Formant Trajectories of English High Tense and Lax Vowels Produced by Korean and American Speakers*

Byunggon Yang
(Pusan National University)

Yang, Byunggon. 2010. **Formant Trajectories of English High Tense and Lax Vowels Produced by Korean and American Speakers.** *Korean Journal of Linguistics*, 35-2, 407-423. Previous studies on the pronunciation of English vowels reported that Korean learners had difficulty producing English tense and lax vowel pairs distinctively. The acoustic comparisons of those studies are mostly based on formant measurements at a single slice of a given vowel section. However, the English vowels usually show dynamic spectral changes across the segment, and only partial data on the vowel segment may be insufficient. The purpose of this paper is to compare the dynamic formant trajectories of English high tense and lax vowel pairs produced by twenty Korean and American males and females. Results showed that the American males and females produced the tense and lax pairs much more distinctly than the Korean counterparts did. Many fine-grained differences along the six measurement points were observed both in the formant trajectories and on the vowel space. These results suggest that more detailed analysis be required in the cross-linguistic comparison of English vowels. Also, the Korean speakers should pay more attention to the dynamic movements of formants in addition to the jaw and tongue positions in order to match those of the American speakers. **(Pusan National University)**

Key words: English phonetics, tense and lax vowels, formant values, formant trajectory, vowel space

1. Introduction

Traditionally the acoustical characteristics of vowels were represented by formant values (Hillenbrand, Getty, Clark, and Wheeler, 1995; Peterson and Barney, 1952; Stevens and House, 1963; Yang, 1990, 1992, 1996; Watson and Harrington, 1999). The formant values were usually determined once at the steadiest portion of the vowel. Peterson and

*This work was supported for two years by a Pusan National University Research Grant.

Barney (1952) noted that the formant values taken from a single section might not represent the acoustic qualities of the given vowel. He stated that the evaluation of the formant transitions at the initial and final segments "is a problem of major importance (p. 184)." From the acoustical analysis of 10 Canadian vowels spoken in isolation by 10 talkers, Nearey and Assmann (1986) attempted to show dynamic movements of vowel formants by determining the first formant (F1) and the second formant (F2) at the initial (20%) and final (80%) section of the vowel. Their results revealed that several monophthongal vowels showed long spectral changes as could be seen in the formant transition of diphthongs. There was a noticeable difference in the tense and lax vowel pairs. The formant transitions of the vowels /i, u/ were not great compared with those for the lax vowels /ɪ, ʊ/. Almost similar patterns of spectral changes for the vowels measured can be observed in the Hillenbrand et al. (1995)'s study from their Southern Michigan data. There seems to be a slight difference in the vowel space because the two studies adopted the origins of the abscissa and ordinate differently. To make up for the fewer measurement points, Nishi, Strange, Akahane-Yamada, Kubo, and Trent-Brown (2008) made three measurements at 25%, 50%, 75% temporal points of the vocalic duration in their cross-linguistic comparison of American English and Japanese. The present study also attempts to examine more detailed measurement of formant values of English high tense and lax vowels and to observe cross-linguistic differences.

Another aspect of the study of English vowels in addition to the vowel production is how native speakers perceive the vowels. Which part or what proportion of the vowel segment are important in the vowel perception? Results of perceptual studies may be helpful to decide an appropriate number of measurement points on the vowel segment produced. Many studies point out the importance of spectral change patterns. From the silent-center study, Jenkins, Strange, and Edman (1983) found higher error rate (13.2%) of vowel identification when the middle portion of the vowel segment was presented to the listeners than the case of the beginning and end portion (7.6%). The error rate of the latter study was comparable to the rate (6.9%) obtained when the whole vowel portion was presented. Nearey and Assmann (1986) found a similar error rate (14.4%) with the nucleus and off-glide portions of the vowel were presented to the listener compared to the rate (12.4%) of the whole vowel. Those studies indicate that the start and end of a vowel segment are important cues for the identification of English vowels. Hillenbrand and Nearey (1999) also experimented with three stimulus categories of flat formants, original formants, and natural conditions. Their results showed that spectral changes were important cues: the natural signal stimuli brought about 95.4% correct identification followed by those with the original formants (88.5%) and those with the flat formants (73.8%). All of those studies mentioned above reveal that any static, partial information of the vowel may not be sufficient to identify each given vowel correctly. In other words, the American English vowels can be better represented by fine-grained dynamic spectral changes or movements on phonetic vowel space.

Previous researches indicate that Koreans cannot distinguish the tense and lax pairs well though the pairs are acoustically different in duration and formant values (Cho, 2003; Edwards, 2003; Koo, 2000; Kent and Read, 2002; Ladefoged, 2001; Yang, 1992, 1996). There are not many studies that specifically focused on the pairs because most subjects produce those vowels with somewhat wider transition along the formant trajectories. That makes a systematic and comparable measurement at the sustained vowel segment almost impossible. After examining the formant values of the English vowels produced by Korean students with an intermediate level of oral fluency, Koo (2000) reported that the six Korean students could not tell the vowel pairs /i-ɪ/ and /u-ʊ/ clearly. He explained that the Korean students could not do so just because they did not imitate the jaw and tongue positions of the native speakers. He measured the formant values at one time point. Thus, some confounding factors from the subjective measurement points might be involved in the final collection of the data. Also, the inappropriate settings to obtain formant values of the back rounded vowels might lead to his distorted vowel space. For example, F1 and F2 of the vowel pair /u-ʊ/ converge and some speech analysis applications usually print out one formant value for the first two possible formants. In Cho (2003)'s study the acoustical formant values produced by five native speakers and forty Koreans living in Chungcheong Province were collected. They produced thirty words five times and he measured durations and formant values of the words. The results showed that the native speakers produced them with a qualitative difference but the Koreans could not tell the vowel pairs

/i-ɪ/ and /u-ʊ/ apart. He concluded that the Koreans had a pronunciation problem in terms of durations and vowel quality. He chose various vowels in various contexts, which must act as a confounding factor and have an influence on the result. Yang (1990, 1996) reviewed nonlinguistic factors in the acoustic measurements and compared the formant values of English vowels produced by forty American and Korean male and female speakers. He suggested that formant comparison be carefully made after controlling the nonlinguistic factors as much as possible, and more measurement points should be selected to capture the acoustical vowel features, i.e., four equidistant vowel points relative to the total duration of the vowel segment. That way he could find a systematic frequency shift at the second time point, which was employed for speaker normalization more efficiently.

The aim of this study is to acoustically examine the English tense and lax vowel pairs produced by twenty Korean and English male speakers. The formant measurements are made on six equidistant time points of the vowel segment so that spectral changes of each vowel can be properly captured. The method of this study may contribute to a comparison of the English vowels produced by native and nonnative speakers in detail.

2. Method

2.1 Subjects and Data Recording

Twenty subjects participated in the study. They consisted of four groups with five members each: American male and female, Korean male and female. They were healthy graduate students of the University of Texas at Austin without any reported history of hearing problems. They majored in linguistics and took a course in phonetics. The average age of the American male group was 30 years old and that of the American female group was 25.6 years old; those of the Korean male and female groups were 32.6 and 29 years old, respectively. The Americans spoke General American. They did not have noticeably strong dialectal variation working as teaching assistants. Four of the subjects were born and spent most of their lives in Texas. The Koreans were born in Seoul and spoke Standard Seoul Korean except one male and one female from Busan. There are fewer vowels in the Busan dialect than Standard Seoul Korean. The two subjects were included in the study after personal interviews because the author evaluated them as not having any strong regional accent. They spent much time in Seoul for their undergraduate and graduate studies. The length of residence in the US for the Korean males was 44.6 months on an average while that of the Korean females was 25.6 months. The English proficiency levels of the graduate Korean students were considered very high, because they satisfied the high entrance requirements of the state university for admission as international students.

2.2 Stimuli and Data Collection

The stimuli set included *had, heed, head, who'd, hawed, hid, hod, hood, and heard.* Only those target words underlined were analyzed acoustically. The subjects produced the set in a sound-attenuated phonetics lab at the university after practicing the set twice. They were instructed to produce each word clearly, as if they would talk to a foreigner or a child without exaggerating the words. They tried again when they felt that they did not pronounce a word clearly. Also, the author monitored their production and sometimes asked them to repeat one or more words. Recording was done in the lab directly on a notebook computer (SAMSUNG SENS X10) with a headset microphone (SENNHEISER PC150) using GoldWave (v. 2.54) at a format of PCM signed 16 bit mono. Recorded data were opened on an edit window of Praat (v. 5.1.11) and the first three formant values were collected using a script after selecting the start and end points from the waveform and spectrogram. The script settings of formant measurement on the edit window were used to determine the first three formants along the six equidistant time points. However, the author sometimes changed the number of formants into higher or lower numbers if the default number 4.5 did not trace the formants correctly along the center of each dark formant band. In order to trace the acoustic trajectories along the equidistant time points, the total duration of the vowel segment of each target word was divided by 5. Then, six time points, including the vowel onset and offset time points, were designated. Because this study determined average formant values within the 25.6 ms window, the onset and offset points were assigned 22.5 ms away from the start and end points, respectively to avoid errors. Figure 1 shows two selections of the first and last time points for the acoustic measurements of the vowel of the target word, *hood*.

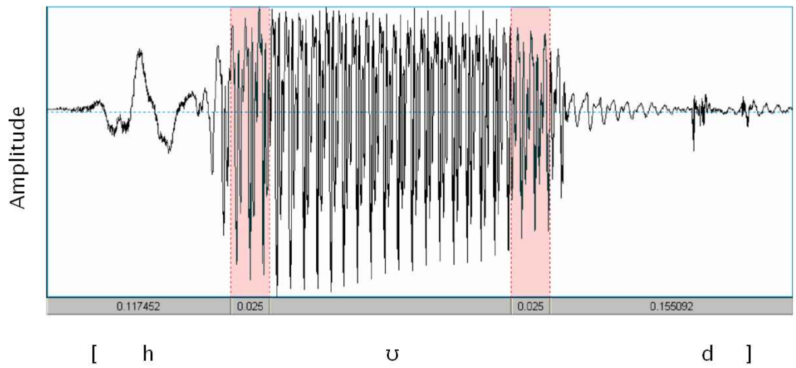

Figure 1. Two signal windows were edited and combined to show the selected measuring points at the first and sixth segments of the English word, *hood* produced by an American subject for an acoustic analysis.

The acoustic values on the six time points along the segment between the start and end points, were saved to a file. Then, all the values were verified visually using spectrogram and waveform as was done in Yang (1990, 1996). Any potential measurement errors deviating from the grand average data or adjacent data by Microsoft Excel 2007 were double checked and revised to collect the most appropriate formant data.

3. Results and Discussion

3.1 Formant Trajectories on Time Axis

In order to compare overall formant patterns of the two language groups, the average male and female formant data of the two groups were plotted on Figure 2.

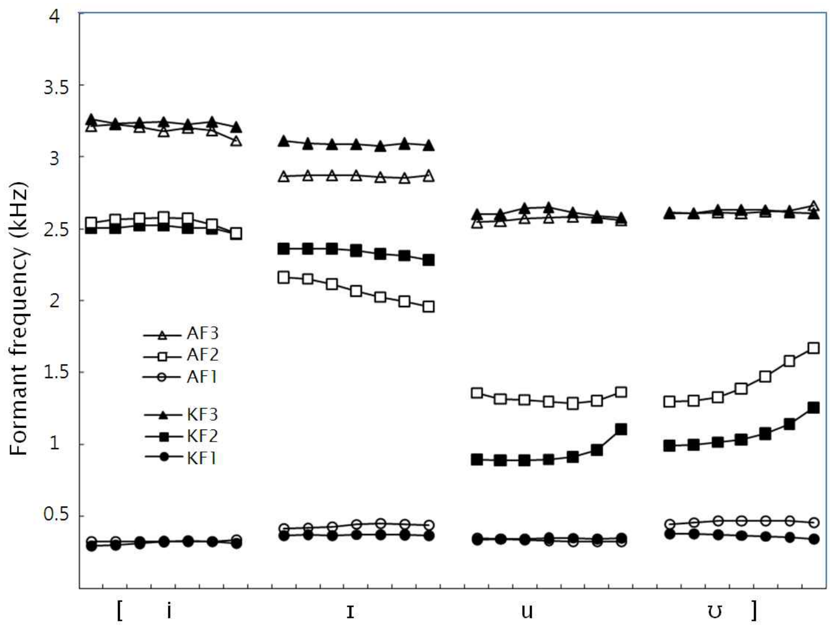

Figure 2. Formant trajectories of American and Korean groups averaged across the male and female data within each language group

There is a noticeable difference in the F2 values and the lax vowels of the two groups in the figure. The differences in F1 and F3 of the tense vowels /i, u/ seem marginal. F1 difference of the back tense and lax vowel pair /u-ʊ/ seems greater than that of the front pair /i-ɪ/. F2 difference of the back vowel pair looks much greater than that of the front pair. Generally, the American speakers differentiated the front vowel pair by the first three formants while the back vowel pair mostly by F1. Even though there is a slight upward shift of F1 for both the front and back vowel pairs, F2 difference for the Korean group does not appear to be as predominant as the American counterpart. Here it is important to mention that the frequency difference in the lower frequency region should not be underestimated because perceptual sensitivity along the audible frequency range is nonlinear. Moreover, the spectrographic display is usually made after applying the pre-emphasis procedure to boost the higher frequency range, thus, one can say that the small difference in the F1 range may be comparable to the large difference in the higher frequency range. Yang(2006) synthesized nine English monophthongs and asked twenty-seven American and Korean male and female listeners to respond to AX discrimination tasks in which the standard vowel was followed by another one with the increment or decrement of the original formant values. The acceptable ranges of the first three formant frequencies of the same vowel quality increased from the lower frequency region to the higher frequency region nonlinearly. For example, the American listeners had an average range of 102 Hz for F1 variation of the nine English vowels followed by 220 Hz for F2 variation, and 457 Hz for F3 variation.

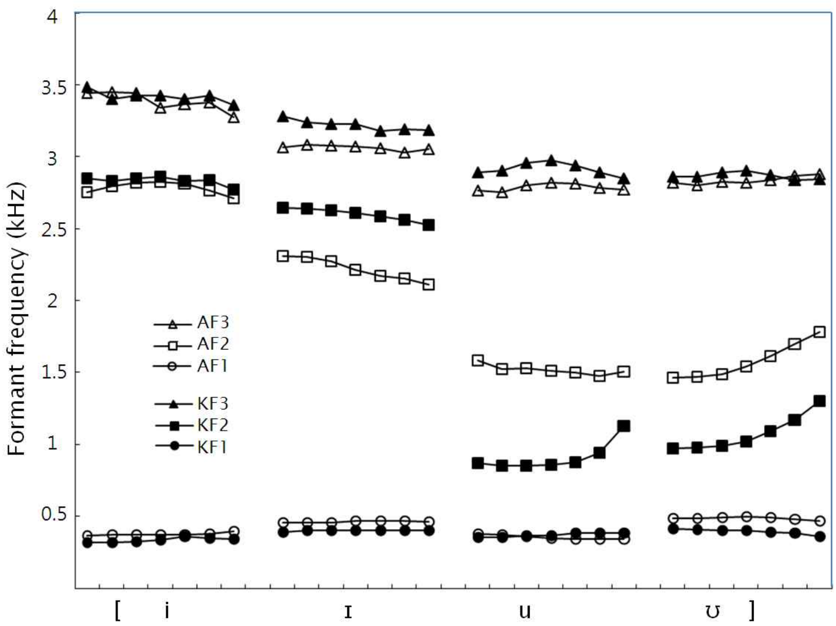

Figure 3. Formant trajectories of the vowels produced by ten American and Korean females

In order to observe any gender difference, the male and female data will be plotted and discussed separately in this paper. Figure 3 displays the formant trajectories of the American and Korean female groups.

Interestingly, one can observe that the onset point of the American female production of the back tense vowel /u/ falls exactly on that of the Korean counterpart. Apparently F1 height of the back vowel pair produced by the Korean female group appears as converging while that of the American group diverging. The trajectory of F2 of the tense back vowel by the Korean group along the six time points forms a high jump at the last two time points. It might be related to the tongue gesture toward the following consonant /d/, which will be discussed later. F3 trajectories of the front vowel /i/ and the back vowel /ʊ/ overlap.

Figure 4 illustrates the formant trajectories of the American and Korean male groups.

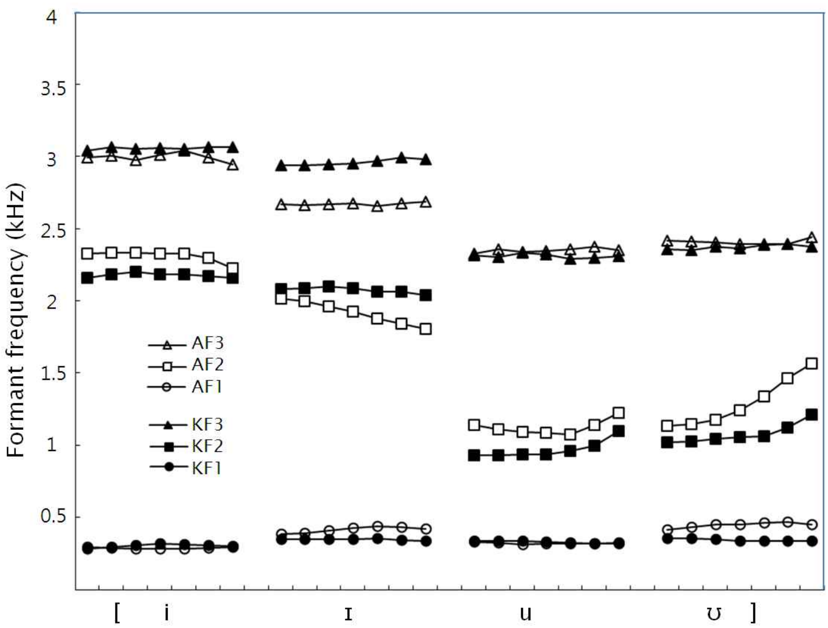

Figure 4. Formant trajectories of the vowels produced by ten American and Korean males

From the figure, one can note that many F1 values along the six time points of the two tense vowels /i, u/ of the two male groups overlap. F2 difference of the back vowel pair of the two male groups seems smaller than that of the female groups in Figure 3. Here again, F1 difference of the tense and lax vowel pairs of the American group seems to diverge from the formant trajectories of the Korean group. The

American group produced the pair quite distinctively while the Korean group did similarly. There is a vast F2 difference of the front vowel pair. F3 values of the Korean group record higher than those of the American group except the lax vowel /ɪ/, which may be attributed to different vocal tract length (Yang, 1990, 1992, 1996).

The author attempted to examine an acoustical difference between the front and back vowel pairs numerically. Figure 5 summarizes the absolute difference sum of each vowel pair produced by the four groups in light of the formant frequencies. The higher the bar, the more separately they produced the vowel pair. The grand total of the absolute difference of the first three formant frequencies between the front tense and lax vowel pair ranks from the highest, the American female group, then, the American male group, followed by the Korean female group, and lastly the Korean male group. There is a slight change in the order of the American male and female groups for the grand total of the back vowel pair. However, the American totals are almost double the Korean counterparts in both the front and back vowel pairs. Seemingly the major source of the difference comes from the second formant frequencies of the back lax vowel.

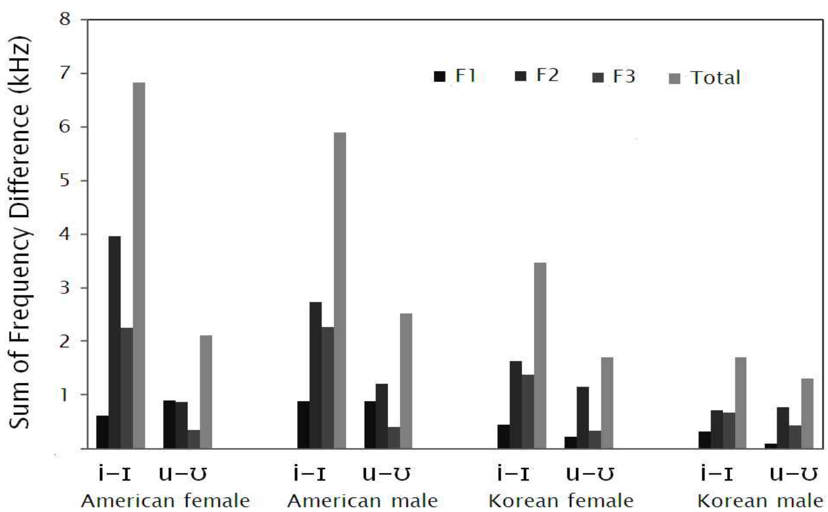

Figure 5. A bar chart of sums of the formant frequency difference (F1, F2, F3, and their total) between the tense and lax vowel pairs produced by each language and gender group

From the observation of the three figures above, one can claim that the Korean groups did not differentiate the front and back tense-lax vowel pairs as distinctly as the American groups. Also, the female speakers produced the pairs more distinctly than the male speakers. Yang (1996)'s data of the tense and lax vowel pairs seem to support the current findings. Whether those absolute differences by the Korean group may lead to higher identification of the vowels requires further research.

3.2 Formant Trajectories on Vowel Space

Generally, the articulatory positions of English vowels can be estimated from the acoustic vowel space of the F1-F2 dimension (Borden, Harris and Raphael, 2003; Ladefoged, 2001; Pickett, 1987). Vowel height is closely related to the first formant frequency while front-back dimension is somewhat related to the second formant frequency (Ladefoged, 2001). Pickett (1987, pp. 50-51) proposes F1 and F2 rules as the following: the frequency of F1 is lowered by any constriction in the front half of the oral cavity while it is raised by the constriction of the pharynx; the frequency of F2 tends to be lowered by a back tongue constriction while it is raised by a front tongue constriction. Thus, we can guess the position or movement of the speaker's jaw and tongue from the acoustic vowel space. Ladefoged (2001, p. 177) points out that F2 reflects lip rounding and suggests to use the difference between the first and the second formant to represent the degree of backness, which would partially remove some factor of lip rounding.

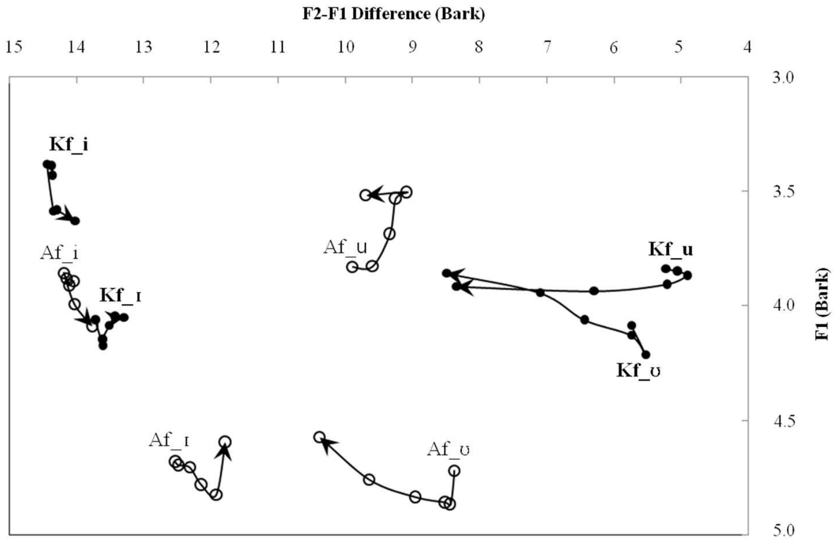

Figure 6. A formant chart showing the trajectories of the first formant (F1) on the ordinate plotted against the second to the first formant (F2-F1) difference on the abscissa for ten American (AF) and Korean female (Kf) speakers. The measurement units are in Bark. The direction of formant movement is shown by the arrows. Each dot indicates the measurement at each six time point. Vowels are shown near the starting points along with the group initials.

Figure 6 illustrates the trajectories of F1 on the ordinate plotted against F2 and F1 difference on the abscissa for ten American and Korean female speakers. In order to reduce speaker difference, the acoustic frequency scale was converted to the auditory unit, Bark, using the equation from Praat, $7*LN(x/650+SQRT(1+(x/650)^2))$. From the figure one can notice that the Korean production of the front and back vowel pairs seems not very distinct. The starting points of the front and back vowel pair produced by the Korean groups appear separated but the Euclidean distance looks shorter than those of the American groups. All the arrow tips in the figure point to a place around 12 Barks, which roughly amounts to 1800 Hz. It must be related to the locus of the consonant /d/ (Delattre, Liberman and Cooper, 1955). Previous studies (Hillenbrand, Getty, Clark and Wheeler, 1995; Peterson and Barney, 1952; Yang, 1992, 1996) employed the stimulus environment /hVd/

claiming that the coarticulation effect would be smaller than any other English consonant. However, we would expect to have more variation or distance if we ask speakers with many way-back vowels in their vowel system to produce the stimuli. Also, there must be some difference in the production of the English consonant /d/ by Americans and Koreans. The English /d/ is produced at the alveolar region, just behind the upper teeth. On the other hand, the Koreans must have produced it touching both the back of the upper teeth and the alveolar region because the Korean language does not have any adjacent consonants comparable to English inter-dental fricatives. Moreover, considering the high English proficiency of the Korean subjects, we would expect greater difference if we had included subjects with lower English proficiency as can be seen in Koo (2000) and Cho (2003). Further studies will be desirable on the subjects with various levels of English proficiency.

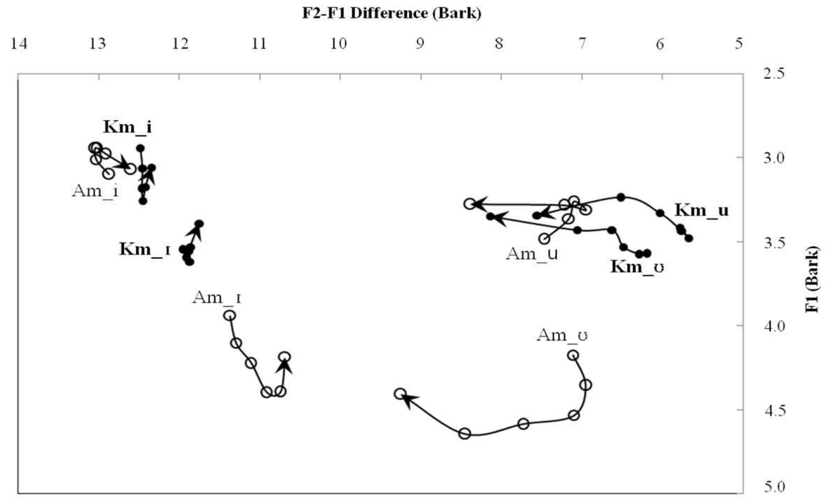

Figure 7. A formant chart showing the trajectories of the first formant (F1) on the ordinate plotted against the second to the first formant (F2-F1) difference on the abscissa for ten American (Am) and Korean male (Km) speakers. The measurement units are in Bark. The direction of formant movement is shown by the arrows. Each dot indicates the measurement at each six time point. Vowels are shown near the starting points along with the group initials.

Figure 7 shows the trajectories of the vowel points produced by the American and Korean male speakers. Here again, the American males differentiated the front and back vowel pairs more distinctly than the Korean males. The final points of the back vowel pair of the Korean male speakers almost overlap. The trajectories of the front and back vowel pairs of the American males seem to move away from each starting point then to change their direction toward the following consonant /d/ at the last two time points. A similar movement pattern can be found in the back vowel pair of the American females but the direction seems slightly different from the front vowel pair. The position of the back vowel pair by the Korean males appears to be in the more back position than that in Figure 6. The American production of the back vowel /ʊ/ is located near the high central Korean vowel /ɨ/, and perceptually the American vowel sounds like the Korean vowel. In Yang (1996)'s study the formant values of the Korean central vowel /ɨ/ almost match those of the American vowel /ʊ/. That may cause some Korean students and teachers to replace the American vowel with the Korean vowel.

The formant chart shows that F2 variations are almost similar in the front tense vowel but much different in the others. The Koreans produced the other vowels at a similar starting point but diverged to the end. Considering that F2 denotes the tongue movement, the Americans must have produced the vowels with more wide gestures than the Koreans had. As Ladefoged (2001) points out, F2 of the front lax vowels descends while that of the back lax vowels ascends. The author listened to the front lax vowel to notice that there was some diphthongal impression like the glide sequence of /ɪ/ to /ɛ /. Whether the transition is the essential quality of the front lax vowel can be judged by a perceptual experiment of synthesized vowels. The reason why F2 of the Korean back vowel pair stays further back may be related with three high vowels in Korean.

As was discussed in Yang (1996), the Koreans produced the vowel at a more back position to secure sufficient perceptual contrast with the central vowel /ɨ/. English vowel space is formed by two high vowels /i, u/, thus the vowel /u/ can be made anywhere on the high position except the frontal region for the vowel /i/. Whether we need to add more time points to the current six can be pursued to do some perceptual studies with more or fewer frequency variation points. If the dynamic variation matters, some models of second or foreign language acquisition (Best, 1995; Flege, 1995; see Nishi, Strange, Akahane-Yamada, Kubo, and Trent-Brown, 2008 for model reviews) may have to be revised to incorporate the subtle cross-linguistic difference along the spectral changes of the vowel segment. Instead of one ideal area for vowel identification, dynamic trajectories maintaining certain perceptual distance may have to be proposed. In addition, the abstract one or two static vector representation of formant values at the start and end points (Nearey and Assmann, 1986) may have to be expanded to at least three or more to represent phonetically appropriate vowel qualities.

4. Summary and Conclusion

This paper examined the acoustic features of four English vowels in /hVd/ context produced by twenty American and Korean males and females. Dynamic formant trajectories of the high front and back vowels were traced at the six time points. Results showed as the following:

First, from the comparison of formant values, the Americans distinguished the tense and lax vowels clearly while the Koreans did similarly. There was a distinct difference in the F1 and F2 values of the lax vowels along the formant trajectories.

Second, the Americans differentiated the front and back vowel pairs more distinctly than the Koreans did in the formant trajectories on the English vowel space. Some coarticulation effects from the vowels toward the consonant /d/ were observed.

From the results above, we can conclude that detailed analyses on the spectral changes may reveal more information in a cross-linguistic comparison. The results may be applicable to the teaching of English tense and lax vowels using the gestural movements on the vowel space. Further studies will be desirable to make up for the production study by a perceptual study of synthesized formant trajectories with various modifications.

References

Best, Catherine T. (1995). "A direct realist view of cross-language speech perception," in Winifred Strange, ed. *Speech perception and linguistic experience: Issues in cross-language research*, 171-204. Timonium, MD: York Press.

Borden, Gloria J., Harris, Katherine S. & Raphael, Lawrence J. (2003). *Speech science primer: Physiology, acoustics, and perception of speech*. Baltimore: Williams & Wilkins.

Cho, Hyeonkwan. (2003). "A comparative study of English vowels," *Linguistic Journal of Korea* 35, 205-230.

Delattre, Pierre, Liberman, Alvin M., and Cooper, Franklin S. (1955). "Acoustic loci and transitional cues for consonants," *Journal of the Acoustical Society of America* 27, 769-773.

Edwards, Harold T. (2003). *Applied phonetics: The sounds of American English*. New York: Delmar Learning.

Flege, James E. (1995). "Second language speech learning: Theory, findings, and problems," in Winifred Strange, ed., *Speech perception and linguistic experience: Issues in cross-language research*, 233-277. Timonium, MD: York Press.

Hillenbrand, J. M. & Nearey, T. M. (1999). "Identification of resynthesized /hVd/ utterances: Effects of formant contour", *Journal of the Acoustical Society of America* 105, 3509-3523.

Hillenbrand, James M., Getty, Laura A., Clark, Michael J., and Wheeler, Kimberlee. (1995). "Acoustic characteristics of American English vowels," *Journal of the Acoustical Society of America* 97, 3099-3111.

Jenkins, James J., Strange, Winifred, and Edman, Thomas R. (1983). "Identification of vowels in 'vowelless' syllables," *Perception Psychophysics* 34, 441-450.

Kent, Ray D., and Read, Charles. (2002). *Acoustic analysis of speech*. San Diego, CA: Singular Publishing Group.

Koo, Hee-San. (2000). "Acoustic features of English vowels produced by Koreans," *Speech Sciences* 7, 107-116.

Ladefoged, Peter (2001). *A course in phonetics*. Boston: Heinle and Heinle.

Nishi, Kanae, Strange, Winifred, Akahane-Yamada, Reiko, Kubo, Rieko, and Trent-Brown, Sonja A. (2008). "Acoustic and perceptual similarity of Japanese and American English vowels," *Journal of the Acoustical Society of America* 124, 576-588.

Nearey, Terrance M., and Assmann, Peter F. (1986). "Modeling the role of vowel inherent spectral change in vowel identification,"*Journal of the Acoustical Society of America* 80, 1297-1308.

Peterson, Gorden E., and Barney, Harold L. (1952). "Control methods used in a study of vowels," *Journal of the Acoustical Society of America* 24, 175-184.

Pickett, J. M. (1987). *The sounds of speech communication: A primer of acoustic phonetics and speech perception.* Austin, Texas: pro-ed.

Stevens, Kenneth N., and House, A. S. (1963). “Perturbation of vowel articulations by consonantal context: An acoustical study,” *Journal of Speech Hearing Research* 6, 111-128.

Watson, Catherine I., and Harrington, Jonathan. (1999). "Acoustic evidence for dynamic formant trajectories in Australian English vowels," *Journal of the Acoustical Society of America 106*, 458-468.

Yang, Byunggon. (1990). *Development of vowel normalization procedure: English and Korean.* Doctoral dissertation. The University of Texas at Austin.

Yang, Byunggon. (1992). “An acoustical study of Korean monophthongs produced by male and female speakers,” *Journal of the Acoustical Society of America* 91, 2280-2283.

Yang, Byunggon. (1996). “A comparative study of English and Korean monophthongs produced by male and female speakers,” *Journal of Phonetics* 24, 245-261.

First author
Byunggon Yang
Department of English Education
Pusan National University
30 Changjundong, Keumjunggu
Pusan, 609-735, Korea
E-mail: bgyang@pusan.ac.kr
Homepage: http://fonetiks.info/bgyang

Received: March 11, 2010
Revised: May 23, 2010
Accepted: June 16, 2010