Harvard International Symposium on Korean Linguistics (HARVARD ISOKL)

Executive Committee

Young-Se Kang, Kookmin University
Young-joo Kim, Hong-Ik University
Susumu Kuno, Harvard University
Ik-Hwan Lee, Yonsei University
Joan Maling, Brandeis University
John Whitman, Cornell University

Program Committee, the preceding and following years

Sang-Cheol Ahn, Kyung Hee University
Young-mee Yu Cho, Rutgers University
Donna B. Gerdts, Simon Fraser University
Young-Seok Kim, Sogang University
Sang-Oak Lee, Seoul National University
James Hye Suk Yoon, University of Illinois at Urbana-Champaign

Edited by

Susumu Kuno
John Whitman
Young-Se Kang

Ik-Hwan Lee
Joan Maling
Young-joo Kim

Department of Linguistics
Harvard University

2001
Speech Perception and Production of the English Final Voicing Contrast by Korean Speakers

Woohyeok Chang and William Idsardi

University of Delaware

1 Introduction

It has been widely observed that English vowels are longer before voiced consonants than before their voiceless counterparts (House & Fairbanks 1953; Peterson & Lehiste 1960; House 1961; Nasby 1970). In addition, perceptual studies have also shown that vowel duration serves as an important perceptual cue to the voicing contrast in the following obstruent (Denes 1955; Raphael 1972; Krause 1982; Lehman & Sharf 1989) and that children as young as three years of age can reliably identify consonant voicing from preceding vowel durations (Krause 1982). This duration cue appears to be learned (Zimmerman & Sapon 1958; House 1961; Lehiste 1970), and the use of this cue has been confirmed even with language-impaired children. In particular, even English children who omit word-final stops as part of a speech production deficit can still perceptually differentiate voiced and voiceless consonants by means of the differential durations of the preceding vowels (Weismer, Dimsen & Elbert 1981).

Although there are many previous studies on the production and perception of the durational difference by L1 speakers, far less is known about L2 learners’ use of vowel duration to identify the voicing contrast in postvocalic consonants. It is, therefore, of interest to compare native speakers with Korean speakers of English in the use of the duration cue in both perception and production. The main goals of the present study are to investigate the English final voicing contrast to confirm what the cues to the contrasts are in coda position, and to examine the acquisition of the vowel duration cue by Korean speakers, by conducting both perception and production experiments.

This paper will show that the Korean group’s performance on an auditory task is much better than their performance on identification and production tasks. This is explained directly if the AX discrimination task makes contact with a different layer of perception. In particular, we conclude that the AX discrimination task can be done at the auditory or phonetic level, where differences in vowel length are still encoded in the (universal) representation. In contrast, the identification and production tasks probe the language-particular mental representation of vowel length and voicing. Analyzing subjects’ poor performance in the auditory discrimination test where the vowel duration before voiced and
voiceless consonants is equalized, we will argue that the durational difference of vowels is important to both Korean and English group as cue for the voicing status of postvocalic consonants. Additionally, comparing the standard model of English with a new model of laryngeal features (following Avery & Idsardi 2000), we will show that the latter is better able to explain vowel-lengthening before contrastive obstruents without invoking opaque derivations. This present study, therefore, provides additional support that English obstruents contrast in Glottal Width, like Korean stops, and do not contrast in voicing.

2 Theoretical preliminaries on English and Korean obstruents

As this paper is primarily concerned with Korean speakers' acquisition of English word-final voicing contrast, we first need to investigate some important facts about English and Korean obstruents.

2.1 Phonetic realization of English obstruents

In (1) and (2) we compare the now standard model of Laryngeal features (Lombardi 1991) with that proposed by Avery & Idsardi (2000, 2001).

(1) Laryngeal
   [voice] [spread] [constricted]
(2) Laryngeal
   Glottal Width (GW) Glottal Tension (GT)
   [spread] [constricted] [slack] [stiff]

While the phonological features of [voice], [spread], and [constricted] are direct dependents of the Laryngeal node in (1), in (2) the antagonistic pairs of features are grouped first into dimensions, such as Glottal Width (GW). Avery & Idsardi go on to analyze obstruents as containing lexical contrasts only in bare dimension nodes, with the terminal features added later to complete the articulation of the segment or to enhance a phonological contrast. Thus, phonological representations are generally underspecified. In English the "voiced" series is unspecified laryngeally, whereas the voiceless series is specified only for GW, which is then completed by adding [spread] (aspiration) or [constricted] (glottalization).

The completion of GW is the only consistent phonetic correlate within the obstruents themselves. For stops, the GW series is generally realized with aspiration in onset position, and with pre-glottalization in coda position. With regard to fricatives, Smith (1997), in her phonetic study on English /s/ and /z/, found that all final fricatives were essentially voiceless; however, the GW nature of /s/ is revealed by the presence of greater airflow, indicating the presence of [spread] phonetically. Smith also found the expected difference in the duration of the preceding vowel, which may still make it possible to differentiate /s/ from /z/ even in the absence of any vocal cord vibration in devoiced [z].

However, the featural cues within the obstruents are not particularly salient and consequently the most important cue is the duration of the preceding vowel. English vowels are substantially longer before "voiced" obstruents, which induces an inverse correlation between the length of vowels and postvocalic consonants. This is illustrated in figure 1, where the voiceless obstruents are longer than the voiced ones (e.g., 170 ms vs. 129 ms and 121 ms vs. 119 ms), and the preceding vowel is shorter before voiceless obstruents (e.g., 166 ms vs. 276 ms and 171 vs. 291 ms).

![Figure 1 Measurement of vowels and consonants in English minimal pairs](image)

2.2 Phonetic realization of Korean obstruents

Most previous work on Korean phonology maintains that Korean has three-way distinction in stops and two-way distinction in fricatives, as shown in (3).

(3) a) stops
   i. plain: unmarked [p, t, k]
   ii. tense: constricted glottis [p', t', k'] (longer closure)
   iii. aspirated: spread glottis [p^, t^, k^] (longer closure)

b) fricatives (based on Avery and Idsardi, 2000)
   i. short: one consonant timing node with spread glottis [s^]
   ii. long: two consonant timing node with spread glottis [s^]

In (3a) the plain and aspirated stop in Korean correspond with English stop distinction in that English stops can be also divided into the unmarked (plain) case and the marked (GW) cases, following Avery & Idsardi (2000). In contrast, the two-way distinction in (3b) does not correspond to the GW contrast in English fricatives. Although there may be small differences in glottal width in Korean fricatives, these differences are not substantial, and instead the contrastive difference within Korean fricatives is length.
It is important to note that vowel length differences similar to those for English have also been reported for Korean. Specifically, Cho (1996) shows that a substantial inverse correlation can be found between the vowel length and the closure duration of the following consonant in CVCV sequences. Vowels before plain stops are longer than vowels before tense and aspirated stops since the plain consonants have the shortest closure duration. Thus the phonetic manifestation of stops in Korean is similar to that in English, with the vowel length of the preceding vowel providing a substantial cue to the laryngeal properties of the following obstruct.

3 Experiment

The goal of this experiment is to investigate whether L2 learners can use the vowel duration cue to distinguish voicing contrasts in word-final consonants in English. Before introducing the methodology of experiment, let us first examine what the possible research questions are and what predictions we can make.

3.1 Research Questions and Predictions

The present study will provide answers for the following questions:

1. Can L2 learners discriminate the final voicing contrast perceptually? That is, do learners perform like English native speakers in discriminating the final voicing contrast in an auditory task (AX discrimination test) and in an identification task?
2. Do L1 and L2 speakers show different levels of performance between the normal condition and the manipulated condition in which vowel length before voiced obstruents is shortened?
3. Do L2 learners use the vowel length cue to distinguish the final voicing contrast in their own production?

Given that Korean also has an inverse correlation in the length of a vowel and the following stop, L2 speakers should show reasonably good performance in the discrimination perception test. Additionally, if vowel length is crucial to the perception of the final voicing contrasts in English, both English and Korean speakers will show poor performance in the manipulated condition where the vowel length difference is neutralized. In general, however, Korean speakers will perform worse than English speakers for the following reasons. First, this is a second language test for Korean speakers. In addition, the set of stop categories in Korean does not match that of English, as Korean has an extra category (3 categories in Korean vs 2 categories in English). Worse still, the fricative contrast is different between Korean and English: Korean fricatives contrast in length (e.g., /s/ vs /ss/ or /s/ vs /s/) whereas English fricatives contrast in Glottal Width (e.g., /s/ with GW vs /s/ without).

3.2 Methods

17 adults from the two language groups participated in this study. The experimental group consisted of 10 Korean ESL speakers who had learned English as their only second language. All of the subjects in this group were raised in Korea, and they had studied English in Korea for between 10 and 16 years. These subjects were between 22 and 34 years of age at the time of testing. The subjects had been residing in North America from 3 to 10 months at the time of testing. The other group in this study consisted of 7 monolingual speakers of American English who served as controls. They ranged in age from 23 to 33 years old.

Table 1 summarizes the background information for each of the two groups.

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean age at testing</th>
<th>Mean age of exposure</th>
<th>Mean years of study</th>
<th>Mean time in North America</th>
</tr>
</thead>
<tbody>
<tr>
<td>Korean</td>
<td>27</td>
<td>14</td>
<td>10</td>
<td>7 months</td>
</tr>
<tr>
<td>Controls</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1 Subject information

The stimuli consisted of 18 minimal pairs exhibiting the final “voicing” contrast, which are limited to monosyllabic words, as illustrated in Table 2.

<table>
<thead>
<tr>
<th>stop</th>
<th>rip</th>
<th>cap</th>
<th>cup</th>
<th>feet</th>
<th>cot</th>
<th>cat</th>
<th>back</th>
<th>task</th>
<th>task</th>
<th>task</th>
</tr>
</thead>
<tbody>
<tr>
<td>rib</td>
<td>cab</td>
<td>cub</td>
<td>feed</td>
<td>cod</td>
<td>cad</td>
<td>tag</td>
<td>tuck</td>
<td>tuck</td>
<td>tuck</td>
<td>tuck</td>
</tr>
<tr>
<td>fricative</td>
<td>safe</td>
<td>leaf</td>
<td>half</td>
<td>peace</td>
<td>face</td>
<td>loose</td>
<td>mouth</td>
<td>teeth</td>
<td>teeth</td>
<td>teeth</td>
</tr>
<tr>
<td>fricative</td>
<td>save</td>
<td>leave</td>
<td>have</td>
<td>peas</td>
<td>phase</td>
<td>lose</td>
<td>mouth (verb)</td>
<td>teeth</td>
<td>teeth</td>
<td>teeth</td>
</tr>
</tbody>
</table>

Table 2 Target words

A native speaker of American English first pronounced a randomized list of 36 sentences containing the test words (see the Appendix), as exemplified in (4).

(4) a She sometimes wears a cap.
b I want to take a cab.
c There is not a moment to lose.
d The fierce dog has broken loose.

The sentences were recorded onto a digital mini disc and then converted to WAV files at a 22 kHz sampling rate. The target words were then edited using the Praat program (version 3.8.26). For the AX discrimination test, stimuli were constructed in the Praat program with uniform intervals of 1500 ms between members of a pair (i.e., an ISI of 1500 ms), and 3000 ms intervals were main-
tained between trials (i.e., an ITI of 3000 ms). For the Identification test, the ITI was also 3000 ms.

**AX Discrimination Task**

The auditory test was designed to test whether subjects could discriminate minimal pairs containing a voicing contrast in the coda. The subjects were asked to respond to each trial by circling either ‘same’ or ‘different’ on a response sheet after they listened to a pair of words. This was done to see whether they could discriminate at the phonetic level. Two hypotheses are relevant to the auditory test. First, the L2 speakers’ performance may be worse than the L1 speakers’. Second, both L1 and L2 speakers should show significantly poorer performance in the manipulated condition in which the length of vowels is equalized. In this task, subjects heard 102 pairs of words, as summarized in the table 3.

<table>
<thead>
<tr>
<th>Contrasts</th>
<th>Number</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voicing and length</td>
<td>36 trials</td>
<td>([kʰ]'ep] vs [kʰ]'eb]</td>
</tr>
<tr>
<td>Voicing and length</td>
<td>36 trials</td>
<td>([pʰ]'vs] vs [pʰ]'ist]</td>
</tr>
<tr>
<td>Voicing and length</td>
<td>18 trials</td>
<td>([pʰ]'vs] vs [pʰ]'is]</td>
</tr>
<tr>
<td>Other (controls)</td>
<td>18 pairs</td>
<td>([kʰ]'ep] vs [kʰ]'eb]</td>
</tr>
<tr>
<td>(length neutralized)</td>
<td>12 trials</td>
<td>([kʰ]'ep] vs [kʰ]'eb]</td>
</tr>
</tbody>
</table>

**Table 3** Test items in the auditory task

The purpose of including the manipulation of vowel length in this task is to investigate whether vowel length is a significant perceptual cue for distinguishing English voiceless and voiced consonants in word-final position. If there is no difference in subjects’ performance between the normal and the manipulated condition, then some other cue in the coda consonant itself provides enough information to distinguish the obstruents. Otherwise, vowel length must be a major significant cue. 18 minimal pairs for voiced and voiceless consonants were created with the same vowel length by shortening the vowels before voiceless consonants to match that found with the voiceless cognate. The steady-state portion of the vowel, which excluded initial and final transition information, was pitch-marked. The vowel continuum was made by simultaneously exciting a glottal pulse from each end of the steady-state vowel. Boundary points for excision were at zero crossings in the stimulus waveform. Finally, subjects heard the minimal pairs such as /trɪ/ vs /trɪp/, with equalized vowel durations.

**Identification Task**

The identification test was designed to test whether subjects can identify the presented words. The subjects heard one member of a minimal pair and circled one of two words on a response sheet. This task is somewhat harder than the discrimination task, and therefore L2 speakers’ performance should be somewhat worse than L1 speakers.

**Production Task**

The production task was designed to investigate whether L2 learners can use vowel length to differentiate final voicing contrasts in their own production. The subjects were asked to pronounce 36 randomized sentences twice apiece. Each sentence contains one target word at the end of the sentence. The duration of the vowels in the target words was then measured using the Praat program. When a vowel is adjacent to a voiced consonant, the transition section was not included in the vowel. That is, the clear appearance of higher formants was considered the edge of the vowel for measurement purposes.

Each subject was tested individually in a quiet room with the experimenter. Each subject completed the production task first, followed by the discrimination task and finally the identification task.

3.2 Results

**AX Discrimination Task**

The overall performance (see Figure 2) shows that there is no difference between the performance of Korean group (85.3%) and that of control group (85.01%). A two-tailed t-test also shows that there is no difference between the Korean and control groups on the auditory test: [t (15) = 1.5, p = 88]

![Figure 2: Auditory task by group](image)

![Figure 3: Auditory task by type](image)

In addition to the overall performance by group, Figure 3 shows the rate of correct response on minimal pairs divided into stops (Korean group: 91.38% vs English group: 84.91%) and fricatives (Korean group: 79.18% vs English group: 71.79%).
group: 85.7%). Paired t-tests reveal that the Korean group is significantly better with stops than with fricatives (p < 0.0001) whereas the control group does not show any significant difference between stops and fricatives (p = 0.8105). The Korean group is also somewhat better than the English group with stops (p = 0.0172), but marginally worse than the English group with fricatives (p = 0.0302).

In the manipulated condition, a major drop for performance is found for both groups. Figure 4 shows that both groups are much worse (below 60%) when the vowel length is neutralized. Vowel length must then be a major cue. A t-test analysis shows that the performance for the manipulated condition is not significantly different between groups (p = 0.5375). But both groups are worse in this condition as compared with the normal condition (p < 0.0001). If we divide the performance in manipulated condition by type, Figure 5 shows that there is no significant difference between stops and fricatives for either group (Korean group: 54.45% in stops, 55.36% in fricatives; English group: 60.34% in stops, 55.58% in fricatives). [Korean: t (18) = 207, p = 8385; English: t (12) = -1.157, p = 2629]

### Production Task

For both groups, vowels are shorter before voiced consonants than before voiceless ones. The mean durations of vowels are summarized in Table 4.

<table>
<thead>
<tr>
<th>Postvocalic consonant</th>
<th>Korean group</th>
<th>English group</th>
</tr>
</thead>
<tbody>
<tr>
<td>voiced</td>
<td>179.59 ms</td>
<td>234.61 ms</td>
</tr>
<tr>
<td>voiceless</td>
<td>164.60 ms</td>
<td>152.74 ms</td>
</tr>
<tr>
<td>difference</td>
<td>14.99 ms</td>
<td>81.87 ms</td>
</tr>
</tbody>
</table>

Table 4 Mean duration of vowel in Korean and English speakers’ production

The difference in vowel duration is much greater for English group (82 ms vs. 15 ms). This, in turn, also means that the ratio of vowel length before voiced consonants to that before voiceless consonants is much greater for English group (1.51 vs 1.09).

### 4 Discussion

In the AX discrimination task, no major differences were found between the Korean and the English group. However, the Korean group did perform significantly worse than the English group on the identification task. Furthermore, the results of the production task show that Korean speakers did not use the vowel duration cue to as great an extent to contrast final voiceless consonants from voiced ones in their production. Based on these findings, we can conclude that the AX discrimination task makes contact with a different layer of perception. While the identification and production task deal with long-term memory and the mental representation of vowel length and voicing, the AX discrimination task can be done at the auditory or phonetic level where differences in vowel length are still encoded in the representation.

In addition, the finding that both groups show significantly poorer performance when vowel length is neutralized suggests that vowel length for both groups is the major cue for the differentiation of voicing contrast in the word.
final consonants. Even though the duration cue is a low level distinction for Korean speakers, they do possess a substantially correct idea of what the English contrasts are. We should also note that the performance in the manipulated condition is still slightly better than chance (Korean: 56.1%; English: 57.9%). This small residual ability to distinguish the two classes indicates that there must still be some acoustic differences between at least some of the items. One interpretation is that this may be evidence for incomplete neutralization as has been found in various experiments on German. Recall that the English voiceless series are pre-glottalized in final position for many speakers, and this difference could be used to discriminate between the stimuli. Nevertheless, the results of this study show clearly that it is the vowel length that is the primary cue.

With all of this as background, the next question, then, is to consider which analysis will give a better account of the phonetic and perceptual findings: the traditional Laryngeal analysis, as in (1), or the GW analysis (2) (Avery and Iddsardi 2000). The voicing analysis requires an opaque derivation to capture the entire picture. The vowels must be lengthened prior to the removal of the voicing feature from codas. That is, vowels need to be lengthened only before the voiced consonants, which are, however, not actually voiced at the surface. In the case of voiceless stops, [constricted glottis] must be added because they are in coda position. Importantly, then, this derivation of the voiceless and voiced case reverses the markedness in the derivation. It is the voiced consonants that are marked in the underlying representation, but by the end of derivation it is instead the voiceless ones that are marked. This is very unusual and undesirable property in a derivation. In contrast, in the GW analysis, no opacity or reversal of markedness is required. Given that only the so-called ‘voiceless’ obstruents are marked with GW node in the underlying representation, we propose that GW consonants themselves lengthen by spreading back onto the vowel, and furthermore that the vowel spreads onto any unmarked obstruent. As a result, we have no flip in the markedness relationship, but an enhancement of the contrast. We are currently investigating the extent of the vowel-lengthening rule by examining the length of vowels before sonorants. Preliminary results indicate an intermediate length before sonorants, indicating a neutral length with sonorants, but these findings must be confirmed through further research.

Finally, as we already predicted in section 3, the Korean group is consistently better with stops than with fricatives. The simplest explanation for this is that the Korean fricative system is impoverished relative to the stop system, as there are 3 contrastive types of stops: /p, p', p"/ but only 2 contrastive fricatives: /s, s'/ In addition, when comparing the Korean systems with the English systems, the Korean stops have just one extra contrast relative to English. That is, the Korean stop system is a superset of the English one. To learn the English stop system the Korean speakers need only suppress the contrastive use of length (or tenseness). In contrast, the Korean fricative system employs a different contrast (length) that of English (GW). So in order to learn the English fricative system, Korean learners must both suppress the use of length and discover that GW is contrastive in fricatives as well. Therefore, the English obstruent system is more congruent with Korean stops than with Korean fricatives, leading to the earlier (and better) acquisition of the English stop system by Korean learners.

5 Conclusion

In this paper, we have examined the acquisition of the English vowel duration cue by Korean speakers. Based on the results from the three different tasks, we conclude that Korean speakers store neither vowel length nor voicing in memorized representations and do not completely internalize the lengthening of the preceding vowel as a rule to differentiate the voicing contrasts of final consonants, even though they can detect the acoustic differences in vowel duration provided that they are tested in an appropriate task. We further found that the vowel length played a crucial role for both Korean and English speakers, as all performance dropped dramatically when the vowel length was neutralized.

Furthermore, this study can provide a refined analysis of the difference in vowel length conditioned by voicing contrast, showing that the GW analysis is superior to the traditional voicing analysis in explaining the full range of phonetic and perceptual facts.

Notes

* We would like to thank Cindy Brown for her comments and suggestions.

References

Appendix: recording list

1. My pants have a rip.
2. This building resembles a rib.
3. She sometimes wears a cap.
4. I want to take a cab.
5. I am fond of the cup.
6. A lion gave birth to a cub.
7. You must learn to stand on your own feet.
8. She has a large family to feed.
9. On a ship, we usually sleep on a cot.
10. My favorite food is fresh cod.
11. We've got three dogs and a cat.
12. He's no gentleman, he's a cad.
13. It takes me an hour to walk there and back.
14. Today, we got a new bag.
15. It would be unwise to change tack.
16. I realize that I lost my name-tag.

17. I just want to have a snack.
18. There must be a snag.
19. The missing child was found safe.
20. It's prudent to save.
21. A fallen leaf is a dead leaf.
22. The old professor has just gone on leave.
23. No goals were scored in the first half.
24. It is a sandwich that I want to have.
25. After fighting, the people longed for peace.
26. This is the soup made of dried peas.
27. His ambition was to meet her face to face.
28. The child is going through a difficult phase.
29. The fierce dog has broken loose.
30. There is not a moment to lose.
31. From time to time, She's got a big mouth.
32. Those are curses that they silently mouth.
33. Finally the employers showed their teeth.
34. Babies like to chew something when they teethe.
35. Finally the employers showed their teeth.
36. Babies like to chew something when they teethe.

Department of Linguistics
University of Delaware
46 East Delaware Avenue
Newark, Delaware, 19716-2551, U.S.A.
e-mail: cwh@udel.edu, jsardi@udel.edu