	4m, 4f	0.5-1.5, mean=1.0
Spanish	21-34	5m, 3f	0.5-1.5, mean=0.7

2.2. Stimuli

To minimize possible effects of vowels or initial/final consonants on VOT [3, 7, 14], as well as the intrinsic *f0* effects of vowels [9], an effort was made to choose words with the same neighboring segments. For comparison of *H*1-*H*2 differences, a low vowel like [a], which has the highest first formant frequency, is generally used to prevent the first formant energy from affecting the amplitude of *H*1 and *H*2. Accordingly, plain, aspirated, and tense stops in /CVn/ words were used as test materials, as given in Table 2.

All nine stimuli were real Korean words except in the case of /k^{h}ang/ (although it may be used as an onomatopoeia), and the test words were read in a carrier sentence, i.e., "igun ___ieyo" (This letter is ___), to avoid any effect on *f0* and VOT related to sentence position [1, 10]. The test materials were written in Korean orthography, and were presented in random order. Each test stimulus was repeated five times, yielding a total of 45 tokens per speaker.

Table 2: Target words with Korean stops

Plain	Aspirated	Tense
pang "room"	p^{h}ang "bang"	p*ang "bread"
tang "group"	t^{h}ang "soup"	t*ang "earth"
kang "river"	k^{h}ang "mimic word"	k*ang "perseverance"

2.3. Procedure

Recordings were made in a quiet office at UC-Berkeley directly into a laptop using a Sennheiser headset microphone. Subjects were asked to read the materials first for practice, then again for recording. Recordings were then digitized and analyzed using the PCquirer software package from SciConRD. For each sentence, synchronized displays of the sound waveform, a wide-band spectrogram, and an *f0* track were produced. For the measurement of *H*1-*H*2, fast Fourier transform (FFT) spectrum was produced. Overall, 1,440 tokens were obtained, 360 tokens (9 target words X 5 repetitions X 8 speakers) for each language group.

2.4. Measurements

f0, VOT, and *H*1-*H*2 were obtained using the PCquirer software package. *f0* values were taken at the onset of the following vowel. Some tokens did not display a pitch track at the onset of the vowel. In such cases, *f0* was measured at near points within 10 ms from the true onset.

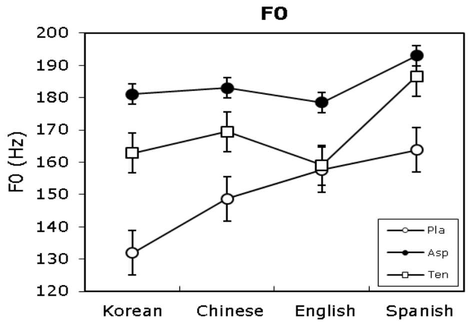
VOT was measured by marking the interval between the beginning of the stop burst and the onset of the periodic portion of the initial vowel of the target word, as displayed in the waveform and the 200 Hz broad bandwidth spectrogram.

*H*1-*H*2 was measured at the onset of the following vowel. The amplitude values were obtained based on the FFT spectrum with a 25 ms window sliced from the voice onset, and the difference of the amplitudes was calculated.

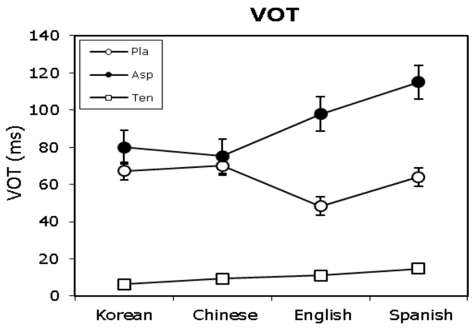
3. RESULTS

Fig. 1 gives the mean values with standard errors for the production of the three stop types (plain, aspirated, and tense) by native speakers of Korean, Chinese, English, and Spanish.

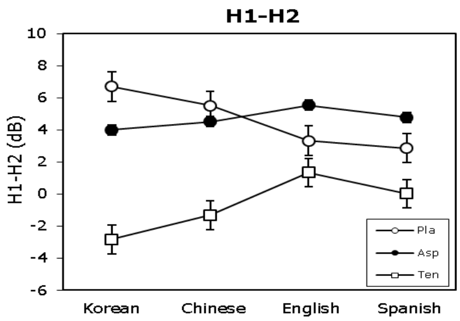
Figure 1: Mean values with standard errors for *F0*, VOT, and *H*1-*H*2, for the production of Korean stops (Plain, Aspirated, Tense) by native speakers of Korean (n=8), Chinese (n=8), English (n=8), and Spanish (n=8)



(a)



(b)



(c)

To test the prediction that the acoustic correlates of encoding the three-way stop contrast would differ according to the speaker's native language group, general linear model repeated measures analysis of variance (ANOVA) was conducted for *f0*, VOT, and *H*1-*H*2, respectively.

3.1. Fundamental frequency (*f0*)

Fig. 1(a) shows a similar *f0* pattern for native speakers of Chinese and Korean, i.e., *f0* is the lowest for vowels following plain stops, higher following tense stops, and highest following aspirated. As for native speakers of English, however, *f0* values are merged for plain and tense stops, and distinctively higher for aspirated stops. Conversely, *f0* was merged for aspirated and tense stops in the case of Spanish speakers.

Repeated measures one-way ANOVA tests for each language group, with *f0* as a dependent variable and stop type as a factor, showed the significant effects for each language [Korean: $F(2, 357)=28.34$, $P<0.001$), Chinese: $F(2, 357)=14.76$, $p<0.001$, English: $F(2, 357)=8.90$, $p<0.001$, Spanish: $F(2, 357)=7.86$, $p<0.001$]. In case of English and Spanish, the interaction might be mostly due to the distinctively higher *f0* values for aspirated and tense stops, respectively.

A pairwise comparison by a post hoc test (LSD, p= .05 level) showed that *f0* patterns of the three types of stops were different from one another for each language group. While there were significant main effects for each pair of stop types for Chinese and Korean speakers, no significant effect was found for English speakers (between plain and tense stops) and Spanish speakers (between aspirated and tense stops), supporting the hypothesis and the graph seen in Fig. 1.

An interesting pattern emerges in that Spanish speakers show a high average for the *f0* values of tense stops. This may be due to their equating of Korean tense stops with stops used in accented words in their native language. It is also important to note that the reason for the especially high *f0* average for Spanish is due to one female subject who has a distinctively higher *f0* range.

3.2. Voice onset time (VOT)

The graph in Fig. 1(b) appears to be consistent with the prediction that English speakers would primarily manipulate the VOT dimension to create the three-way stop contrast. Repeated measures one-way ANOVA tests for each language group, with VOT as a dependent variable and stop type as a factor, showed the significant effects for each language [Korean: $F(2, 357)=831.12$, $P<0.001$), Chinese: $F(2,357)=65.66$, $p<0.001$, English: $F(2, 357)=293.69$ $p<0.001$, Spanish: $F(2, 357)=194.33$, $p<0.001$].

A pairwise comparison by a post hoc test (LSD, p= .05 level) showed the significant effect for each comparison of stops, except for the plain and aspirated stops for the Chinese speakers' group. In addition, the gender effect was found among the

English [F (1, 358)=3.87, p<0.001] and Spanish speakers' groups [F (1, 358)=16.06, p<0.001], with females exhibiting longer VOT on average when compared with males.

3.3. H1-H2

Fig. 1(c) shows the similar *H*1-*H*2 pattern between Chinese and Korean, and between English and Spanish. Repeated measures one-way ANOVA tests for each language group, with *H*1-*H*2 as a dependent variable and stop type as a factor, showed the significant effects for each language [Korean: F (2, 357)=133.93,P<0.001), Chinese: F(2, 357)=399,81, p<0.001, English: F(2, 357)=293. 68, p<0.001, Spanish: F(2, 357)=39.39, p<0.001].

A pairwise comparison by a post hoc test (LSD, p= .05 level) showed the significant effect for each comparison of stops in Korean and Spanish speakers' groups. However, in the Chinese speakers' group, the effect was significant between tense and plain/aspirated stops, and in the English speakers' group, the significant effect was found only between tense and aspirated stops.

4. DISCUSSION

The results indicate that production of stops in L2 predominantly relies on L1, and accordingly, if features used to mark phonetic categories in L2 are also used in L1, more effective phonetic acquisition of L2 can be expected. Besides the fact that Chinese is a tone language, recent studies have shown that Mandarin Chinese speakers were significantly more sensitive to changes in *H*1-*H*2 values than were English and Spanish listeners [8], and it is also reported that the low-dipping Tone 3 in Mandarin is often produced with creaky voice [4]. Therefore, native Chinese speakers can benefit from the use of all three of the aforementioned acoustic cues (*f0*, VOT, *H*1-*H*2) to produce the three-way stop contrast in a way similar to native speakers of Korean. In English and Spanish, however, tone (and thus *H*1) is not manipulated on a lexical level, and thus native speakers of those languages lack sensitivity to both *f0* and *H*1-*H*2 contrasts, and therefore generally showing less similarity in their stop production to the Korean L1 speakers. Lastly, gender effect on VOT was found only among English and Spanish speakers. At least in the case of English, this can be understood to be the influence of L1, as it has been recognized that the females generally exhibit longer VOT on average than males for English voiceless stops [11].

5. ACKNOWLEDGEMENTS

The authors would like to thank Professor John Lie (Chair, Center for Korean studies, UC Berkeley) for his support to this study.

6. REFERENCES

[1] Bran, J.A., Lauder, M.Z., Daniloff, R. 1977. Phonological contrastivity in conversation: A comparative study of voice onset time. *Journal of Phonetics* 5, 339-350.
[2] Chang, B. 2009. The implementation of Laryngeal Contrast in Korean as a Second Language, *UC Berkeley Phonology Lab Annual Report (2009)*, 321- 329.
[3] Cho, T., Ladefoged, P. 1999. Variation and universality in VOT: Evidence from 18 languages, *Journal of Phonetics* 27, 207-229.
[4] Davison, D.S. 1991. An acoustic study of so-called creaky voice in Tianjin Mandarin. *UCLA Working papers in Phonetics* 78, 50-57.
[5] Kang, K-H, Guion, S. 2008. Clear speech production of Korean stops: Changing phonetics targets and enhancement strategies. *Journal of Acoustical Society of America* 124, 3909-3917.
[6] Kim, M, Stoel-Gammon, C. 2009. The acquisition of Korean word-initial stops. *Journal of Acoustical Society of America* 125, 3950-3961.
[7] Klatt, D.H. 1975. Voice onset time, frication, and aspiration in word-initial consonant cluster. *Journal of Speech and Hearing Research* 18. 129-140.
[8] Kreiman, J, Gerratt, B.R. 2010. Perceptual sensitivity to first harmonic amplitude in the voice source. *Journal of Acoustical Society of America* 128, 2085-2089.
[9] Lehiste, I., Peterson, G.E. 1961. Some basic considerations in the analysis of intonation. *Journal of Acoustical Society of America* 33, 419-425.
[10] Lisker, L, Abramson, A.S. 1964. A cross-language study of voicing in initial stops: Acoustic measurements. *Word* 20, 384-422.
[11] Morris, R.J. et al. 2008. Voice onset time differences between adult males and females: Isolated syllables. *Journal of Phonetics* 36, 308-317.
[12] Shin, E. 2007. How do non-heritage students learn to make the three-way contrast of Korean stops? *Korean Language in America* 12, 85-105.
[13] Silva, D.J. 2006. Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology* 23, 287-308.
[14] Weismer, G. 1979. Sensitivity of voice-onset time (VOT) measures to certain segmental features in speech production. *Journal of Phonetics* 7, 197-204.
[15] Wright, J.D. 2007. *Laryngeal Contrast in Seoul Korean*, Ph.D. thesis, University of Pennsylvania, Philadelphia, PA.