Discrimination of Synthesized English Vowels by American and Korean Listeners

Byunggon Yang*

ABSTRACT

This study explored the discrimination of synthesized English vowel pairs by twenty-seven American and Korean, male and female listeners. The average formant values of nine monophthongs produced by ten American English male speakers were employed to synthesize the vowels. Then, subjects were instructed explicitly 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 highest and lowest formant values of the same vowel quality were collected and compared to examine patterns of vowel discrimination. Results showed that the American and Korean groups discriminated the vowel pairs almost identically and their center formant frequency values of the high and low boundary fell almost exactly on those of the standards. In addition, the acceptable range of the same vowel quality was similar among the language and gender groups. The acceptable thresholds of each vowel formed oval to maintain perceptual contrast from adjacent vowels. The results suggested that nonnative speakers with high English proficiency could match native speakers' performance in discriminating vowel pairs with a shorter inter-stimulus interval. Pedagogical implications of those findings are discussed.

Keywords: English vowels, discrimination, formant synthesis, speech perception

1. Introduction

In a daily conversation the identification of speech sounds is the first crucial stage in the listener's task to decode the incoming speech signal into a meaningful representation of the speaker's intended message (Cutler, Smits & Cooper, 2005). This stage will be more difficult when the listener comes from a different language background. Researchers on cross-linguistic perception revealed that young infants could discriminate both native and

* Professor, English Education Department, Pusan National University

non-native phonetic contrasts (Aslin et al., 1981; Lasky, Syrdal-Lasky & Klein, 1975; Trehub, 1976) whereas children and adults often have problems discriminating non-native contrasts (Goto, 1971; MacKain, Best & Strange, 1981; Trehub, 1976; Polka, 1995; Polka & Bohn, 1996; Tsukada et al., 2005). Thus, some people hypothesized that infants possess innate ability to discriminate the universal set of phonetic distinctions but this ability declines as a function of specific listening experience. Furthermore, non-native adults may improve their discrimination ability by intensive training or natural language experience by a long residence in the native country (MacKain, Best & Strange, 1981; Jamieson & Morosan, 1986; Lively et al., 1994) but even bilinguals cannot match native listeners (Mack, 1989). Mack compared the English speech perception and production of 10 fluent adult English-French bilinguals who acquired their two languages in early childhood with that of 10 adult English monolinguals. Mack found that the two groups made nearly indistinguishable discrimination of the vowel continua of /i-ɪ/. However, there was significant difference in their identification responses, specifically in the location of the group's crossover points. The bilinguals identified significantly fewer vowel stimuli as /i/ than did the monolinguals, though the groups' functions were equally monotonic. Mack interpreted the result as evidence of a restructuring of the bilinguals' perceptual system due to the phonetic properties of the French /i/. Mack concluded that the phonetic system of early adult bilinguals approximates, but does not match, that of monolinguals. Similarly, Polka (1995) examined perception of natural productions of two German vowel contrasts, /u/-/y/ and /U/-/Y/ in native English adults. She found that the English adults failed to attain native like discrimination accuracy for the lax vowel pair while all the subjects showed native-like performance for the tense vowel pair. She attributed the result to the degree of difference in category goodness of the two vowel pairs after conducting a key word identification and rating task. She suggested that linguistic experience shapes the discrimination of vowels. Tsukada et al. (2005) examined 108 Korean immigrants to America to find that the Korean children discriminated better than the adults. However, the children's performance was less accurate than the age-matched native children. Dividing and observing the Korean children into two groups of the length of residence in America, three and five years, the researchers suggested that the Korean children were in the process of learning to perceive English vowels in a native-like way. On the other hand, Polka and Bohn (1996) investigated language-specific influences on the vowel discrimination of English and German adults. They found that both native and non-native listeners discriminated English and German vowel contrasts with a high level of accuracy in the AxB task. German adults discriminated 100% correctly on the German /u/-/y/ contrast while American adults did 99.7%. For the English vowels pairs /ɛ/-/æ/ English adults scored 99.2% while German adults did 94.2%. They concluded that both native and non-native listeners demonstrated uniformly high accuracy in the vowel discrimination. The extremely high ratios of correct discrimination are understandable when one considers the two vowel pairs have a different phonetic quality, and cannot be classified in the same category. What would be the result if we ask the non-native speakers to discriminate vowel pairs of the same category? One motivation of this study is to conduct a discrimination experiment among native and non-native listeners by modifying the formant frequency values of standard vowels.

There were several attempts to explore the vowel discrimination of native listeners by varying single or more formant frequency values because vowels work as building blocks in everyday speech communication (Hawks, 1994; Kewley-Port & Watson, 1994; Kewley-Port & Zheng, 1998; Kewley-Port, 2001; Watson & Kelley, 1981) Kewley-Port and Watson (1994) obtained thresholds for formant frequency discrimination for ten synthetic English vowels. They found that those thresholds for formant frequency stayed constant at about 14 Hz in the F_1 frequency region and increase linearly in the F_2 region at a rate of 10 Hz within which the resolution for formant frequency was about 1.5%. They also observed that the increment and decrement conditions led to similar results except that there was higher subject variability for some higher frequency formants particularly, increment conditions. So they combined threshold values from the increment and decrement conditions in their final analysis. They also obtained "nearly identical" thresholds for the vowels /a, o, u/ which have converging F_1 and F_2 values, and /i/ which has closer F_2 and F_3 though Flanagan (1955) suggested that asymmetries might occur when formants fall close together. On the other hand, Hawks (1994) compared difference limens for formant patterns of synthetic vowel sounds by varying parallel or opposing simultaneous variation of F_1 and F_2 as well as single formant variation of F_1 and F_2. They obtained smaller difference limens for both single and multiple-formant changes than the previous studies. They suggested that subjects discriminated significantly better in the parallel multi-formant variation than opposing multi-formant or single-formant variation.

The present study was designed to assess how closely non-native speakers can approximate native speakers' vowel discrimination by comparing the formant frequency boundary of the same phonetic vowel quality. We adopted the AX discrimination task that may be implemented in very similar forms to both native and non-native speakers and may yield finer perceptual resolution of vowels than identification or subjective scaling procedures. We employed synthetic stimuli so that distinctive acoustic dimensions could be systematically varied. To avoid any erroneous results from non-native subjects with various English levels, this study chose a group of Korean graduate students with a higher English listening ability and experience staying in America. Research results may offer some insights into acceptable vowel pronunciation for non-native speakers.

2. Method

2.1 Stimuli

The stimuli were 376 pairs of synthesized vowels. Each pair consisted of a standard vowel followed by a modified one with a different formant value. The standard vowels were synthesized from the average formant values of the nine English vowels with more sustained patterns on the spectrogram produced by ten American males (Yang 1996, Table II) using *SenSynPPC1.0*, a Klatt88 formant synthesizer. The fourth formant was fixed at 3,700 Hz with a bandwidth of 350 Hz. The fifth formant and its bandwidth were set at 4,300 Hz and 400 Hz. Those for the sixth formant were set at 4,990 Hz and 500 Hz, respectively. The other parameters were set at default values except formant bandwidths. Bandwidths were difficult to determine from the produced vowels because of an irregular glottal source spectrum (Fujimura & Lindquivist, 1971; Klatt, 1980). The synthetic bandwidth values were generated by Fant's equations (Fant, 1972:47) in order to avoid any overloaded output from converging formants. Although bandwidth had some influence on the intensity and quality of synthesized vowel outputs, a small variation of the bandwidth within a certain frequency range did not affect perception significantly in a pilot study. The duration of each synthesized vowel was set at about 300 ms, which has been found in most vowels in clear speech (Yang, 1996). If the duration is too short, the subjects may not capture the stimulus pair phonetically. The amplitude and fundamental frequency values across time were modeled after analyzing those of nine vowels produced by two native Americans to simulate naturally produced single vowels. Its intensity and f0 ascended sharply to 79 dB and 128 Hz at 100 ms, and descended slowly to 77 dB and 117 Hz at 200 ms, respectively. Then, sets of stimuli were synthesized in which F_1 was varied by a step of 30 Hz below or above the standard formant values for F_1 while F_2 and higher formants were held constant. Also, sets of F_2 variation by a step of 40 Hz and sets of F_3 variation by a step of 50 Hz were created. The steps were adopted after several pilot studies to reduce the total duration of each session and those sets were maximally created but not running into the adjacent formants. The inter-stimulus interval (ISI) was set to 600 ms. That interval was arbitrarily adopted after reviewing the previous study (Crowder, 1982), because it was sufficient time to listen to the phonetic difference with normal memory retention from several pilot studies. If the ISI becomes longer, the memory cannot hold the preceding stimulus, and accordingly subjects' responses may result in random selection.

Originally, 1,109 files were synthesized and saved on a computer to conduct a pilot study. Two judges among the American male subjects listened to all the pairs of a synthesized standard vowel followed by another synthesized one with a modified formant value and determined the range within which they heard the same vowel quality. These pairs within the range were included in the test sets. In addition, synthesized vowels with eight to twelve steps below or above the boundary of the judges were included to capture the various ranges of the other subjects. Finally, a set of 362 synthesized vowel pairs well below or above the range of the two judges were chosen for use as perceptual stimuli. Five practice pairs were placed in the beginning of the stimuli block so that subjects could adjust to a comfortable listening level. Also, nine pairs of the same standard vowels were randomly placed in the test block to screen unreliable subjects. Only the three corner vowels /i, ɔ, u/ were included in the F_3 stimuli to reduce testing time because sets of F_3 variation showed more similar patterns of perception across a wider range than those of F_1 and F_2.

2.2 Subjects

Twenty-seven subjects participated in discrimination tests. Five additional subjects were excluded from the final data because they failed to mark the same for eight out of the nine pairs of the same vowels that were randomly inserted in the test block. All were graduate students. The average age of the native speakers was 29 years for males and 25 years for females while those of the Korean males and females were 33 and 30 years, respectively. The Korean subjects were considered well above the average in English listening ability because they scored high (mean=613, s.d.=15) on TOEFL (Test of English as a Foreign Language) and had spent around two years as graduate students in America. The length of residence in the US for the Korean males was 28 months on average (s.d.=9) while the Korean females had stayed 40 months on average (s.d.=22). Ten of the thirteen Korean speakers were born and raised in Seoul, the capital city of Korea. The others were from Pusan and Gwangju, but they did not speak with any strong regional accent. All the Korean subjects had started learning English from the age of 12 when they entered middle school. They had learned English as a foreign language accompanied by tapes of native speakers and passed the very high admission criteria of language ability provided by the University. Nine of the fourteen American speakers were born and spent most of their lives in Texas. The others came from Minnesota, New York, and North Carolina and moved around. The majority of the American subjects worked as teaching assistants and did not show any strong dialectal variation. None of the American and Korean subjects reported any history of speaking or hearing problems or experience with synthetic speech.

2.3 Procedures

The participants were individually tested for 30 minutes in the sound-attenuated Phonetics Lab at the University. Each subject listened to the stimuli binaurally through a set of headphones on a digital player at a comfortable level. The AX discrimination method was adopted so that the subjects could concentrate on each pair and judge its similarity easily. The author instructed explicitly that their task was to respond "same" or "different" to each pair of synthetic vowels in phonetic quality and not in other parameters such as pitch, loudness, or duration. Also, the subject was instructed either to circle the stimulus number if the pair sounded the same in vowel quality or to put a slash mark if it sounded different. It was expected that a blocked presentation and phonetically explicit instructions would increase the subject's sensitivity and reliance on phonetic memory in the discrimination task (Cowan & Morse, 1986).

The highest and lowest marked values in each formant of the vowel, i.e., the highest and lowest formant frequency boundary, within which each subject perceived the same phonetic vowel quality, were collected from their answers. The following interpolation procedure was applied to capture the subtle perceptual range of each subject. When a subject indicated "same", then "different", then "same" responses to three stimulus pairs in ascending or descending order, the frequency value of the middle "different" was taken for the highest or lowest boundary, respectively. However, if a subject responded "same" followed by two "different" responses (i.e., one circle after two slashes from the center) without any further "same" response, then, the highest or lowest "same" response was discarded because it deviated too far away from the center value. The center formant frequency in this study refers to the center point of the highest and lowest formant boundary of the same vowel quality. The acoustical distance between the boundary denotes the acceptable range.

3. Results

3.1 Formant boundary values of the same vowel quality

Tables I and II list the highest and lowest of the first three formant frequencies of the same vowel quality for the American and Korean listeners.

Table I. The highest and lowest value of the first three formant frequencies of the same vowel quality for the seven American male and seven American female listeners. Their standard deviations are given in parentheses. F_nf denotes the n-th formant values of females while F_nm does those of males. Fave indicates the average of the female formant values in the same column. Mave denotes the average of the male values. The units are in Hz.

Vowel	F_1mlo	F_1mhi	F_2mlo	F_2mhi	F_3mlo	F_3mhi
i	230(32)	316(0)	2186(60)	2488(183)	2862(147)	3183(132)
ɪ	362(16)	448(29)	1949(39)	2138(28)		
ɛ	495(64)	606(47)	1809(94)	1986(36)		
æ	636(88)	794(141)	1643(45)	1860(73)		
u	277(40)	380(34)	1216(83)	1587(78)	2089(53)	2482(58)
ʊ	386(35)	476(30)	1245(36)	1405(43)		
o	477(23)	567(23)	1047(40)	1270(76)		
ʌ	536(27)	643(38)	1211(65)	1420(75)		
ɔ	616(34)	706(16)	983(35)	1117(28)	2198(158)	2856(202)
Mave	446	548	1477	1697	2383	2840

Vowel	F_1flo	F_1fhi	F_2flo	F_2fhi	F_3flo	F_3fhi
i	222(21)	316(0)	2186(68)	2511(212)	2883(50)	3251(206)
ɪ	379(39)	478(38)	1903(94)	2121(92)		
ɛ	491(26)	619(43)	1791(82)	2037(100)		
æ	588(41)	788(110)	1623(65)	1863(92)		
u	282(33)	376(45)	1233(131)	1582(92)	2061(144)	2496(114)
ʊ	403(16)	497(33)	1234(45)	1428(73)		
o	455(16)	562(32)	1036(20)	1264(73)		
ʌ	541(29)	648(27)	1200(55)	1437(79)		
ɔ	603(30)	717(34)	946(45)	1103(35)	2091(141)	2984(237)
Fave	440	556	1461	1705	2345	2910

Table II. The highest and lowest of the first three formant frequencies of the same vowel quality for the six Korean male and seven Korean female listeners. Their standard deviations are given in parentheses. Fave indicates the average of the female formant values in the same column. Mave denotes the average of the male values. The units are in Hz.

Vowel	F_1mlo	F_1mhi	F_2mlo	F_2mhi	F_3mlo	F_3mhi
i	221(23)	321(12)	2157(91)	2544(194)	2733(122)	3258(88)
ɪ	377(33)	452(31)	1892(36)	2192(22)		
ɛ	546(44)	606(13)	1787(30)	2027(39)		
æ	620(67)	770(86)	1656(47)	1863(36)		
u	288(45)	393(42)	1200(111)	1573(70)	2074(49)	2540(92)
ʊ	386(33)	516(36)	1211(67)	1451(57)		
o	463(12)	553(23)	1047(0)	1240(39)		
ʌ	537(44)	632(31)	1191(42)	1488(23)		
ɔ	608(29)	713(24)	973(64)	1096(45)	2202(227)	2902(227)
Mave	449	551	1457	1719	2336	2900

Vowel	F_1flo	F_1fhi	F_2flo	F_2fhi	F_3flo	F_3fhi
i	205(15)	320(11)	2157(118)	2551(78)	2897(48)	3319(175)
ɪ	383(27)	473(21)	1886(49)	2138(36)		
ɛ	505(59)	617(47)	1769(68)	2009(55)		
æ	576(33)	726(86)	1634(30)	1846(60)		
u	260(42)	389(32)	1233(113)	1513(57)	2038(96)	2561(111)
ʊ	395(23)	523(29)	1234(31)	1451(57)		
o	472(40)	575(34)	1041(15)	1253(43)		
ʌ	532(35)	639(16)	1217(54)	1460(25)		
ɔ	603(57)	723(35)	957(36)	1103(45)	2120(130)	2956(229)
Fave	437	554	1459	1703	2352	2945

One can note that the grand average values of the four groups were almost identical. General trends indicate that the higher the formant number, the bigger the standard deviations. No subjects in each language group responded the highest or lowest boundary of the randomized test stimuli set as the final perceptual boundary. Some boundary values exactly matched within and across the language or gender groups. Within the language groups, the highest F_1 and lowest F_2 boundary values of the vowel /i/ perceived by the American male and female groups were the same. Also, those of the Korean males and females came out almost the same. Between the language groups, the lowest F_1 boundary values of the vowel /ɔ/ between the American and Korean females were the same as well as that of the vowel /ʊ/ between the American and Korean males. The highest F_2 boundary of the American and Korean male groups came out the same. In addition, the lowest F_2 boundary across American and Korean female groups was exactly the same. The formant frequency difference between the two language groups ranged from 0 to 129 Hz while that between the two gender groups ranged from 0 to 164 Hz. The maximum difference in F_1 and F_2 between the language groups of the same gender was 69 Hz in the highest F_2 boundary of the female vowel /u/. That of F_3 was 68 Hz in the highest boundary of the female vowel /i/, and 129 Hz in the lowest boundary of the same male vowel. Within the same language group, the maximum difference of 128 Hz occurred in the highest boundary of the vowel /ɔ/ between the American males and females. The difference between the Korean male and female groups in the lowest F_3 boundary of the vowel /i/ was 164 Hz. On average, the differences of the lowest boundary in F_1 between the American males and females were 8 Hz and 6 Hz, respectively. Those for the Korean male and female groups were 3 Hz and 12 Hz. Also, the cross-linguistic difference between the American and Korean males or between the American and Korean females was 3 Hz. The difference of the lowest boundary in F_2 between the American male and female groups was 16 Hz while that of the Korean groups was 2 Hz. Again, the cross-linguistic difference of the lowest boundary was less than 20 Hz. In F_3, the difference within each group was 38 Hz for the Americans and 16 Hz for the Koreans. The difference across the two languages was 7 Hz between the male groups and 47 Hz for the female groups. All of the differences above seem to be marginal when one considers the wide acceptable ranges between the highest and lowest thresholds of the same vowel quality.

We determined the center formant frequency values to conduct one-way ANOVAs between the two language groups using *SPSS 10.0 for Windows*. The results came out as not significant: F (1, 241)=0.000, p=0.994 for F_1c; F (1, 241)=0.000, p=0.988 for F_2c; F (1, 79)=0.040, p=0.841 for F_3c. Also, results from the one-way ANOVAs between the two gender groups were obtained as follows: F (1, 241)=0.011, p=0.918 for F_1c; F (1, 241)=0.10, p=0.920 for F_2c; F (1, 79)=0.102, p=0.750 for F_3c.

Overall, these analyses indicate that there exist no significant differences across the language and gender groups in the discrimination of the synthesized vowel pairs. The present findings may relate to the use of the same stimuli but it is quite interesting to observe converging results even though the presentation of the stimuli was done randomly.

3.2 Acceptable ranges of the nine English vowels

Tables III and IV denote the acceptable ranges of the first three formant frequencies of the same vowel quality on the AX discrimination test by the American and Korean male and female listeners.

Table III. The acceptable ranges of F_1-F_3 of the nine English vowels discriminated by the American listeners.

Vowel	F_1mr	F_2mr	F_3mr
i	86(32)	303(197)	321(185)
ɪ	86(36)	189(30)	
ɛ	111(51)	177(124)	
æ	159(69)	217(105)	
u	103(49)	371(124)	393(93)
ʊ	90(55)	160(73)	
o	90(35)	223(100)	
ʌ	107(49)	209(69)	
ɔ	90(35)	134(51)	657(348)
Average	102	220	457

Vowel	F_1fr	F_2fr	F_3fr
i	94(21)	326(269)	368(233)
ɪ	99(69)	217(142)	
ɛ	128(61)	246(162)	
æ	199(130)	240(142)	
u	94(70)	349(184)	436(246)
ʊ	94(47)	194(96)	
o	107(42)	229(89)	
ʌ	107(52)	237(120)	
ɔ	114(57)	157(73)	893(358)
Average	115	244	565

Table IV. The acceptable ranges of F_1-F_3 of the nine English vowels discriminated by the Korean listeners.

Vowel	F_1mr	F_2mr	F_3mr
i	100(31)	387(249)	525(186)
ɪ	75(37)	300(49)	
ɛ	60(46)	240(51)	
æ	150(114)	207(69)	
u	105(73)	373(155)	467(133)
ʊ	130(56)	240(110)	
o	90(19)	193(39)	
ʌ	95(44)	297(56)	
ɔ	105(41)	123(37)	700(315)
Average	101	262	564

Vowel	F_1fr	F_2fr	F_3fr
i	116(21)	394(112)	421(193)
ɪ	90(24)	251(72)	
ɛ	111(41)	240(101)	
æ	150(114)	211(82)	
u	129(29)	280(95)	523(163)
ʊ	129(41)	217(56)	
o	103(49)	211(38)	
ʌ	107(24)	243(48)	
ɔ	120(46)	146(46)	836(356)
Average	117	244	593

The acceptable ranges were quite comparable between each language and gender group as shown in their average values. Within the same language groups, the difference between American male and female groups was within the range of 0 to 236 Hz while that of the Korean groups amounted to 0 to 136 Hz. Across the language groups, the difference ranged from 0 to 204 Hz. The average standard deviation of F_1 for the four groups was 51 Hz followed by 104 Hz for F_2 and 189 Hz for F_3. In F_1 the vowel [æ] showed the widest acceptable range across the four different groups. Also its standard deviation was greater. On the other hand, the standard deviations of the high front vowel [i] were relatively smaller. In F_2 the high vowels [ɪ, u] showed wider acceptable ranges in the four groups whereas the open vowel [ɔ] had the narrowest range across the four groups. The standard deviations of the high vowels of F_2 were relatively greater than those of the others. In F_3, the high vowels [ɪ, u] had narrower acceptable ranges than the open vowel [ɔ]. From Tables III and IV we found that the female groups had slightly wider ranges than the male groups except in the F_2 range of the Korean group. This might be related to their higher formant values arising from a shorter vocal tract length (Fant, 1972). Diehl et al. (1996) pointed out that the male/female difference exists in the females' sparse resolution in the higher frequency range because of their high fundamental frequency. The acceptable ranges were much wider than the vowel formant thresholds suggested by Kewley-Port and Watson (1994) because we asked the subjects to judge the phonetic vowel quality in the synthesized vowel pairs.

Results on the analysis of variance between the two gender groups also revealed non-significant main effects for the first three formant ranges: F (1, 241)=3.784, p=0.053 for F_1r, F (1, 241)=0.064, p=0.800 for F_2r and F (1, 79)=1.238, p=0.269 for F_3r. Between the languages we obtained the following results: F (1, 241)=0.017, p=0.896 for F_1r; F (1, 241)=1.532, p=0.217 for F_2r and F (1, 79)=1.088, p=0.300 for F_3r. Here again, we found that there was no significant difference in the discrimination ranges for either language or gender.

4. Discussion

In the previous section we observed that there were no significant differences in the center formant frequency and threshold ranges of both native and non-native listeners. What does the vowel space look like if we plot the F_1 center formant frequency values of the highest and lowest formant frequency values for the nine vowels and the synthesis standard vowels against those of F_2? Figure 1 shows the nine vowel points of the American and Korean groups. The synthesis standards are inscribed on the vowel space near its center marked "+".

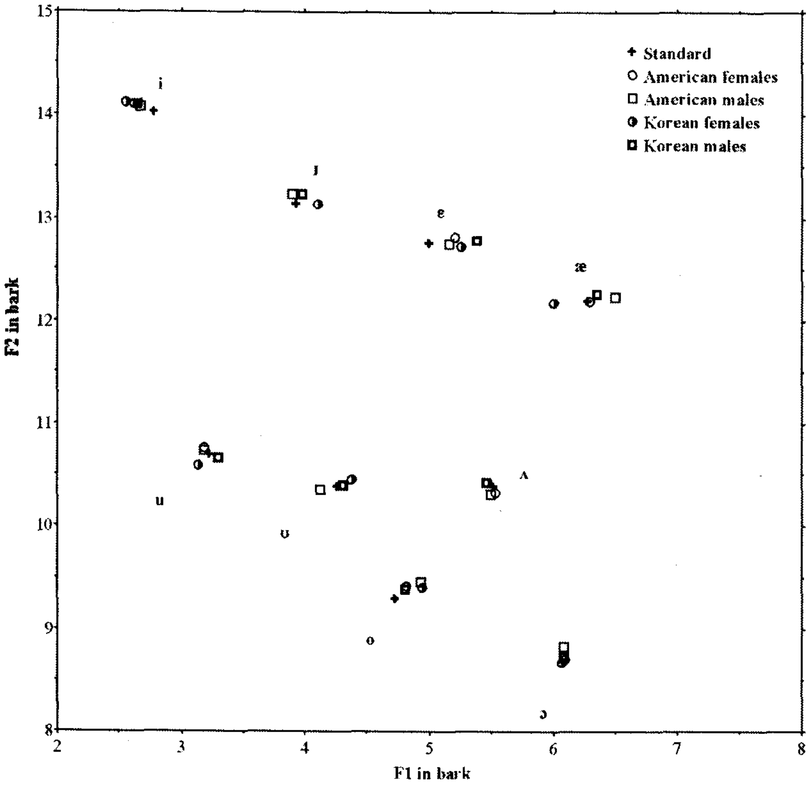

Figure 1. Vowel chart of the first formant frequency (F_1) against the second formant frequency (F_2) in bark. · The center formant frequency values were determined from the highest and lowest thresholds of the same vowel quality.

As was observed in the statistical comparison of the previous section, there were some marginal shifts from the standard vowel points. All the four groups clustered around the standards. One can note that the perceptual distance seems to be maintained in the perceptual vowel space. The distance among adjacent front or back vowels is quite similar. The reason can be attributed to the original stimuli with evenly-spaced vowel points which were obtained from the average formant values of the American male speakers. Interestingly, the listeners had captured the target vowel points almost exactly at the standards. For the vowel /æ/ the Korean females show smaller distance. It appears that the four groups perceived the synthesized vowel pairs almost the same even though their formant values of production might be considered quite different (Yang, 1996). Figure 2 shows the bark difference between the formant values of the standard vowels and those of the four groups.

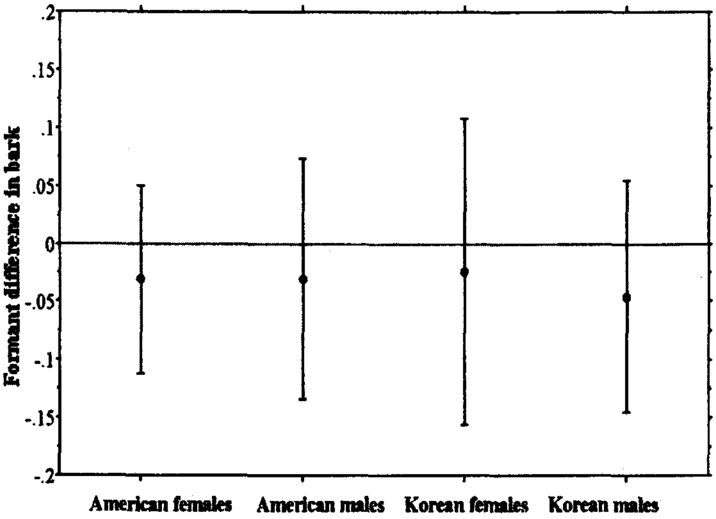

Figure 2. Formant difference in bark between the standard and the four groups

The maximum and minimum difference range in bark was 0.372 bark between the standard and the American females, 0.356 bark between the standard and the American males, 0.54 bark between the standard and the Korean females and 0.53 bark between the standard and the Korean males, one-half a critical band which Kewley-Port (2001) indicated as a typical threshold for formant discrimination by listeners with little training. One may note that the Korean groups discriminated the synthesized vowel pairs slightly more broadly away from the targets than the native speakers.

How can we apply these findings to English vowel education? Specifically, will American males and females judge English vowels produced by Korean males and females the same as theirs in phonetic quality? If we can define a certain area of the same vowel quality by native speakers in the vowel space, it can be used to judge non-native speakers' vowel production level. In other words, American listeners would accept those vowels within the range as normal English ones. Then, non-native English teachers may offer immediate feedback to non-native learners by asking them to open the jaw further or to move the tongue forward or backward. Some phonetic trainers tend to demand students to exactly imitate what the native speakers produce despite their vocal tract size differences. Figure 3 illustrates the highest and lowest vowel points of the 18 American male and female speakers. The formant frequency is transformed into bark.

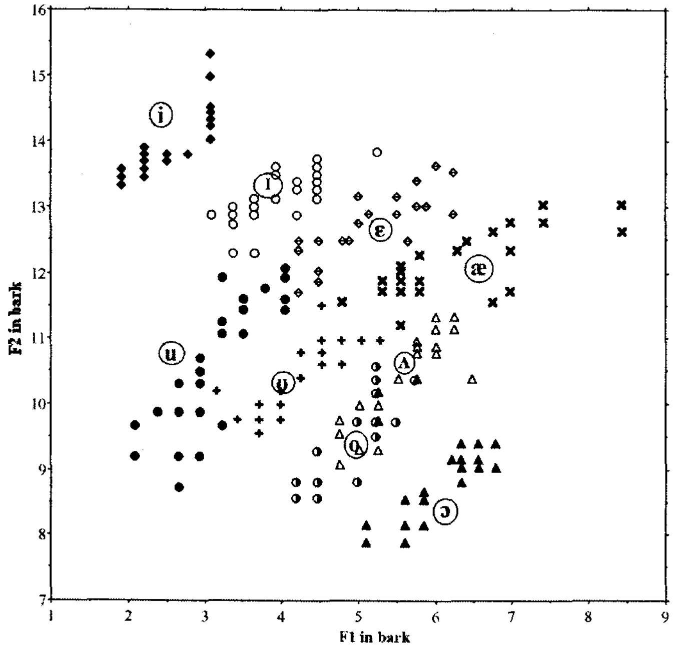

Figure 3. Vowel chart of the first formant frequency (F_1) against the second formant frequency (F_2) in bark. Each point denotes the highest or lowest thresholds of the same vowel quality.

Interestingly, the boundary points form a certain oval region for each vowel. The region appears to be squashed by the adjacent vowel points so that they spread diagonally. Also, the boundary points of the same vowel form a certain oval area not overlapping adjacent vowel regions except the vowel /o/ and /æ/. Those two vowels can be clearly distinguished from each other if we include F_3, which generally represents the lip rounding feature. A wider spread can be observed in the vowels /u/ and /æ/. This shape might be related to the maintenance of perceptual contrast in the neighboring vowels. Also, these asymmetrical shapes conform very closely with the oval areas of the same American English vowels (Peterson & Barney, 1952: Fig. 8). Cowan and Morse (1986: Fig. 4) proposed a model of decay in the memory representation of a vowel. They observed that the boundary of the confidence region for representation was narrowly defined around the perceived vowel quality within the vowel space in an ISI of 500 ms after the vowel. However, as the ISI became longer, the confidence region expanded to the direction of the schwa. They speculated that the movement might occur because there was little room for expansion toward the edges of the vowel direction and predicted that the perceived quality presumably shifted so as to remain centered within the expansion region. If they examined the confidence regions with more vowels, they might illustrate the directions of each vowel perception like the current study.

Before we continue our discussion on the pedagogical implications, we may have to consider some additional points. This study employed the average formant values of ten American males to get the current acceptable ranges and the ranges of the four different groups. We would expect to obtain similar ranges but different center formant frequency values if we employed a different synthesis standard. In other words, we may still need to find different ranges of various synthesis standards to evaluate whether those non-native productions fall well above or below the threshold ranges set by native listeners. This may be related to the normalization in which the acoustical formant values of males and females are significantly different but average people can easily identify the given vowel considering the range of each speaker's variation. Further experiments with a stimulus set of different formant models may provide criteria for that purpose.

In the previous result section, we noted that the acceptable ranges were quite similar among the speakers and that the higher formants had the wider ranges. We will explore the relationship between the acceptable ranges and the formant frequency values. First, the high and low frequency endpoints were converted to bark. Then the center formant frequency values were collected to compare the production data transformed into bark units as shown in Figure 4.

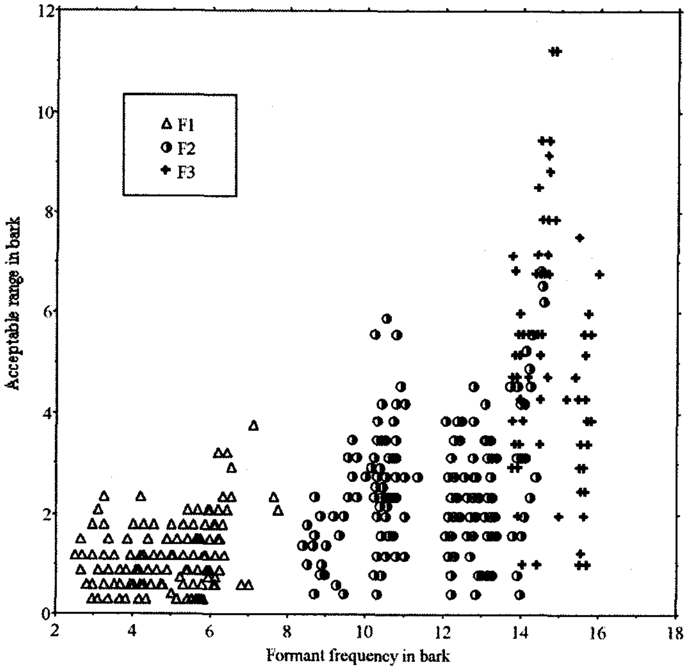

Figure 4. The formant frequency against the acceptable range in bark.

From Figure 4 we observe that the acceptable ranges become wider as the formant frequency increases. Generally, the lower formant values have narrower acceptable ranges and the higher formants have wider acceptable ranges. These trends may be related to the distance between adjacent formants. Correlation coefficients were obtained between the center formant frequency values and the acceptable ranges as well as the distance among the three formant values. There was a moderate negative correlation ($r=-0.537$, $p<0.05$) between the acceptable range of F_2 and F_3c-F_2c, while a moderate positive correlation ($r=0.579$, $p<0.05$) existed between the acceptable range of F_3 and F_3c-F_2c. However, the coefficients between the acceptable range of F_1 and F_2c-F_1c ($r=-0.068$), F_3c-F_1c ($r=-0.062$) or F_3c-F_2c ($r=-0.086$) were almost negligible as well as non-significant ($p<0.05$). The other comparisons were significant but with low coefficients: F_2r and F_2c-F_1c ($r=0.347$), or F_3c-F_1c ($r=0.359$); F_3r and F_2c-F_1c ($r=-0.475$), or F_3c-F_1c ($r=-0.314$). These results suggest that the acceptable ranges do not strongly depend on the distance between adjacent formant values. In addition, the figure reflects a high resolution in the lower frequency range while there is a coarse resolution in the higher frequency range as shown in the human auditory scale, bark (Scharf, 1970; Zwicker & Terhardt, 1980). The critical bandwidth is defined as “that bandwidth at which subjective responses rather abruptly change” (Sharf, 1970:159). It reflects the bandwidth of the filters in the human auditory system but the scale may not be comparable to the current perceptual range because we could not derive a reliable regression function from our data.

In the previous study by Yang (1996), the non-native Korean speakers could not distinguish between the tense and lax vowel pairs including the [e, æ] pair in their production, but in this study the Korean subjects discriminated those vowels very accurately, almost comparable to the native speakers. This may be related to the fact that the Korean subjects had spent around two years in the United States and possessed a higher English listening ability. As Werker (1994) suggested, the Korean subjects might have reorganized the innate discriminative ability by being frequently exposed to an English environment (Werker & Tees, 1984a, 1984b). Kewley-Port (2001) also reported that some subjects achieved discrimination performance similar to that of the trained listeners in just an hour. From a pilot study comparing Korean production of English vowels with that of Americans, we found that the Koreans could not match the native speakers' target formant frequency.

5. Summary and Conclusion

The present study was designed to compare the discrimination of synthesized English vowels. In this study, 27 American and Korean listeners participated in the discrimination experiment of the synthesized English vowels. The average formant values of the nine vowels produced by ten American English speakers were employed to synthesize the vowels. The center formant values were determined from the highest and lowest formant thresholds of the same vowel quality. Then, those values of the four groups were compared to examine patterns of vowel discrimination. Results revealed that the non-native subjects had perceived the synthesis standard almost exactly as the native speakers. Statistical comparison indicated that the American and Korean groups discriminated the vowel pairs almost identically. Also, their center formant frequency values fell almost exactly on the standards. Furthermore, the acceptable range of the same vowel quality turned out to be almost the same across the language and gender groups. The central formant frequency points of the American and Korean groups matched the standard vowel points. The highest and lowest boundary points appeared oval not interfering with formant values of adjacent vowels. Finally, the higher formant resulted in the wider acceptable ranges. Taken together, the results suggested that nonnative speakers with high English ability could match native speakers' performance in discriminating vowel pairs with a shorter inter-stimulus interval.

However, it should be noted that we have looked at the AX discrimination of synthesized English vowels by native and non-native speakers after varying one of the first three formant values while controlling the other two formants. Also, there was not much difference in the vowel discrimination with a shorter inter-stimulus interval. This might not occur in actual speech production. The three formant values usually interact with each other because the volume of the human tongue is relatively constant and the oral cavity will vary according to its movement. Consequently, it may be desirable to investigate some possible combinations of formant modification in future studies.

References

Aslin, R. N., Pisoni, D. B., Hennessy, B. L. & Perey, A. J. 1981. Discrimination of voice onset time by human infants: New findings and implications for the effect of early experience. *Child Development*, 52, 1135-1145.

Crowder, R. G. 1982. Decay of auditory memory in vowel discrimination. *Journal of Experimental Psychology: Human Learning and Memory* 8, 153-162.

Cowan, N. & Morse, P. A. 1986. The use of auditory and phonetic memory in vowel discrimination. *Journal of the Acoustical Society of America* 79, 500-507.

Cutler, A., Smits, R. & Cooper, N. 2005. Vowel perception: Effects of non-native language vs. non-native dialect. *Speech Communication*, 47, 32-42.

Diehl, R. L., Lindblom, B., Hoemeke, K. A. & Fahey, R. P. 1996. On explaining certain male-female differences in the phonetic realization of vowel categories. *Journal of Phonetics*, 24, 187-208.

Fant, G. 1972. Vocal tract wall effects, losses, and resonance bandwidths. *STL-QPSR*, 2-3, 28-52.

Flanagan, J. 1955. A difference limen for vowel formant frequency, *Journal of the Acoustical Society of America*, 27, 288-291.

Fujimura, O. & Lindquivist, J. 1971. Sweep-tone measurements of vocal-tract characteristics. *Journal of the Acoustical Society of America*, 49, 541-558.

Goto, H. 1971. Auditory perception by normal Japanese adults of the sounds "L" and "R". *Neuropsychologia*, 9, 317-323.

Hawks, J. W. 1994. Difference limens for formant patterns of vowel sounds. *Journal of the Acoustical Society of America*, 95(2), 1074-1084.

Jamieson, D. G. & Morosan, D. E. 1986. Training non-native speech contrasts in adults: Acquisition of English /θ/-/ð/ contrast by franco-phones, *Perception & Psychophysics* 40, 205-215.

Kewley-Port, D. 2001. Vowel formant discrimination II: Effects of stimulus uncertainty, consonantal context, and training. *Journal of the Acoustical Society of America*, 110(4), 2141-2155.

Kewley-Port, D. & Watson, C. S. 1994. Formant-frequency discrimination for isolated English vowels. *Journal of the Acoustical Society of America*, 95(1), 485-496.

Kewley-Port, D. & Zheng, Y. 1998. Auditory models of formant frequency discrimination for isolated vowels. *Journal of the Acoustical Society of America*, 103(3), 1654-1666.

Klatt, D. 1980. Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, 67, 971-995.

Lasky, R. E., Syrdal-Lasky, A. & Klein, R. E. 1975. VOT discrimination by four to six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology*, 20, 215-225.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y. & Yamada, T. 1994. Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories, *Journal of the Acoustical Society of America*, 96, 2076-2087.

Mack, M. 1989. Consonant and vowel perception and production: Early English-French bilinguals and English monolinguals, *Perception & Psychophysics*, 46, 187-200.

MacKain, K. S., Best, C. T. & Strange, W. 1981. Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*, 2, 368-390.

Peterson, G.E. & Barney, H. L. 1952. Control methods used in the study of the vowels. *Journal of the Acoustical Society of America*, 24(2), 175-184.

Pisoni, D. B. 1973. "Auditory and phonetic memory codes in the discrimination of consonants and vowels, *Perception & Psychophysics*, 13, 253-260.

Polka, L. 1995. Linguistic influences in adult perception of non-native vowel contrasts. *Journal of the Acoustical Society of America*, 97, 1286-1296.

Polka, L & Bohn, O. -S. 1996. A cross-language comparison of vowel perception in English-Learning and German-learning infants. *Journal of the Acoustical Society of America*, 100(1), 577-592.

Sharf, B. 1970. Critical bands. In J. V. Tobias (Ed.) *Foundations of Modern Auditory Theory Vol. I.* New York: Academic Press.

Stevens, K. 1998. *Acoustic Phonetics.* Cambridge, MA: The MIT Press.

Trehub, S. E. 1976. The discrimination of foreign speech contrasts by infants and adults. *Child Development*, 47, 466-472.

Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H. & Flege, J. 2005. A developmental study of English vowel production and perception by native Korean adults and children. *Journal of Phonetics*, 33, 263-290.

Watson, C. S. & Kelly, W. J. 1981. The role of stimulus uncertainty in the discrimination of auditory patterns. In D. J. Getty & J. N. Howard (Eds.) *Auditory and Visual Pattern Recognition*, (pp. 37-59). Hillsdale, NJ: Erlbaum.

Werker, J. F. 1994. Cross-language speech perception: Development change does not involve loss. In J. C. Goodman & H. C. Nusbaum (Eds.) *The development of speech perception: The transition form speech sound to spoken words* (pp. 93-120). Cambridge, MA: The MIT Press.

Werker, J. F. & Tees, R. C. 1984a. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. Infant Behavior and Development, 7, 49-63.

Werker, J. F. & Tees, R. C. 1984b. Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America*, 75(6), 1866-1878.

Yang, B. 1996. A comparative study of American English and Korean vowels produced by male and female speakers. *Journal of Phonetics*, 24, 245-261.

Zwicker, E. & Terhardt, E. 1980. Analytical expressions for critical-band rate and critical bandwidth as a function of frequency. *Journal of the Acoustical Society of America*, 68, 1523-1525.

received: January 30, 2006
accepted: March 13, 2006

▲ Byunggon Yang
English Education Department, Pusan National University
30 Changjundong, Keumjunggu, Pusan, 609-735, Korea
Homepage://fonetiks.info/bgyang
Tel: 010-9618-7636
E-mail: bgyang@pusan.ac.kr