Speech Perception and Production of English Postvocalic Voicing by Korean and English Speakers*

Woohyeok Chang**

ABSTRACT

The main purpose of this study is to investigate whether Korean learners can use the vowel duration cue to distinguish voicing contrasts in word-final consonants in English. Given that the Korean group's performance on the auditory task was much better than their performance on the identification task or on the production task, we conclude that the AX discrimination task makes contact with a different layer of perception. In particular, 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 contrast, the identification and production tasks are probing the mental representation of vowel length and voicing. It was also founded that Korean speakers stored neither vowel length nor voicing in memorized representations and did not internalize the lengthening of the preceding vowel as a rule to differentiate the voicing contrasts of final consonants, even though they were able to detect the acoustic differences in vowel duration provided that they were tested in an appropriate task.

Keywords: Vowel duration, Postvocalic, Voicing Contrasts, Perception, Discrimination Task, Identification Task, Production Task

1. Introduction

It has been widely accepted in the literature that VOT, which refers to a timing lag between the articulatory burst release of word-initial stop consonants and the onset of voicing for the following vowel, is one of the important perceptual or acoustic cues distinguishing between voiced-voiceless pairs of word-initial stops (Lisker & Abramson, 1964).

---

* I would like to thank two anonymous reviewers for their valuable comments on this paper.

** Department of English, Dankook University

For example, in English word-initial voiceless stops show long lag VOT values, whereas word-initial voiced stops show short lag VOT values, such as [t] versus [d].

For the word-final voicing contrast, however, vowel lengthening before final voiced consonants provides an important perceptual cue distinguishing between voiced-voiceless pairs of word-final consonants in English (House & Fairbanks, 1953; Peterson & Lehiste, 1960; House, 1961; Naeser, 1970). 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, Dinnsen & Elbert, 1981).

The acquisition of VOT by L2 learners, which plays a powerful acoustic cue for discriminating English voicing contrasts in the word-initial position, has been quite extensively studied (Kuijpers, 1996; Curtin, Goad, & Pater, 1998). 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 English 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 could be accounted for by assuming that 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 lan-

guage-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.

2. Research Questions and Predictions

2.1 Research Questions

The present study attempts to 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?

2.2 Predictions

Given that Korean also has an inverse correlation in the length of a vowel and the following stop, one might predict that L2 speakers will show reasonably good performance in the discrimination perception test.[1] 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.

---

1) 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 in Korean. 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 obstruent.

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^{h}] vs. [s:]) whereas English fricatives contrast in Glottal Width (e.g., /s/ with wide spread of glottis vs. /z/ without). Smith (1997), in her phonetic study on English /s/ and /z/, found that all final fricatives were essentially voiceless, and the physical nature of /s/ is revealed by the presence of greater airflow, indicating the presence of [spread] phonetically.

3. Method

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, and they 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. The background information for each of the two groups is summarized in Table 1.

Table 1. Subject information

Group	Mean age at testing	Mean age of exposure	Mean years of study	Mean residence in North America
Korean	27	14	10	7 months
American	29	-	-	-

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

Table 2. Target words

stop	rip	cap	cup	feet	cot	cat	back	tack	snack
	rib	cab	cub	feed	cod	cad	bag	tag	snag
fricative	safe	leaf	half	peace	face	loose	mouth	teeth	teeth
	save	leave	have	peas	phase	lose	mouth (verb)	teethe	teethe

A native speaker of American English first pronounced a randomized list of 36 sentences containing the test words (see the Appendix), as exemplified in (1).

(1) 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 in a sound-treated booth 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 maintained between trials (i.e., an ITI of 3000 ms). For the Identification test, the ITI was also 3000 ms.

3.1 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 over headphone. 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 Table 3.

Table 3. Test items in the auditory task

Contrasts		Examples	Number
Voicing and length	Different	[k^{h}æp] vs. [k^{h}æ:b] [p^{h}is] vs. [p^{h}i:z]	36 trials (18 pairs with 2 repetitions)
Voicing and length	Same	[k^{h}æp] vs. [k^{h}æp] [k^{h}æ:b] vs. [k^{h}æ:b] [p^{h}is] vs. [p^{h}is] [p^{h}i:z] vs. [p^{h}i:z]	36 trials
Voicing (length neutralized)	Different	[k^{h}æp] vs. [k^{h}æb] [p^{h}is] vs. [p^{h}iz]	18 trials (18 pairs)
Other (controls)	Different	[k^{h}æp] vs. [k^{h}at]	12 trials (6 pairs with 2 repetitions)

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 voiced 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 excising 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 /rib/ vs. /rip/, with equalized vowel durations.

3.2 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 over headphone and circled one of two words on a response sheet. Each member of a minimal pair was presented to the listeners twice in random order. The stimuli thus consist of 72 trials for target words and 12 trials for control ones. This task is somewhat harder than the discrimination task, and therefore L2 speakers' performance should be somewhat worse than L1 speakers'.

3.3 Production Task

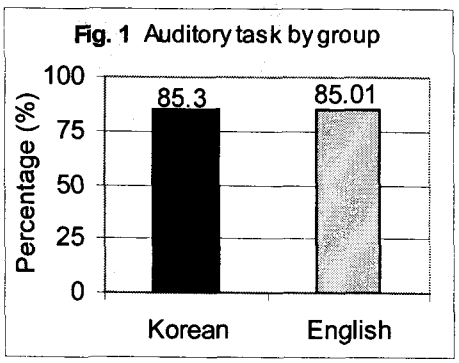
The production task was designed to investigate whether L2 learners can use vowel length to differentiate final voicing contrasts in their own production. The subjects were asked to pronounce 36 randomized sentences twice apiece. Each sentence contains one target word at the end of the sentence (See the appendix for the list of sentences). 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 appearance of steady-state formants was considered the edge of the vowel for measurement purposes.

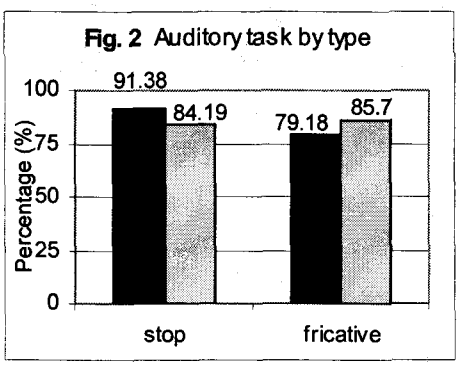
Each subject was tested individually in a sound-treated booth with the experimenter. Each subject completed the production task first, followed by the discrimination task, and finally the identification task.

4. Results

4.1 AX Discrimination Task

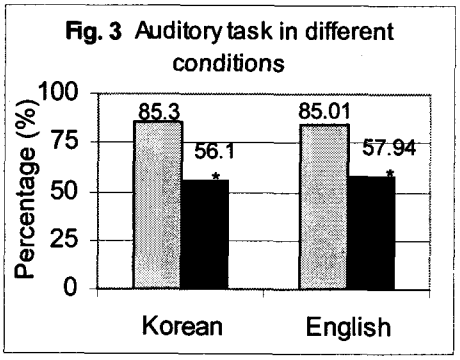
The overall performance (see Figure 1) shows that there is no difference between the performance of Korean group (85.3%) and that of English (control) group (85.01%). A two-tailed t-test also shows that there is no difference between the Korean and control groups on the auditory test: [t (15) = 0.15, p = 0.88].

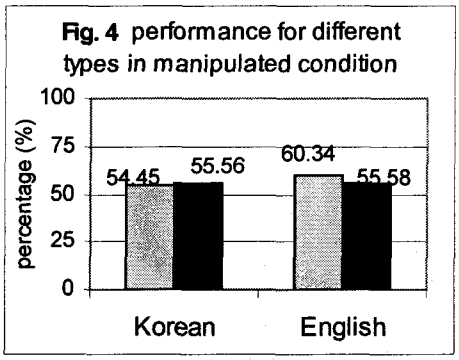




In addition to the overall performance by group, Figure 2 shows the rate of correct response on minimal pairs divided into stops (Korean group: 91.38% vs. English group: 84.19%) and fricatives (Korean group: 79.18% vs. English group: 85.7%). Paired t-tests reveal that the Korean group is significantly 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 3 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 4 shows that there is no significant difference between stops and fricatives for either group (Korean group: 54.45% in stops, 55.56% in fricatives; English group: 60.34% in stops, 55.58% in fricatives): [Korean: $t(18) = 0.207$, $p = 0.8385$; English: $t(12) = -1.157$, $p = 0.2629$].



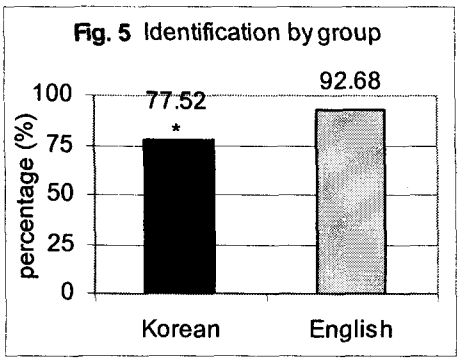


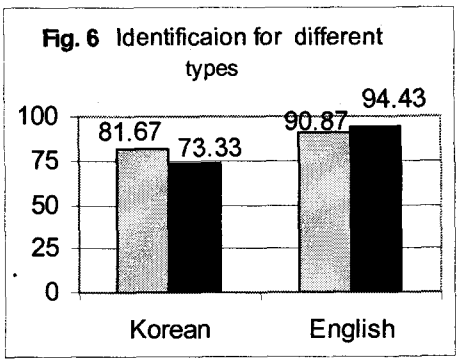
4.2 Identification Task

The overall performance on the identification task is shown in Figure 5. The Korean group (77.52%) did perform worse than the English group (92.68%), and this difference is statistically significant: [$t(15) = 15.328$, $p < 0.0001$].

Figure 6 shows the performance broken down by stops and fricatives. The Korean group shows somewhat better performance with stops (81.67%) than with fricatives

(73.33%), and this is moderately significant (p = 0.0111). In contrast, the English group shows no significant difference between stops (90.87%) and fricatives (94.43%): (p = 0.0928).





4.3 Production Task

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

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

Postvocalic consonant	Korean group	English group
voiced	179.59 ms	234.61 ms
voiceless	164.60 ms	152.74 ms
difference	14.99 ms	81.87 ms

Different from Korean speakers' performance, the difference in vowel duration is much greater for English group (82 ms *vs.* 15 ms). This is also confirmed when we analyze the result by consonant type (stops vs. fricatives). Specifically, much difference in vowel duration between voiced and voiceless ones is observed in both stops and fricatives for English group. That is, the vowels before voiced stops are longer than the ones before voiceless stops by 83.09 ms, and the vowels before voiced fricatives are longer than the ones before voiceless fricatives by 80.65 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 (1.54 *vs.* 1.09).

5. 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 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 (Port & O'Dell, 1985). 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.

Finally, 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^h/ but only 2 contrastive fricatives: /s, s:/. In addition, when comparing the Korean systems with the English sys-

tems, 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 tensity). In contrast, as Avery & Idsardi (2000) mentioned, the Korean fricative system employs a different contrast (length) than that of English (Glottal Width). So in order to learn the English fricative system, Korean learners must both suppress the use of length and discover that Glottal Width 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.

6. 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.

References

Avery, P. & Idsardi, W. 2000. Laryngeal dimensions, completion and enhancement. ms., York University and University of Delaware.

Cho, T. H. 1996. Vowel correlates to consonant phonation: An acoustic-perceptual study of Korean obstruents. Arlington, TX: M.A. Thesis, University of Texas at Arlington.

Curtin, S., Goad, H. & Pater, J. 1998. Phonological transfer and levels of representation: the perceptual acquisition of Thai voice and aspiration by English and French speakers. *Second Language Research*, 14, 389-405.

Denes, P. 1955. Effect of duration on the perception of voicing. *Journal of the Acoustical Society of America*, 27, 761-764.

House A. S. & Fairbanks, G. 1953. The influence of consonant environment upon the sec-

ondary acoustical characteristics of vowels, *Journal of the Acoustical Society of America*, 25, 103-113.

House, A. S. 1961. On vowel duration in English. *Journal of the Acoustical Society of America*, 33, 1174-1178.

Krause, S. E. 1982. Vowel duration as a perceptual cue to post-vocalic voicing in young children and adults. *Journal of the Acoustical Society of America*, 71, 990-995.

Kuijpers, C. T. L. 1996. Perception of the voicing contrast by Dutch children and adults. *Journal of Phonetics*, 24, 367-382.

Lehiste, Llse. 1970. *Suprasegmentals*. Cambridge. MA: M.I.T. press.

Lehman, M. E. & Sharf, D. J. 1989. Perception/production relationships in the development of the vowel duration cue to final consonant voicing, *Journal of Speech and Hearing Research*, 32, 803-815.

Lisker, L. & Abramson, A. S. 1964. A cross-language study of voicing in initial stops: acoustical measurement. *Word*, 20, 384-442.

Naeser, M. A. 1970. The American child's acquisition of differential vowel duration, Doctoral dissertation, University of Wisconsin, Madison, WI.

Peterson, G. E. & Lehiste, I. 1960. Duration of syllabic nuclei in English, *Journal of the Acoustical Society of America*, 32, 693-703.

Port, R. F. & O'Dell, M. L. 1985. Neutralization of syllable-final voicing in German, *Journal of Phonetics*, 13, 455-471.

Raphael, L. J. 1972. Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English, *Journal of the Acoustical Society of America*, 51, 1296-1303.

Smith, Caroline L. 1997. The devoicing of /z/ in American English: effects of local and prosodic context. *Journal of Phonetics*, 25, 471-500.

Weismer, G., Dimmsen, D. & Elbert, M. 1981. A study of the voicing distinction associated with omitted, word-final stops, *Journal of Speech and Hearing Disorders*, 46, 320-328

Zimmerman, S. A. & Sapon, S. M. 1958. Note on vowel duration seen cross-linguistically. *Journal of the Acoustical Society of America*, 30. 152-153.

received: May 2, 2006
accepted: May 24, 2006

▲ Woohyeok Chang
Department of English, Dankook University
San #29, Anseo-dong, Cheonan-si, Choongnam, Korea 330-714
Tel: +82-41-550-3139
E-mail: woohyeok@dankook.ac.kr

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.