Perception of English High Vowels by Korean Speakers of English

Lee, Ji-Yeon¹⁾

ABSTRACT

This study compares the perception of English high tense and lax vowels (/i, I, u, U/) by English speakers and Korean speakers of English. The four vowels were produced in /hVd/ context by a native speaker of English, and each word's vowel duration was manipulated to range from 170ms to 290ms in 30ms increments. Two English speakers and six Korean speakers of English were asked to listen to pairs of tense and lax vowel words with manipulated vowel durations and to identify the pair by choosing either heed-hid or hid-heed for front vowels and either who'd-hood or hood-who'd for back vowels. The results show that English speakers distinguished tense vowels from lax vowels with 100% accuracy regardless of the different durations, compared to 62% accuracy for Korean speakers of English. Most errors occurred for lengthened lax vowels and shortened tense vowels. 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 theoretical and pedagogical implications of this finding are discussed.

Keywords: English tense and lax vowels, perception, duration, Korean speakers of English

1. Introduction

English high tense vowels differ from their lax counterparts both in quality and duration (e.g., Peterson & Barney, 1952; Hillenbrand et al., 1995). The high tense vowels, /i/ and /u/, sound longer than the high lax vowels, /I/ and /U/, and the formants of the tense vowels are also different from those of the lax counterparts. The distance between the first formant (F1) and the second formant (F2) is greater for the high tense vowels than for the high lax ones. However, previous studies on English vowel production by Korean speakers showed that Korean speakers do not make a clear quality distinction between English high tense and lax vowel pairs, /i/- /I/ and /u/-/U/, when they pronounce these vowels (Koo, 2000, 2005; Yang 2008).

In fact, a common misconception about English pronunciation

1) University of Kansas, jylee9@ku.edu

Received: August 1, 2009 Revision received: November 9, 2009 Accepted: November 28, 2009 by Korean speakers is that English high tense vowels are different from high lax vowels only in terms of duration without being aware that they are also different in quality. Korean speakers commonly perceive the English high front lax vowel /I/ as a short high front tense vowel /i/, and the high back lax vowel /u/ as a short high back tense vowel /u/. This is what Korean students are taught about English high vowels in English classrooms: that they can make the /I/ sound by pronouncing it like a shorter /i/ sound, and that they can make the /o/ sound by pronouncing it like a shorter /u/ sound (Kim, 2004). Moreover, English-Korean dictionaries transcribe the pronunciation of those vowels in the same manner. For example, heed is transcribed as [hi:d] and hid as [hid], and who'd is transcribed as [hu:d] and hood as [hud] (Dong-a prime English dictionary, 1990; Naver internet English dictionary). These dictionaries use the same phonetic symbol, [i], for both heed and hid and [u] for both who'd and hood, and they only distinguish them with the length mark, [:]. Kang(2001) also reported that best-selling books on English pronunciation in Korea do not properly explain the quality difference between English high tense and lax vowels.

Since Korean learners of English are taught that English tense and lax vowels only differ in terms of duration, Korean students treat the two different sounds as longer and shorter versions of the same vowel. This misconception leads to the assumption that Korean students rely mainly on duration, not quality or formant differences, as a cue to discriminate English high tense vowels from their lax counterparts, because that is what they learn in the classroom and what they find in the dictionary.

The purpose of this study was to investigate if English speakers and Korean speakers of English perceive these vowels differently. It was hypothesized that Korean learners of English use mainly duration as a cue to discriminate English high vowels, whereas English speakers mainly use vowel quality.

2. Previous studies

Hillenbrand *et al.* (2000) summarized several studies on vowel duration that showed similar patterns of duration across the twelve English vowel categories despite their differences in absolute duration due to different carrier words. These studies reported consistently longer durations for English high tense vowels than for their lax counterparts. The role of duration in English vowel perception also has been examined. Hillenbrand *et al.*(2000) claimed that duration had a small overall effect on vowel identity among English native speakers since most of the vowels were identified correctly not only at the original durations but also at all three altered durations of 144 ms, 272 ms, and 400ms. Interestingly, perception of the vowel contrasts that vary systematically in duration, such as /i/-/I/ and /u/-/o/, was minimally affected by different durations among English native speakers.

In contrast, non-native speakers of English are affected by duration in vowel identification. Flege *et al.* (1997) conducted a perception experiment with both experienced and inexperienced German, Mandarin, Korean and Spanish speakers of English by using *beat-bit* and *bat-bet* spectral continua with three different durations. With the *beat-bit* continuum, the statistical analysis showed that the inexperienced Korean, the experienced Korean, and the inexperienced Mandarin subjects had significantly smaller spectral effect scores than the native English subjects. In addition, they had significantly larger temporal effect scores than the native

English subjects. In other words, their responses were mostly affected by different durations of the stimuli and only minimally by their spectral qualities. Non-native speakers of English identified long vowels as tense and short vowels as lax.

Unlike previous studies, the present study examines the perception of English high vowels by Korean speakers by presenting a *pair* of tense and lax vowels, with systematically manipulated durations. If Korean speakers do not properly weigh the quality difference between the two vowels, they may misidentify a long lax vowel as a tense vowel and a short tense vowel as a lax vowel even when they listen to the different vowels in a pair, not one at a time. Therefore, this study conducted a perception experiment using pairs of English high tense and lax vowel words, such as *heed-hid*, and *who'd-hood*, which had been manipulated to vary in vowel duration.

3. Method

3.1 Subject

Two male English native speakers and six Korean learners of English at a University in the United States participated in the perceptual experiment. <Table 1> shows the Korean subjects' background information including gender, length of residency in the United States in years, and age of arrival in the United States. No subject had a known history of hearing impairment.

Table 1: Background information of the six Korean subjects

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

3.2. Stimuli

Four words (heed, hid, who'd, and hood) were produced at a normal speech rate without a carrier sentence 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: 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 a similar frequency of English usage (Francis & Kucera, 1982). The frequency was considered because these stimuli were to be used in the perception experiment, where results could be influenced by word frequency. The stimuli were recorded in an Anechoic Chamber at a university in the United States with a Marantz PMD671 solid-state recorder and an Electro-Voice RE20 microphone. The recordings were saved as wave files for the purpose of analysis using Praat (Boersma and Weenink, 2005).

In order to confirm that these words' acoustic characteristics fell within the range of typical English high vowels, two properties of each word were measured using Praat: 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 each vowel using wide-band spectrograms. As for vowel duration, vowel onset was taken to be the onset of F1 in the spectrogram, and vowel offset was indicated by the loss of F2. /h/ duration was defined from the onset of high frequency energy to the vowel onset: /d/ duration was defined from the vowel offset to the offset of high frequency energy. As shown in <Figure 1>, the four words presented different spectrograms.

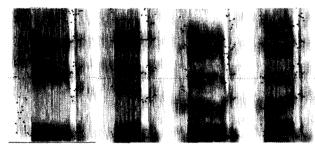


Figure 1. Spectrograms of (from left to right) *heed, hid, who'd,* and *hood* with formant lines.

The values of the first three formants are shown in <Table2>. The F2-F1 value was additionally calculated in order to investigate the difference between tense vowels and lax vowels regardless of the frontness of the vowels. When they were compared as a pair of /i/-/I/ and /u/-/ U/, tense vowels had larger F2-F1 values than lax vowels (i.e. /i/=2133Hz > /I/ = 1515 Hz, and /u/=1013 Hz > /U/=776 Hz).

Table 2: Formant frequency values (Hz) of the English high vowels

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

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

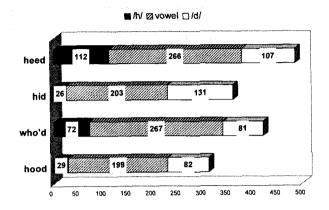


Figure 2. Duration (ms) of each segment of heed, hid, who'd, hood

These measurements confirm that there are acoustic differences in both the formant values and duration of the four English high vowels. Even though only one speaker produced all the stimuli for this study, the formant values and the durations of the four English high vowels fall within 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 are longer when followed by high tense vowels as compared to high lax vowels.

Table 3. Five different durations for each word

heed	hid	who'd	hood
170 ms	170 ms	170 ms	170 ms
	200 ms		200 ms
200 ms	(original	200 ms	(original
	duration)		duration)
230 ms	230 ms	230 ms	230 ms
260 ms		260 ms	
(original	260 ms	(original	260 ms
duration)		duration)	
290 ms	290 ms	290 ms	290ms

These four English words *heed, hid, who'd,* and *hood* were saved as individual wave files. The vowel duration of each word was manipulated to create five durations ranging from 170 to 290 ms in 30 ms increments, as shown in <Table 3>, using Praat software.

The manipulation procedures used to create, for example, a 290ms *hid* sound file from the original 200 ms file are summarized in Figure 3. Using Praat software, the onset and the

offset points of the vowel duration were defined, and the duration tier was extracted and saved as a text file. Since the target duration of 290 ms is 1.45 times the original duration of 200 ms, the values in the text file were changed from 1 to 1.45. The new values were saved, and this text file was substituted for the original duration tier, which resulted in creating a 290ms hid sound file. The same procedure with different values depending on the target durations was applied to create all 20 stimuli (4 words \times 5 different durations).

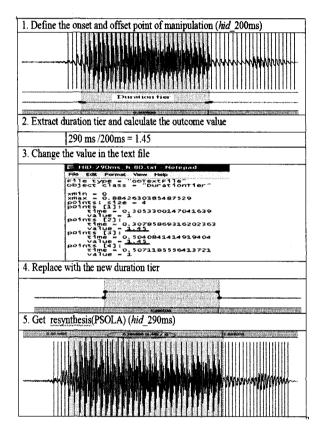


Figure 3. Procedures to manipulate duration using Praat software

Our acoustic analysis demonstrated that the /h/ durations before the tense vowels were 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 sought to assess the effect of vowel duration on the perception of tense and lax vowels, it was necessary to control other possible cues. /h/ segment durations were therefore changed to the mean duration of 80 ms for heed and hid, and 60 ms for who'd and hood in the same manner as described above.

In sum, 20 stimuli were created, consisting of 4 target words with 5 different vowel durations. Next, stimulus pairs were generated, which consisted of one file from the *heed* set and the

other from the *hid* set for the front vowel stimulus pairs. Similarly, one file from the *who'd* set and the other from the *hood* set comprised the high back vowel stimulus pairs. In order to prevent possible response bias, the word order was counterbalanced. In other words, *hid-heed* sets were produced as well as *heed-hid* sets, and not only *who'd-hood* word order sets were created but also *hood-who'd* word order sets. Intervals of approximately 500-700 ms were inserted between the two members of a pair. <Figure 4> shows waveforms of two stimulus pairs as examples. A total of 100 pairs 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.

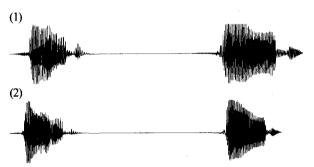


Figure 4. Sample waveforms of the stimulus pairs

1) heed_170ms hid_260ms
2) hood 230ms who'd 230ms

Using Paradigm perception experiment software, the stimulus pairs were presented to listeners in random order. Two separate blocks were designed: Block 1 had 50 front vowel stimulus pairs, and Block 2 had 50 back vowel stimulus pairs.

3.3 Procedure

The experiment was conducted in a quiet room. 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 definitions and transcriptions of these words were provided to Korean subjects before the experiment.

The subjects wore headphones. They were instructed to respond by clicking one of the mouse buttons. For example, in Test 1, they were asked to press the left button when they thought that they heard the pair in the order of *heed-hid*, and the right button when they heard *hid-heed*. English native speakers (E1, E2) and the first three Korean subjects (K1, K2, K3) took Block 1 with *heed-hid/hid-heed* sets and Block 2 with *who'd-hood/hood-who'd* sets. The other three Korean subjects (K4, K5, K6) took the tests

in the opposite order.

Before the actual test, all subjects took a practice test with four sample stimulus pairs and had the opportunity to ask questions about the procedure. No feedback was provided during the tests, and the response time was limited to 5 seconds. Upon test completion, results were saved in Microsoft Excel. The correct response rates and the number of errors for each stimulus were examined.

4. Results

The correct response rates for the English and Korean subjects are presented in <Figure 5>. Both English subjects (E1, E2) had 100% correct response rates, whereas Korean subjects' rates (K1-K6) varied, with 62 % on average. Korean subjects' accuracy was 53.7% for the front vowel pairs, and 69.7 % for the back vowel pairs. A paired samples T test showed that the number of errors for the front and back vowel pairs was not significantly different t(5)=1.103, p=0.32.

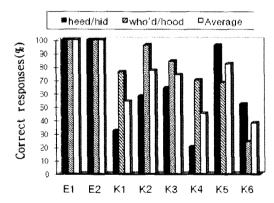


Figure 5.Correct response rates for the two English (E1, E2) and six Korean (K1–K6) subjects

<Table 4> shows the number of errors for each stimulus pair in detail. Error values were summed, regardless of the presentation order. For example, a value of 10 in the shaded 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 shortened tense vowels and lengthened lax vowels. For example, Korean speakers made the most (ten) errors each for heed _170 ms -hid_260 / hid _260 ms - heed _170ms and for heed_200ms - hid_290 / hid_290 ms -heed_200ms.

Table 4. The number of errors for each stimulus pair (Front vowel pairs on top and back vowel pairs at the bottom). The maximum number of errors possible is 12 (6 subjects X 2 alternative orders).

		1			
hid (ms) heed	170	200	230	260	290
170	7	8	9	10	9
200	5	4	6	7	10
230	1	4	8	7	9
260	4	2	3	5	5
290	0	3	6	2	5

hood (ms) who'd	170	200	230	260	290
170	2	5	(5)	4	6
200	1	5	2	(5)	6
230	3	2	3	4	(5)
260	3	3	5	5	3
290	3	2	2	2	5

Further analyses were performed to confirm that the Korean subjects made errors, indeed, because they overused durational cues. First of all, a "non-typical value" was assigned to each stimulus. Since tense vowels are typically longer than lax vowels, the shorter a tense vowel is, the less typical the stimulus is. Likewise, the longer a lax vowel is, the less typical the stimulus is. Therefore, as shown in <Table 5>, the biggest "non-typical value" of 5 was assigned to the shortest tense vowel stimuli (heed 170 ms and who'd 170 ms respectively) and to the longest lax vowel stimuli (hid 290 ms and hood 290ms respectively). The smallest "non-typical value" of 1 was assigned to the other endpoints and the remaining values were also assigned accordingly. This "nontypical value" of each stimulus marked in parentheses in <Table 5> was added up to generate the "non-typical value" of stimulus pairs. For example, the non-typical value of 3 in the shaded box is the sum of the value of 1 (tense vowel 260ms) and the value of 2 (lax vowel 230ms). In other words, this value becomes bigger when a stimulus pair consists of shorter tense vowels and longer lax vowels.

Table 5. Non-typical value of each pair stimuli

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	l	2	3	4

<Figure 6> shows that the mean number of errors increases as the "non-typical values" increase from 0 to 8. Each number of errors in <Table 4> was matched with a "non-typical value" in <Table 5>. The errors of the same non-typical value were summed and averaged. For example, the number of errors, 5 (marked with the arrow), for the non-typical value of 6, is the average of 5, 5, and 5 in parentheses in Table 4. The correlation between non-typical value and average number of errors was 97.3%, which is very high.

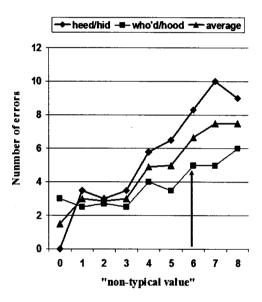


Figure 6. The mean number of errors of each non-typical value (see text)

In sum, the results indicated that the native speakers of English distinguished tense vowels from lax vowels with 100% accuracy regardless of the different durations. By contrast, Korean speakers of English only had a 62% correct response rate. Most errors occurred among the Korean speakers of English when lax vowels were lengthened and tense vowels were shortened. Korean speakers of English misidentified the sounds when they perceived long vowel sounds as tense vowels even though that was a lengthened lax vowel, and short ones as lax vowels even though that was a shortened tense vowel, without paying attention to the quality differences. Therefore, the findings of this study demonstrate that Korean speakers mainly rely on vowel duration as a cue to discriminate the tense from the lax vowels.

5. Discussion and Conclusion

The results show that the spectral differences between English high tense and lax vowels are perceptually clear to English native speakers. The results also support the hypothesis that Korean learners of English mainly use duration as a cue to discriminate English high tense vowels, /i/ and /u/, from corresponding lax vowels, /I/ and /U/. Korean speakers of English thus misidentified the sounds when they perceived long vowel sounds as tense vowels, although they were lengthened lax vowels, and they misidentified short vowels as lax vowels, although they were shortened tense vowels. Apparently, they did not pay attention to the qualitative differences between these sounds.

It is worth noting that previous studies on the perception of English tense-lax vowels among non-native speakers (e.g. Flege et al, 1997) used stimuli that were spectrally ambiguous, whereas this study used original spectral stimuli without any manipulation of the formant values. Because the subjects still made errors even though the spectrum was not ambiguous, the results of this study strongly indicate that Korean speakers of English pay more attention to vowel duration than to vowel quality when they are asked to discriminate /i/ from /I/ and /u/ from /O/.

The question remains: why do Korean subjects use duration rather than spectral differences as cues to discriminate between English high tense and lax vowels? Based on the existing literature, there are two possible answers to the question: (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 (p.49)". In other words, the Korean vowel inventory does not have /i/-/I/ or /u/-/U/ distinctions even though /i/ and /u/ are available in Korean. Thus, Korean learners of English seem to categorize both English high front vowels as one Korean high front vowel /i/, and they seem to categorize both back vowels as one Korean high back vowel /u/. Then, when they are asked to discriminate the tense vowels from lax vowels, they rely on the duration cues as they learned in school. Moreover, Korean learners seem to have difficulty mastering these vowels because, based on Flege's (1987) classification, they are "similar"

In contrast to Flege's results, Bohn (1995) claimed that the use

of duration as a cue in vowel perception by non-native speakers of English cannot be explained solely by L1 transfer. Bohn argued that it was observed that the native Spanish and Mandarin listeners also relied on durational cues to differentiate vowel contrasts, which are not used in their native language. Bohn thus suggested that listeners use duration cues 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, one could argue whether this practice of using duration cues in English vowel perception is important or not when learning English. One could argue that it should not be problematic whether nonnative speakers of English use either duration or quality as a cue as long as they are able to discriminate the target vowels. Even though Korean students are not aware of the quality difference between the English high tense and lax vowels, they still may be able to discriminate between those sounds when they listen to them in everyday speech because tense vowels are usually longer than lax vowels in natural speech. Relying on duration cues may be a good strategy for Korean students to discriminate the English high vowels.

However, relying on duration cues would be problematic for three reasons. First, it is a misconception that English high tense vowels are just longer than their lax counterparts. Korean EFL teachers should realize this fact, and they should be made aware of this aspect of English pronunciation. Second, vowel durations can change in utterance. Especially, 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. Finally, the possibility exists that this perception strategy is transferred to the production when Korean students pronounce the target sounds. Given that production is closely related to perception (Jamieson & Morosan, 1986, Rochet, 1995, Flege, 1995), Korean learners of English might produce tense vowels and lax vowels differently only in terms of duration in the same way they perceive them. It is commonly observed that Korean speakers of English pronounce the English word slip as a shorter [slip], which sounds like sleep to native speakers of English, or full as a shorter [ful], which sounds like fool to native speakers of English. Consequently, relying only on duration may

not be a perfect strategy, and the spectral differences as well as durational differences between English high tense vs. lax vowels should be explicitly taught to Korean speakers of English for intelligible and accurate communication in English.

References

- Boersma, P. & Weenink, D. (2005). Praat (Version 4.3.12 for Windows) [Computer software]. Amsterdam: University of Amsterdam, Institute of Phonetic Sciences. Available from http://www.fon.hum.uva.nl/praat/
- Bohn, O. (1995). "Cross-language speech perception in adults: First language transfer doesn't tell it all." In W. Strange (Ed.), Speech perception and linguistic experience: Issues in cross-language research (pp. 279-304). Timonium, MD: York Press, Inc.
- Dong-a prime English dictionary. (1990). Seoul, Korea: I B D Ltd.
- 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*, Vol.15, No.1, pp. 47-65.
- Flege, J. E. (1995). "Second language speech learning: Theory, findings, and problems." In W. Strange (Ed.), Speech perception and linguistic experience: Issues in cross-language research (pp. 233-277). Timonium, MD: York Press, Inc.
- Flege, J. E., Bohn, O. & Jang, S. (1997). "Effects of experience on non-native speakers' production and perception of English vowels." *Journal of Phonetics*, Vol.25, No.4, pp. 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. & Wheeler, K.(1995).
 "Acoustic characteristics of American English vowels."
 Journal of Acoustical Society of America, Vol. 97, pp. 3099-3111
- Hillenbrand, J. M. & Clark, M. J.(2000). "Some effects of duration on vowel recognition." *Journal of Acoustical Society* of America, Vol. 108, pp. 3013-3022
- Jamieson, D. & Morosan, D. (1986). "Training non-speech contrasts in adults: Acquisition of the English /θ/ and /δ/ sounds by francophones." *Perception and Psychophysics*, Vol.40, pp. 205-215.
- Kang, N-J. (2001). "Problems with English pronunciation education and their solution." *Malsori*, Vol. 41, pp. 1-18.
- Kim, J-E. (2004). Native Language and Non-linguistic Influence on the Production of English Vowels by Speakers of Korean: An Acoustic Study. Doctoral dissertation, The University of Texas at Arlington.
- Koo, H-S. (2000). "Characteristics of English vowels spoken by Koreans." Speech Sciences, Vol.7, pp.99-108
- Koo, H-S.(2005). "A study of the pronunciation of English vowels between male and female speakers." Speech Sciences, Vol.12, pp.7-16
- Naver internet English dictionary. Retrieved March 3, 2007 from the Website: http://dic.search.naver.com/.
- Peterson, G. E. & Barney, H. L. (1952). "Control methods used in a study of the vowels." Journal of Acoustical Society of

America, Vol. 24, pp. 175-184.

- Rochet, B. L. (1995). "Perception and production of second-language speech sounds by adults." In W. Strange Ed. Speech perception and linguistic experience: Issues in cross-language research (pp. 379-410). Baltimore, MD: York Press.
- Yang, B. (1996). "A comparative study of American English and Korean vowels produced by male and female speakers." *Journal of Phonetics, Vol. 25*, pp. 245-262.
- Yang, B. (2008). "An acoustical comparison of English tense and lax vowels produced by Korean and American males." *Speech Sciences, Vol.15*, pp.19-27.

• Lee, Ji-Yeon

Department of East Asian Languages and Cultures University of Kansas Lawrence, KS 66044 U.S.A.

E-mail: jylee9@ku.edu