

Perception of English High Vowels : Duration as a Cue by Korean Speakers of English*

Ji-Yeon Lee
University of Kansas

Abstract

This study examines the acoustic characteristics of English high tense and lax vowels (/i, I, u, U/) and the perceptual difference of those vowels between English speakers and Korean speakers of English. In Experiment I, the first three formant values and duration of the four vowels in /hVd/ carrier were measured. The result shows that the high tense vowels have larger F2–F1 values and longer durations than high lax vowels. In Experiment II, each word's vowel duration was manipulated into the range from 170ms to 290ms in 30ms increments. Two English speakers and six Korean speakers of English were asked to listen to a pair of tense and lax vowel words (e.g., *heed*_230ms–*hid*_170ms for the front vowel test and *hood*_290ms –*who'd*_200ms for the back vowel test), and to discriminate the pair by choosing one of the options either *heed-hid* or *hid-heed* for front vowels and either *who'd-hood* or *hood-who'd* for back vowels. The results demonstrate that English speakers discriminated tense vowels from lax vowels 100% correctly regardless of the different durations, compared to 62% accuracy rate in Korean speakers of English. Most errors were found when the duration of lax vowels was lengthened and the duration of tense vowels was shortened. Therefore, the results of this study demonstrate that Korean speakers mainly rely on vowel duration as a cue to discriminate the tense and lax vowels. The pedagogical implications of this phenomenon will be discussed.

1. Introduction

One of the misconceptions about English pronunciation in Korean English classrooms is that English high tense vowels are different from high lax vowels only in terms of duration. In other words, English high front tense vowel /i/ is regarded as a long high front lax vowel /ɪ/, and high back lax vowel /ʊ/ is considered as a shorter version of high back tense vowel /u/. Korean students are frequently taught that they can make /ɪ/ sound if they pronounce /i/ shorter, and that this is also true of the high back vowels. Moreover, English-Korean dictionaries transcribe the pronunciation of those vowels in this manner (e.g., *heed* is transcribed as [hi:d] and *hid* as [hid]).

Given that Korean students learn English as a foreign language, in which classroom instruction and dictionaries are the most influential sources on their English learning, it is reasonable to assume that they would rely mainly on duration as a cue to discriminate English high tense vowels from lax counterparts. Therefore, in this study two experiments are conducted in order to answer the two following research questions: (1) How English high tense vowels are acoustically different from lax vowels? (2) Do English speakers and Korean speakers of English perceive these differently? If so, it is hypothesized that Korean learners of English use mainly duration as a cue to discriminate those vowels. Experiment I examines the acoustic

* I deeply appreciate Dr. Allard Jongman's thoughtful advice and comments throughout this study.

characteristics of the four English high vowels using Praat software. Experiment II performs a perceptual experiment using the stimuli manipulated with vowel durations.

2. Previous Research

Hillenbrand, *et al.* (2000) summarizes several studies on vowel duration measurement which show similar patterns of duration across the twelve English vowel categories despite differences in absolute duration due to different carrier words. These studies report consistently longer durations for English high tense vowels than lax counterparts.

The role of duration in English vowel perception also has been examined. In cases of English native speakers, Hillenbrand *et al.* (2000) claimed that duration had a small overall effect on vowel identity since the most of the signals were identified correctly not only at the original durations but also at all three altered durations of 144 ms, 272 ms, and 400ms. He mentioned that the vowel contrasts that vary systematically in duration, such as /i/-/ɪ/ and /u/-/ʊ/, were minimally affected by duration.

However, non-native speakers of English were affected by the duration on vowel identification. Flege, *et al.* (1997) conducted a perception experiment with experienced /inexperienced German, Mandarin, Korean and Spanish speakers of English using *beat-bit* and *bat-bet* spectral continuum with three different durations. The statistical analysis showed that with the *beat-bit* continuum, the inexperienced Koreans, the experienced Koreans, and the inexperienced native Mandarin subjects had significantly smaller spectral effect scores than did the English native subjects, and had significantly larger temporal effect scores than did the native English subjects. In other words, their responses were mostly affected by different durations of the stimuli rather than spectral qualities. When they listened to a long stimulus, they identified it as a tense vowel word and a short one as a lax vowel word.

3. Experiment I--acoustic measurement

3.1. Stimuli

Four words (*heed, hid, who'd, and hood*) were produced by an adult male native speaker of English who did not have a known history of either speech or hearing disorders. The /hVd/ word set was chosen based on two reasons: one is because the vowel does not exhibit coarticulatory effects of the preceding consonant, /h/, and alveolar /d/ also has relatively little influence on the formants of the preceding vowel (Yang, 1996). Additionally these words have similar frequency of English usage (Fancis & Kucera, 1982). The frequency was considered because these stimuli also would be used in the perception experiment, where results could be influenced by the word frequency if not controlled. The stimuli were recorded in the Anechoic Chamber at the University of Kansas with Marantz PMD671 solid-state recorder and Electro-Voice RE20 microphone. The recording was stored as wave files for the purpose of analysis using the Praat software program.

3.2. Procedure

Each word was displayed in the waveform and spectrogram using Praat software program. For each stimulus, two properties were measured: the first three formants of the vowels and the

duration of the each phonetic segment (i.e. /h/, vowel, and /d/). The formant values were taken in the middle of the each vowel. As for the vowel duration, vowel onset was taken to be the onset of periodicity in the waveform, and vowel offset was indicated by the loss of this periodicity. Consonant duration was defined from the onset of distinct darkness to the vowel onset for in case of /h/, and from the vowel offset to loss of darkness in case of /d/.

3.3. Result

As shown in Figure1, the four words presented different spectrograms.

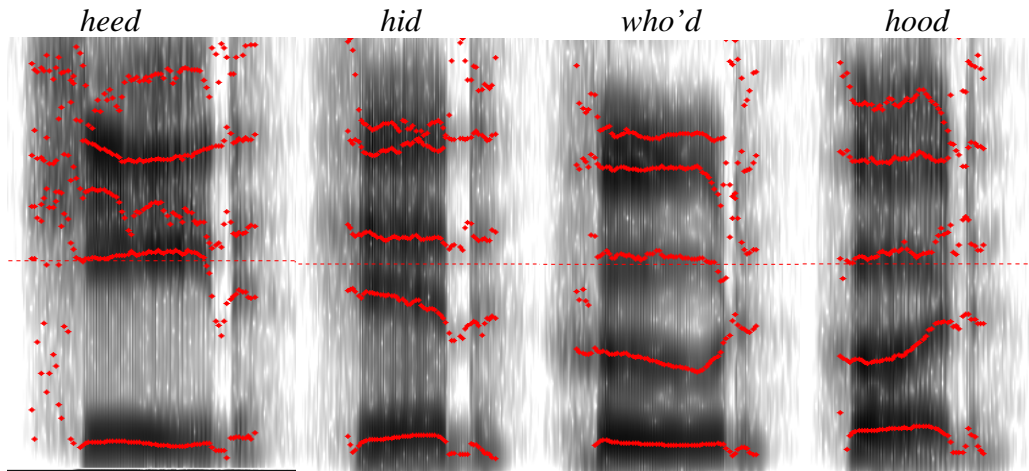


Figure 1: Spectrogram of *heed*, *hid*, *who'd*, and *hood* with formant lines in red

The values of first three formants are shown in Table1. The F2-F1 value was additionally calculated in order to investigate the difference of the vowels between tense vowels and lax vowels regardless of the frontness of the vowels. When they were compared as a pair of /i/-/ɪ/ and /u/-/ʊ/, tense vowels have larger F2-F1 values than lax vowels (i.e., /i/=2133Hz > /ɪ/ = 1515 Hz, and /u/=1013 Hz > /ʊ/= 776 Hz).

	F1	F2	F3	F2-F1
/i/	330	2463	2988	2133
/ɪ/	424	1939	2662	1515
/u/	379	1392	2526	1013
/ʊ/	509	1285	2524	776

Table 1: Formant frequency values of the English high vowels in Hz

Figure 2 illustrates the durations of each segment of the four stimuli. As expected, tense vowels are longer approximately by 60 ms than lax vowels. Moreover, /h/ segments preceding the tense vowels are pronounced longer than that preceding the lax ones.

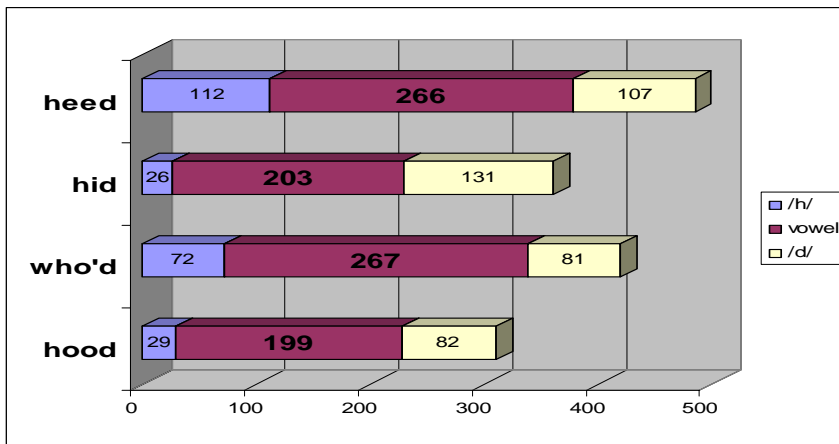


Figure 2: Duration of each segment of *heed*, *hid*, *who'd*, *hood* in ms

4. Experiment II--perceptual experiment

4.1. Subject

Two male English native speakers and six Korean learners of English at the University of Kansas participated in the perceptual experiment. Table 2 shows the Korean subject's background information including gender, length of residency in the U.S. in years, and age of arrival in the U.S. No subject has a known history of hearing impairment.

Subject	Gender	Length of residence in the U.S.	Age of arrival in the U.S.
K1	Male	2 years	34 years
K2	Female	5 years	16 years
K3	Male	5 years	16 years
K4	Male	3.5 years	28 years
K5 ¹	Female	10 months	37 years
K6	Male	4 years	31 years

Table 2: Background information of the six Korean subjects

4.2. Stimuli

The four English words of *heed*, *hid*, *who'd*, and *hood* described earlier were saved as an individual wave file respectively. Vowel duration of each word was manipulated to create five kinds of durations from 170--290ms in a 30 ms increment using Praat software². The manipulation procedures used to create a 290ms *hid* sound file from the original 200 ms file are summarized in Figure 3.

¹ Even though K5 has been in the U.S. for the shortest time among the Korean subjects, it should be noted that she is an English teacher for more than 10 years in Korea.

² I thank Yuwen Lai for instructing on sound duration manipulation with Praat software.

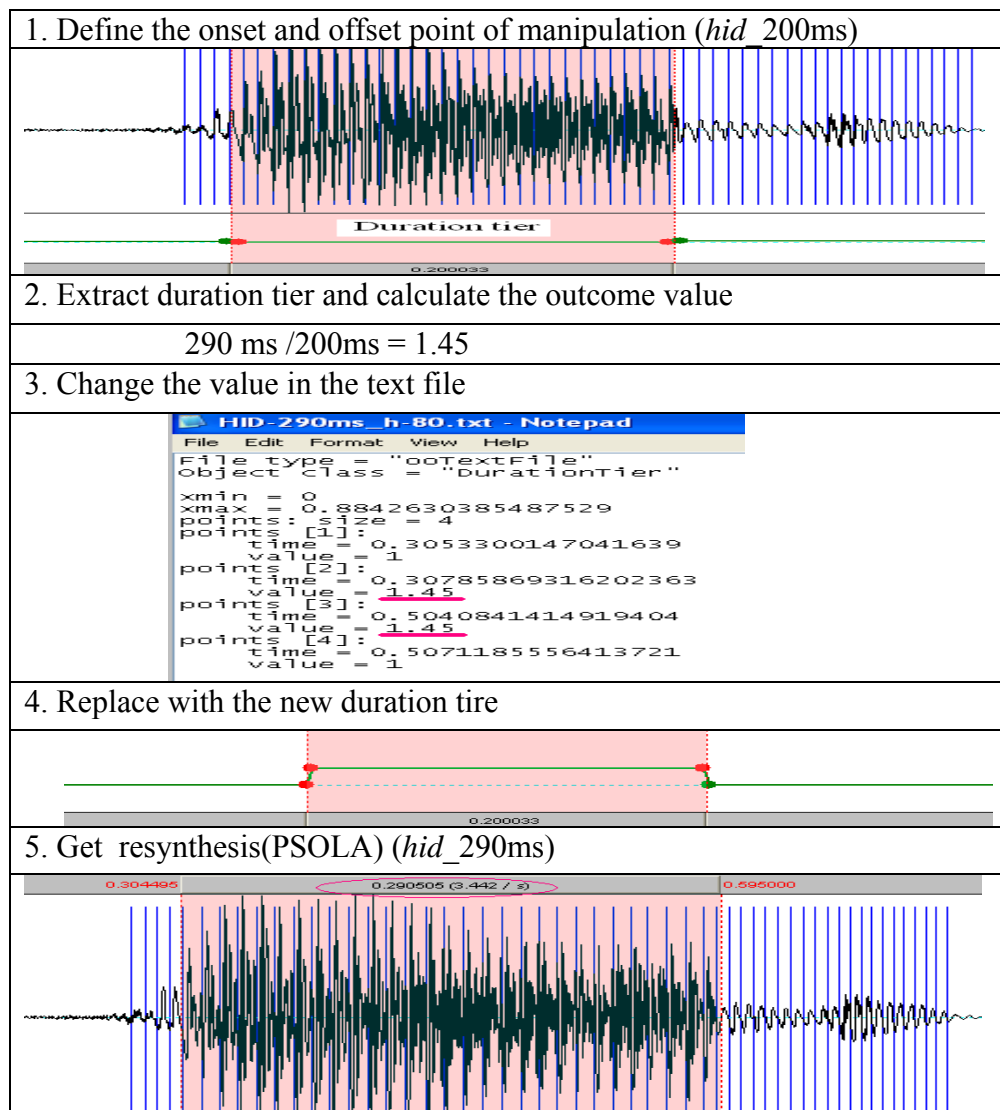


Figure 3: Procedures to manipulate duration using Praat software

In the same manner, /h/ segment durations were changed into the mean length of 80 ms for *heed* and *hid*, and 60ms for *who'd* and *hood* in order to make the /h/ sound neutral. From the Experiment I, it was found that the /h/ durations before the tense vowels are consistently longer than those before the lax vowels. Therefore, it is possible that /h/ duration affects the discrimination between the tense and lax vowels. Since the present study examines the effect of vowel duration on discriminating tense from lax vowels, it is necessary to control other possible cues.

After creating a set of five files for each word with ambiguous /h/ duration (i.e., 20 files total = 5 durations × 4 words), the “pair stimuli” were generated, which consisted of one file from the *heed* set and the other from the *hid* set for the front vowel pair stimuli. Similarly, one file from the *who'd* set and the other from the *hood* set comprised the high back vowel pair stimuli. The word order was alternated in order to prevent possible influence on response. In other words, *hid-heed* sets were produced as well as *heed-hid* sets, and not only *who'd-hood* sets but also

hood-who'd word order sets were created. A second file was attached to the firstly selected file with approximately 500–700 ms interval. Figure 4 shows waveforms of two pair stimuli as examples. A total of 100 “pair stimuli” were produced: 25 *heed-hid* sets (5 durations of *heed* × 5 duration of *hid*), 25 *hid-heed* sets, 25 *who'd-hood* sets, and 25 *hood-who'd* sets.

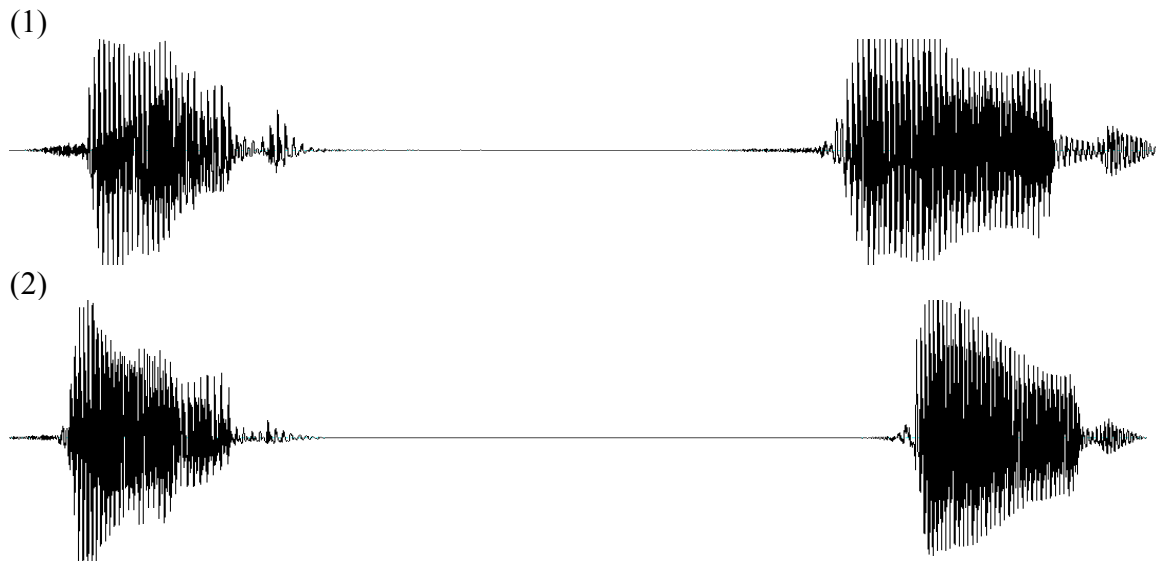


Figure 4: Sample waveforms of the “pair stimuli”
1) *heed*_170ms–*hid*_260ms 2) *hood*_230ms–*who'd*_230ms

Using the Paradigm perception experiment software, the pair stimuli were designed to be presented to the listener at random. Two separate tests were designed: Test 1 has 50 front vowel pair stimuli, and Test 2 has 50 back vowel pair stimuli.

4.3. Procedure

The experiments were performed in a quiet room with PCs. After brief explanation about the experiment, subjects signed the Human subject consent form. Based on the author’s observation that some Korean learners of English may not realize the pronunciation difference between *heed* and *hid* or between *who'd* and *hood*, English-Korean dictionary’s definitions and transcriptions of these words were provided to Korean subjects before the experiment.

The subjects wore headphones. They were instructed to make a response by clicking one of the mouse buttons. For example, in Test 1 they were asked to press left button when they thought that they heard in the order of *heed-hid*, and right button when they heard in the order of *hid-heed*. English native speakers (E1, E2) and the first three Korean subjects (K1, K2, K3) took Test 1 with *heed-hid/hid-heed* sets and Test 2 with *who'd-hood/hood-who'd* sets. The other three Korean subjects (K4, K5, K6) took the tests in the opposite order because it was found that all of the K1–3 yielded higher correct response rate in Test 2 than in Test 1, and a training effect was suspected for the result.

Before the actual test, they took a practice test with four sample pair stimuli, and had an opportunity to ask questions about the procedures. No feedback was provided during the tests,

and the reaction times were limited to 5 seconds. Upon test completion, results were saved as an Excel file. The correct response rates and the number of errors for each stimulus were examined.

4.4. Result

The correct response rates by the English and Korean subjects are presented in Figure 5. Both English subjects had 100% correct response rates, whereas Korean subjects' rates vary with 62 % mean value.

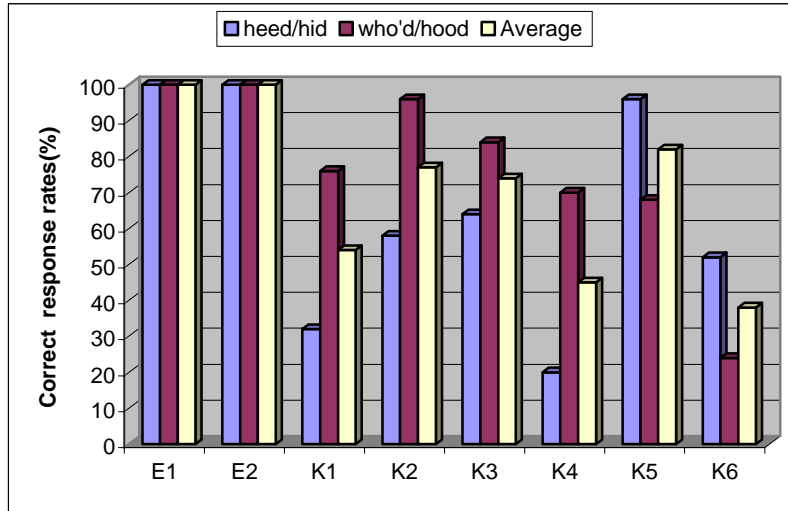


Figure 5: Correct response rate by two English (E1, E2) and six Korean (K1–K6) subjects.

Table 3 shows the number of errors for each pair stimuli in detail. Error values were summed, regardless of the presentation order. For example, a value of 10 in the green box means that the Korean subjects made 10 errors for the pair of *heed* _170 ms – *hid*_260 ms and the pair of *hid* _260 ms – *heed* _170ms. Not surprisingly, most errors were made for the pairs of shorten tense vowels and lengthen lax vowels.

<i>hid</i> (ms) \ <i>heed</i>	170	200	230	260	290	<i>hood</i> (ms) \ <i>who'd</i>	170	200	230	260	290
170	7	8	9	10	9	170	2	5	5	4	6
200	5	4	6	7	10	200	1	5	2	5	6
230	1	4	8	7	9	230	3	2	3	4	5
260	4	2	3	5	5	260	3	3	5	5	3
290	0	3	6	2	5	290	3	2	2	2	5

Table 3: The number of errors for each pair stimuli. The most number of errors possible is 12 (6 subjects X 2 alternative orders).

Further analyses were performed to confirm that the Korean subjects made errors due to overuse of durational cues. First of all, a “non-prototype value” was assigned to each pair stimuli as shown in Table 4. The value was calculated by adding up two “non-prototype value” of single stimuli marked in parentheses in Table 4. For example, the non-prototype value, 3, in the green box equals to the value of 1 (tense vowel_260ms) plus the value of 2 (lax vowel_230ms). In other words, the value becomes bigger when a pair stimulus is consisted of shorter tense vowels and longer lax vowels.

Lax (ms) \ Tense	170 (0)	200 (1)	230 (2)	260 (3)	290 (4)
170 (4)	4	5	6	7	8
200 (3)	3	4	5	6	7
230 (2)	2	3	4	5	6
260 (1)	1	2	3	4	5
290 (0)	0	1	2	3	4

Table 4: Non-prototype value of each pair stimuli

Figure 6 presents that the mean number of errors increases as the “non-prototype values” become larger from 0 to 8. Every number of errors in Table 3 was matched to a “non-prototype value” in Table 4. The errors of the same non-prototype value were summed and averaged. For example, the number of errors, 5 (marked with a green arrow), for the non-prototype value of 6, is the average of 5, 5, and 5 colored red in the Table 3.

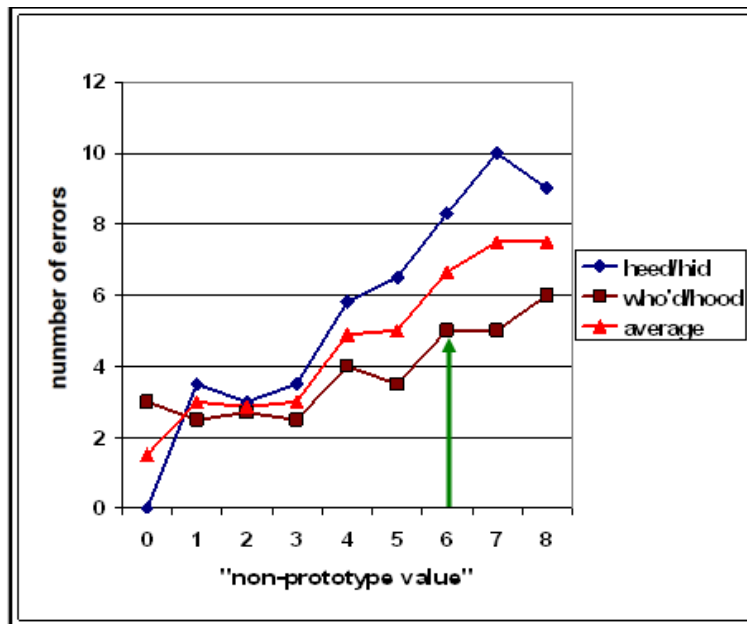


Figure 6: The mean number of errors of each non-prototype value

5. Discussion and Conclusion

Experiment I confirms that there are acoustic differences in the formant values and duration of the four English high vowels. Even though the stimuli were produced by only one speaker in the present study, the formant values and the durations of the four English high vowels fall in the range of other large scale studies such as Peterson and Barney (1952) and Hillenbrand, *et al.*(1995). Interestingly not only vowel durations but also /h/ segment durations become longer when followed by high tense vowels than when followed by high lax vowels.

The data from the Experiment II show that the spectral differences between English high tense vowels and lax vowels are perceptually clear to English native speakers. On the other hand, the result supports the hypothesis that Korean learners of English mainly use duration as a cue to discriminate English high tense vowels from corresponding lax vowels. It is worth noting that the stimuli used in previous studies (e.g., Flege *et al.*, 1997) on non-native speakers' perception are spectrally ambiguous, whereas this study used original spectral stimuli without any manipulation of the formant values. Therefore, the results of this study more strongly indicate than the previous studies that Korean speakers of English pay more attention to the vowel duration than to the vowel quality when they are asked to discriminate /i/ from /ɪ/ or /u/ from /ʊ/ because they still made errors even when the spectrum is not ambiguous.

The question remains why Korean subjects use duration as a cue rather than spectral difference between English high tense and lax vowels. From the literature review, two possibilities exist: (1) Equivalence classification (Flege, 1987), and (2) Desensitization hypothesis (Bohn, 1995). Flege (1987) defined equivalence classification as "a basic cognitive mechanism which permits humans to perceive constant categories in the face of the inherent sensory variability found in the many physical exemplars which may instantiate a category." In other words, Korean vowel inventory does not have /i/-/ɪ/ or /u/-/ʊ/ distinction even though /i/ and /u/ are available in Korean. Thus, Korean learners of English categorize both English high front vowels into one Korean high front vowel /i/ and both back vowels into one Korean high back vowel /u/. Then, they rely on duration cues (as they learned in school) when asked to discriminate the tense vowels from lax vowels. Additionally, since these vowels are considered "similar" to their L1 in terms of Flege (1987)'s classification, Korean learners would have difficulty in mastering them.

On the other hand, based on the results of previous studies, Bohn (1995) claimed that using duration as a cue in vowel perception by non-native speakers of English can not be explained solely by L1 transfer because it was observed that the native Spanish and Mandarin listeners also relied on a durational cues, that are not used to differentiate vowel contrasts in their native language. Therefore, he suggested that duration cues were used in vowel perception because they are easier to access than spectral cues if listeners have been desensitized to spectral differences in a particular area of the vowel space.

From an EFL teacher's point of view, another question can be raised: whether this practice of using duration cues in English vowel perception is important when learning English. One could argue that it should not be problematic whether they use duration as a cue or quality as a cue as long as English learners are able to discriminate the vowels. However, it could be problematic when the lax vowel is stressed and if this perception strategy is transferred to the production. When a vowel is stressed, its pitch becomes higher, it sounds louder, and its duration becomes longer. Therefore, it is possible for nonnative speakers to misperceive, for example, stressed *hid*

as *heed* or stressed *hood* as *who'd*. Given that English has a large number of minimal pairs of high tense and lax vowels, relying on duration as a cue would be not a perfect strategy. Moreover, it is very likely that Korean learners of English produce tense vowel and lax vowel differently only in terms of duration in the same way they perceive them. Consequently, it is reasonable to claim that the spectral differences as well as durational differences between English high tense vs. lax vowels should be explicitly taught to non-native speakers of English for intelligible and accurate communication in English.

References

- Bohn, O.-S. (1995). Cross-language speech perception in adults: first language transfer doesn't tell it all. In a W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Baltimore: York Press. 279-304.
- Flege, J.E. (1987). The production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics* 15: 47-65
- Flege, J.E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics* 25: 437-470.
- Francis, W.N. & Kucera, H. (1982). *Frequency analysis of English usage: lexicon and grammar*. Boston: Houghton Mifflin.
- Hillenbrand, J., Getty, L.A., Clark, M.J., and Wheeler, K.(1995). Acoustic characteristics of American English vowels. *Journal of Acoustical Society of America* 97: 3099-3111
- Hillenbrand, J.M., & Clark, M.J.(2000). Some effects of duration on vowel recognition. *Journal of Acoustical Society of America* 108: 3013-3022
- Peterson, G. E., and Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of Acoustical Society of America* 24: 175-184.
- Yang, B. (1996) A comparative study of American English and Korean vowels produced by male and female speakers. *Journal of Phonetics* 25: 245-262.