Journal of Phonetics (1996) **24,** 245–261

A comparative study of American English and Korean vowels produced by male and female speakers

Byunggon Yang

Department of English, College of Humanities, Dongeui University, 24 Kaya-dong, Pusanjin-gu, Pusan 614-714, Korea

Received 16th February 1993, and in revised form 26th October 1995

F_1–F_3 and f_0 of 10 Korean vowels and 13 American English vowels produced by 10 male and 10 female speakers of each language group were studied while holding dialectal factors as homogeneous as possible in each group. Within- and across-language comparisons of the collected data revealed considerable variation in vocal tract length between male and female speakers and between Korean and American English speakers. For a more precise comparison, the variation was drastically reduced by applying uniform scaling within and across the two languages. In the cross-language comparison of the normalized data, which were converted to a perceptual dimension, it is argued that adaptive dispersion is operating within a language's system of contrasts to fulfill a condition of sufficient contrast. *t*-tests were conducted on those vowels transcribed using the same or similar IPA symbols in the two languages to assess the statistical significance of those comparisons.

1. Introduction

The main acoustic correlates of vowel quality are formant frequencies. However, vowels spoken by different speakers show great variation in formant values. If one wanted to compare the vowel qualities produced by native speakers of different languages, such as English and Korean, by measuring vowel formant frequencies, the following questions would arise:

1. What aspects of the acoustic measurements reflect genuine phonetic quality differences between the two languages?

2. What aspects reflect linguistically unimportant idiosyncrasies? How and to what extent can these linguistically irrelevant differences be reduced?

3. Once these differences are reduced, what are the theoretical implications of any cross-linguistic differences that remain?

Failure to address the first two questions will make it impossible to compare acoustic measurements of vowels from any two speakers or language populations. Resolving those questions requires diagnosis of sources of speaker variation. Speaker variation has been attributed to (1) linguistic factors such as dialectal and sociolectal differences and (2) non-linguistic factors such as physical anatomy, age,

0095-4470/96/020245 + 17 $18.00/0

gender, and emotional state of the speaker (Ladefoged & Broadbent, 1957; Traunmüller, 1988). Some of the non-linguistic factors are systematic and their effects may be theoretically separable from linguistically relevant properties of speech by systematic transformations; other factors may be minimized by methods of statistical inference (Fujisaki, 1972). The goal of factoring out these nonlinguistic factors is to allow a linguistically relevant acoustic specification of the vowel qualities of any given language. This procedure has been called ''normalization'' (Fant, 1968). Efforts to normalize the vowel qualities of different speakers can generally be divided into auditorily-based (Syrdal & Gopal, 1986; Miller, 1989) and articulatorily-based (Nordström & Lindblom, 1975; Fant, 1975) proposals for speaker normalization. The former methods focused on transforming the raw data by using logarithmic scaling, the Bark (critical-band–rate scaling), the mel, or the König scales. The latter ones attempted to remove anatomical differences in vocal tract length or in the ratio of pharynx to mouth cavity. A detailed discussion of their merits and demerits is found in Yang (1990).

An appropriate normalization procedure can be a ''valuable tool for the classification of vowels'' (Disner, 1980: 253). If the confounding elements introduced by heterogeneity among speakers are effectively removed by normalization, the resulting acoustic vowel chart becomes an accurate representation of the linguistic aspects of the vowels, facilitating both across-speaker and across-language comparison. For foreign language teachers, a comparative study of vowel qualities in the foreign and the native language may prove valuable in teaching the language and correcting student's errors. Also, quantitative criteria, handled properly, could be used to evaluate student's achievement of correct pronunciation skills in the foreign language.

Because the ultimate goal of this study is to achieve a cross-linguistic comparison of normalized vowels, this study will review three non-linguistic differences between male and female speakers which potentially interfere with this comparison. These are fundamental frequency, vocal tract length, and the ratio of front cavity to back cavity.

First, the rate of vocal fold vibration (acoustically, f_0 range) is inversely proportional to the mass and length of the vocal folds and proportional to the tension of the folds. Negus (1949) reported that vocal cord length averaged 12.5 to 17 mm in adult females and 17 to 23 mm in adult males; the average rate of vibration of males' folds is about 125 Hz compared to about 200 Hz for females. f_0 also varies as the speaker changes the tension of laryngeal muscles and to some extent the subglottal pressure (Lieberman, 1967). Boothroyd (1986) observed that f_0 varied between 70–200 Hz in men, 140–400 Hz in women, and 180–500 Hz in children.

Second, formant frequencies are inversely related to the overall length of the speaker's vocal tract, which varies according to age and gender (with females usually having shorter vocal tracts than males). Overall vocal tract length can be estimated directly from formant frequency measurements. Assuming the cross-sectional vocal tract area to be almost uniform for the vowel [ʌ] (as in English *Hudd*), one can obtain the length of the speaker's vocal tract (L) by specifying F_3 of [ʌ] in the well-known formula (Fant, 1970: 292):

$$L = 5C/4F_3 \quad (1)$$

(C = 34 000 cm/sec = speed of sound in air)

Using Equation (1), the vocal tract ratios of female to male in three European languages were found to be 0.89 for Swedish (Fant, 1975), 0.89 for Dutch (van Nierop, Pols & Plomp, 1973; Pols, Tromp & Plomp, 1973), and 0.86 for English (Peterson & Barney, 1952). These ratios corroborate Chiba & Kajiyama's (1941) estimates of a female/male ratio of 0.87 and indicate that, on average, female vocal tracts are 11–14% shorter than those of males.

Based on overall vocal tract difference, Nordström & Lindblom (1975) proposed a uniform scaling method for gender normalization. Their method involved estimating the total length of a subject's vocal tract from an average of F_3 in vowels with F_1 greater than 600 Hz. Then, the ratio k of the length of the average male vocal tract (L_m) to the average female vocal tract length (L_f) is determined by

$$k = F_3m_{av}/F_3f_{av} \tag{2}$$

in which F_3m_{av} and F_3f_{av} indicate an average of the third male and female formant values. Equation (2) is used for the uniform scale factor k in this paper. In Nordström & Lindblom's procedure, the female formant frequencies are adjusted along a trajectory to a position within the male reference system by multiplying by the scale factor k. Thus, each scaled *n*th female formant frequency is denoted as F_nf_{sc} and can be determined according to

$$F_nf_{sc} = k \times F_nf \tag{3}$$

This study adopts the uniform scaling method proposed by Nordström & Lindblom (1975) for a cross-linguistic study because it is one of the simplest procedures with a good normalization result.

A third non-linguistic factor which introduces variation between speakers is the ratio of pharynx to mouth cavity lengths. Chiba & Kajiyama (1941: 188–193) stated that the mouth cavity length of an eight-year old girl was 30% shorter than that of an adult male, while the length of the girl's pharynx was 56% shorter than that of the male. The length of a pharynx and mouth cavity can be estimated from the formant frequencies of the vowel [i]. In a two-cavity simplified model of [i], F_2 depends on the back cavity or pharynx while F_3 depends on the front or mouth cavity (Fant, 1973). The length of the back cavity (LB) and that of the front cavity (LF) can be approximated by the following:

$$LB = C/2F_2 \tag{4}$$

(C = speed of sound)

$$LF = C/2F_3 \tag{5}$$

These are only approximate values given the simplicity of the model. For Swedish speakers, Fant (1973: 90–91) reported that according to the above formulas the female back and front cavities were 2.1 cm and 1.25 cm shorter, respectively, than the corresponding male cavities; these predictions fitted the physiological data well.

2. Method

2.1. *Subjects*

A total of 40 subjects were chosen from 59 students participating in recording and listening sessions at the University of Texas at Austin (UT). They were selected from an age range of 18 to 27 years and formed four groups of 10 subjects each:

Korean males, Korean females, American males, and American females. The subjects were students attending UT and all had normal hearing and health. All the Korean subjects were born and educated in Seoul and spoke Standard Korean. The mean length of time in the U.S. was 3 years for male students, and 5 years for female students. American subjects were limited to those who indicated that the American South or Southwest was the area where they spent most of their lives and all spoke Southern or Southwestern dialects.

These 40 subjects were selected from the larger pool of 59 in two steps. First, the subjects were grouped homogeneously on the basis of collected information from a questionnaire which included subjects' age, sex, height, native language, fluency in other languages, dialect, and history of speech and hearing disorders. Second, scores by 8 judges (2 males and 2 females for each language group) were employed to screen out those subjects who were perceived as having deviant dialects within each language group. These 8 judges were a subset of an original group of 20 judges (10 from each language group). The 20 judges listened to a randomized recording of words containing the vowels [i, ɛ, a, u] (one token per vowel) produced by all speakers from that language group. While listening, judges put a check mark by any token perceived as deviant from their language group. The 4 Korean and 4 American judges having the fewest marks were selected as judges; scores for each subject from the 4 judges in each language group were then calculated. Subjects with more than 80% of their tokens perceived as being of a different dialect from the other members in the group were excluded from the study.

2.2. *Stimuli*

The speech samples consisted of 67 American English (hereafter (AE)) and 52 Korean words. Each AE vowel occurred in an /h(V)d/ context in which the vowel should not exhibit coarticulatory effects of the preceding consonant because /h/ is the voiceless variant of the following vowel; alveolar /d/ also has relatively little influence on the formants of the preceding vowel. In AE, 13 vowels /æ, a, ɔ, e, ɛ, i, ɚ, ɪ, ɑ, o, u, ʌ, ʊ/, as in *had, hard, hawed, hayed, head, heed, herd, hid, hod, hoed, who'd, Hudd,* and *hood,* were chosen. Each Korean vowel occurred in an /h(V)da/ context to approximate the frame chosen for English. In Korean that is a typical form. For example, each verb stem combines with the particle *da* or *h(V)da* to form a root infinitive. The 10 Standard Korean vowels investigated were /a, ɛ, e, i, o, ø, u, y, ʌ, ɨ/, as in *hada, hɛda, heda, hida, hoda, høda, huda, hyda, hʌda,* and *hɨda.*

These 10 Korean and 13 AE vowels appeared five times on the reading list in random order. Assuming that subjects needed time to adapt to the circumstances, extra tokens were added at the beginning and ending of each AE and Korean word list. These extra tokens and any unnaturally-produced tokens were discarded in the analysis. Three out of the five productions of each vowel for each subject were chosen for the average data set.

The recording was done in a sound-proof booth in the UT Phonetics Lab. Subjects read each word from a printed word list at a normal rate. The experimenter monitored the recording level throughout the session to avoid weak or overloaded signals. The recording took 2–3 minutes per subject.

2.3. *Data collection*

The input samples were low pass filtered at 4 kHz and digitized at a sampling rate of 10 kHz. A spectrogram of each word was made using a 256-point discrete Fourier transform (DFT) analysis with a 6.4-ms Hamming window once every millisecond. Most spectrograms showed steady states between vowel onset and offset points, but some showed continuous changes in the formant frequencies across the entire vowel making it difficult to identify a consistent time point for spectral analysis. Furthermore, AE vowels were substantially longer than Korean vowels: the average duration of AE vowels was 251 ms, with a standard deviation (SD) of 61 ms, while that of Korean vowels was 86 ms, with an SD of 32 ms. (This difference might result from syllable structure differences. The medial consonant in the Korean CVCV words was realized as voiceless [t] by a few subjects when they produced the word with two distinct syllables.) In view of these temporal differences, this study adopted a proportionate time point for spectral analysis to make comparison of the two vowel systems meaningful. Vowel onset and offset were determined by observing both the spectrogram and the amplitude tracing. On the spectrogram, each vowel tended to begin with a glottal pulse and clear formant bars following the weak noise of [h]. On the amplitude tracing, each vowel was represented by a periodic oscillation at about 40 dB preceded and followed by a nonperiodic consonant waveform, as illustrated in Fig. 1. Vowel onset was identified as the point where the 40 dB threshold was crossed. Vowel offset was assigned to the point where the amplitude fell and the formant bars terminated on the spectrogram.

Vowel onset and offset were used to determine total vowel duration. Formant frequency measures were taken one-third into the vowel (i.e., at the point determined by adding one-third of the total duration to vowel onset). Formant values were both automatically computed by a spectral analysis tool and visually verified using the spectrographic display; these methods almost always converged. f_0 was gathered from computer estimates by an autocorrelation method while checking its validity against the duration of a vocal fold pulse on the waveform. When formant values of the same vowel and subject showed wide variation, the author double-checked them by listening to and comparing the spectrograms of the three tokens. For reliability, measurements of selected tokens were independently made by a phonetician; there was less than 5% disagreement. Any systematic errors were immediately corrected after discussion.

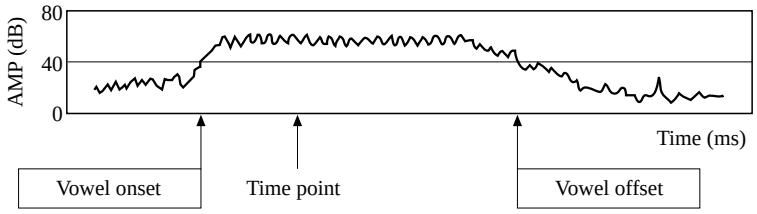

Figure 1. Illustration of waveform display and temporal labels. "Time point" refers to the location of the spectral analysis (one-third into the vowel).

TABLE I. Average values of f_0, the first three formants (F_1, F_2, F_3), and their standard deviations (in parentheses) for the American male speakers' vowels. ($n = 30$)

Vowel	f_0		F_1		F_2		F_3	
æ	126	(16)	687	(83)	1743	(113)	2497	(137)
a	125	(15)	638	(46)	1051	(74)	2318	(185)
ɔ	128	(20)	663	(62)	1026	(57)	2527	(171)
e	128	(18)	469	(36)	2082	(130)	2636	(168)
ɛ	132	(24)	531	(52)	1900	(84)	2561	(148)
i	136	(21)	286	(32)	2317	(104)	3033	(191)
ɚ	130	(20)	490	(32)	1363	(99)	1787	(165)
ɪ	130	(15)	409	(32)	2012	(110)	2671	(148)
ɑ	127	(15)	694	(89)	1121	(85)	2548	(136)
o	129	(18)	498	(41)	1127	(93)	2375	(131)
ʊ	135	(21)	446	(46)	1331	(102)	2380	(125)
ʌ	127	(15q	592	(45)	1331	(71)	2494	(167)
u	135	(17)	333	(33)	1393	(213)	2282	(114)

TABLE II. Average values of f_0, the first three formants (F_1, F_2, F_3), and their standard deviations (in parentheses) for the American female speakers' vowels. ($n = 30$)

Vowel	f_0		F_1		F_2		F_3	
æ	209	(22)	825	(81)	2059	(208)	2928	(95)
a	205	(19)	782	(106)	1287	(97)	2563	(173)
ɔ	206	(19)	777	(86)	1140	(91)	2895	(143)
e	209	(19)	521	(70)	2536	(138)	2991	(77)
ɛ	211	(20)	631	(57)	2244	(190)	2968	(84)
i	221	(23)	390	(32)	2826	(140)	3416	(162)
ɚ	218	(22)	523	(69)	1550	(110)	1927	(254)
ɪ	216	(24)	466	(51)	2373	(164)	3014	(94)
ɑ	205	(18)	857	(92)	1255	(85)	2877	(168)
o	207	(17)	528	(73)	1206	(183)	2824	(143)
ʊ	214	(20)	491	(56)	1486	(172)	2836	(154)
ʌ	206	(18)	701	(75)	1641	(89)	2901	(108)
u	228	(27)	417	(29)	1511	(326)	2796	(169)

3. Data analysis and discussion

Three repetitions of each AE and Korean vowel by each subject were averaged: the AE data consisted of (20 subjects) × (13 vowels) × (4 estimates: three formants and one f_0) = 1040; the Korean data included (20 subjects) × (10 vowels) × (4 estimates) = 800. Tables I and II list f_0 and the average formant values for the American male and female speakers. Tables III and IV list those values for the Korean speakers.

The percent difference (*Diff.*) between the male and female f_0 and formant frequencies was calculated to identify patterns of gender variation according to

$$\textit{Diff.}\ (\%) = \{(F_n f - F_n m)/F_n m\} \times 100 \tag{6}$$

TABLE III. Average values of f_0, the first three formants (F_1, F_2, F_3), and their standard deviations (in parentheses) for the Korean male speakers' vowels. ($n = 30$)

Vowel	f_0		F_1		F_2		F_3	
a	162	(25)	738	(87)	1372	(124)	2573	(127)
ɛ	165	(22)	591	(75)	1849	(106)	2597	(110)
e	167	(26)	490	(105)	1968	(150)	2644	(94)
i	172	(24)	341	(29)	2219	(176)	3047	(146)
o	170	(25)	453	(47)	945	(134)	2674	(156)
ø	166	(24)	459	(69)	1817	(163)	2468	(134)
u	174	(27)	369	(43)	981	(141)	2565	(173)
y	174	(26)	338	(30)	2114	(140)	2729	(213)
ʌ	165	(25)	608	(76)	1121	(110)	2683	(145)
ɨ	174	(26)	405	(37)	1488	(176)	2497	(80)

TABLE IV. Average values of f_0, the first three formants (F_1, F_2, F_3), and their standard deviations (in parentheses) for the Korean female speakers' vowels. ($n = 30$)

Vowel	f_0		F_1		F_2		F_3	
a	264	(26)	986	(107)	1794	(108)	2957	(227)
ɛ	263	(29)	677	(108)	2285	(169)	3063	(141)
e	263	(26)	650	(113)	2377	(77)	3068	(117)
i	271	(29)	344	(48)	2814	(168)	3471	(177)
o	269	(31)	499	(60)	1029	(143)	3068	(159)
ø	265	(29)	602	(109)	2195	(152)	3013	(132)
u	278	(28)	422	(83)	1021	(139)	3024	(138)
y	272	(30)	373	(62)	2704	(95)	3222	(108)
ʌ	263	(28)	765	(125)	1371	(108)	3009	(183)
ɨ	279	(30)	447	(68)	1703	(106)	2997	(173)

Diff. in f_0 varied from an average of 63% (SD = 3%) for the AE speakers to 59% (SD = 2%) for the Korean speakers. For formant frequencies, the English *Diff.* was 18% (SD = 8%) for F_1, 16% (SD = 6%) for F_2, and 15% (SD = 4%) for F_3; Korean *Diff.* was 18% (SD = 11%) for F_1, 20% (SD = 9%) for F_2, and 17% (SD = 3%) for F_3. Average standard deviation for F_1 to F_3 was 6% for the English speakers and 8% for the Korean speakers, suggesting that there is about a 6–8% variation within each language group which may not be removed by uniform or non-uniform normalization procedures. *Diff.* of F_1 varied from 7% in [ɝ] to 36% in [i] in AE, while in Korean it ranged from 10% in [ɨ] to 34% in [a]. If we use any formant-specific normalization procedure proposed by Fant (1975), we may get about 2% more reduction since the difference between F_1–F_3 is 2% for English and Korean. In other words, if the AE female data were uniformly scaled down by 16%, female F_2 would be exactly scaled to that of male speakers but F_1 would deviate by 2%, while F_3 would deviate by 1%. Similarly, an 18% shift of the Korean female data to those of males would lead to 2% deviation of F_2, and 1% deviation of F_3. However, whether the 2% improvement by the non-uniform scaling is necessary or not should be

evaluated perceptually because human listeners can recognize by context some vowels with only partial acoustic information (e.g., Strange, 1989).

3.1. *Non-linguistic factors*

We begin by applying the three non-linguistic factors considered in the Introduction to the current data set. Across vowels, the average f_0 for AE speakers was 130 Hz (SD = 18 Hz) for males and 212 Hz (SD = 21 Hz) for females; for Korean speakers it was 169 Hz (SD = 25 Hz) for males and 269 Hz (SD = 29 Hz) for females. For open vowels with higher F_1, f_0 was lower. This phenomenon is the well-known "vowel inherent pitch" effect (Lehiste, 1967). A question arises as to whether f_0 can serve in part as an independent source of speaker normalization. Statistically, the present data showed a strong negative correlation between f_0 and F_1 in males ($r = -0.84$ for the Americans, and $r = -0.92$ for the Koreans). The correlation among the females was weaker ($r = -0.79$ for the Americans; $r = -0.73$ for the Koreans). Thus, f_0 may be used for F_1 scaling with a 70–84% possibility of correctness (i.e., r^2 range). But the correlation was almost negligible ($r < 0.39$) between f_0 *vs.* F_2 or F_3.

The second factor is overall vocal tract length. Using Equation (1), overall vocal tract length was estimated from F_3 of the vowel in *Hudd* for American subjects and from F_3 of the vowel [ʌ] for Korean subjects. Estimated average vocal tract length was 17.1 cm (SD = 1.1 cm) for American males and 14.7 cm (SD = 0.5 cm) for females, while the corresponding Korean values were 15.9 cm (SD = 0.9 cm) for males and 14.2 cm (SD = 0.9 cm) for females. Thus, the language groups differ by roughly 1 cm for males and less than 1 cm for females. However, from Equation (1), one centimeter difference in vocal tract length leads to roughly a 200 Hz difference in the frequency of F_3 (e.g., F_3 is 3036 Hz for a vocal tract 14 cm long compared to 2833 Hz for a 15 cm vocal tract). Consequently, comparison of the two language groups should incorporate the vocal tract length difference because it should result in higher formant frequencies for the Korean speakers.

The third factor is the difference in size of mouth and pharynx cavities. The lengths of the back and front cavities were estimated from AE and Korean [i] using Equations (4) and (5). The average LF was 7.3 cm and 7.7 cm for American and Korean males respectively, and 6.0 cm for both American and Korean females. The average LB was about 5.6 cm for both American and Korean males, and 5.0 cm for American females and 4.9 cm for Korean females. Fig. 2 shows the average values and +/− one standard deviation bar.

In Fig. 2, there are clear gender and language differences. Male and female speakers have a back cavity that is longer than the front cavity, but the difference isn't as great for the female speakers. Across languages, the average difference for front cavity length of males and females is small, but the difference for back cavity length is almost twice that of the front cavity. The average *Diff.* of all the Korean speakers' F_1 for the Korean vowel [i] was less than 1% whereas for F_2 it was about 27%. Since F_2 of [i] depends on the length of the pharynx, this difference may provide evidence for the non-uniform shortening of this cavity in the Korean data. Another point to observe in the figure is that the total length of LB and LF is slightly longer for the Korean males than for the American males and the opposite is true between the Korean and American female groups. This may be surprising in

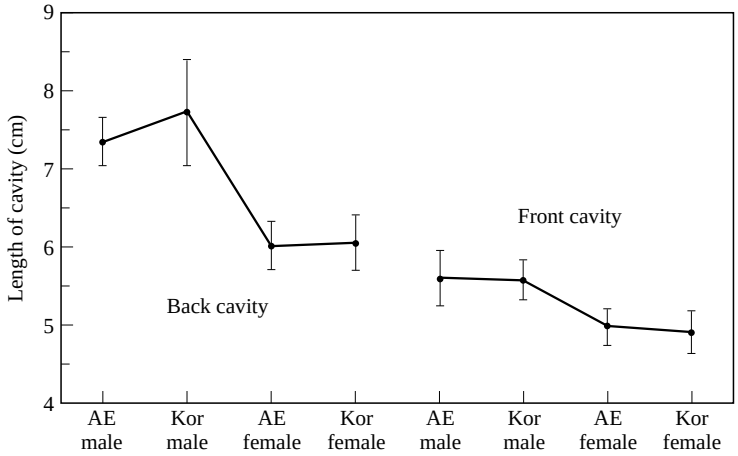

Figure 2. Average length of front and back cavities of American English and Korean groups. (Vertical bars indicate +/− one standard deviation.) The estimates are based on formant patterns of [i]. Kor = Korean; AE = American English.

view of the larger vocal tract estimates for the American males. This disagreement might come either from idiosyncratic anatomical configurations of the Korean male speakers for vowel [i] or from wide group variation, which is shown by greater standard deviation of the Korean male LB. Further studies on the relationship between physiological and acoustical aspects of vowels of the two languages is needed to clarify the problem.

3.2. *Normalization*

Now, what would be required for a more precise cross-linguistic comparison? Specifically, how can the linguistically irrelevant differences be reduced? In view of the large differences across groups observed here in vocal tract length, we adopted the Nordström & Lindblom model (1975) for this purpose. The AE male vowels were the reference for normalization. First, the AE and Korean female data were uniformly scaled to those of the male data (within-language normalization). Second, the Korean male and female data were uniformly scaled to those of the reference AE male data (across-language normalization). Here an average F_3 of open vowels whose F_1 is greater than 600 Hz for the AE male data (F_3mAE) and the equivalent for the female data (F_3fAE) are calculated by

$$F_3\text{mAE} = (2318 + 2527 + 2497 + 2548)/4 = 2472.5$$

$$F_3\text{fAE} = (2968 + 2901 + 2895 + 2563 + 2928 + 2877)/6 = 2855.3$$

Then, a uniform scale factor k was determined according to

$$k = F_3\text{mAE}/F_3\text{fAE}$$
$$= 2472.5/2855.3 = 0.86593$$

This scale factor was applied uniformly to the AE female data. As the scale factors indicate, we may expect the result to be a uniform reduction of the AE female

TABLE V. The first three formants (F_1, F_2, F_3) of the American female speakers scaled to those of the American males

Vowel	F_1	F_2	F_3
æ	714	1783	2535
a	677	1114	2219
ɔ	673	987	2507
e	451	2196	2590
ɛ	546	1943	2570
i	338	2447	2958
ɚ	453	1342	1669
ɪ	404	2055	2610
ɑ	742	1087	2491
o	457	1044	2445
ʊ	425	1287	2456
ʌ	607	1421	2512
u	361	1308	2421

TABLE VI. The first three formants (F_1, F_2, F_3) of the Korean female speakers scaled to those of the Korean males

Vowel	F_1	F_2	F_3
a	857	1560	2571
ɛ	589	1987	2664
e	565	2067	2668
i	299	2447	3018
o	434	895	2668
ø	524	1909	2620
u	367	888	2630
y	324	2351	2802
ʌ	665	1192	2617
ɨ	389	1481	2606

vowel space. Similarly, a uniform scale factor between the Korean male and female data was calculated to be 0.8696. Tables V and VI list the scaled data cross–linguistically.

Next, the vocal tract ratio k between the AE males and the equivalent of the Korean males (F_3mKor) was calculated by

$$k = F_3mAE/F_3mKor$$
$$= 2472.5/2628 = 0.94083$$

Table VII lists the Korean male data scaled cross-linguistically. Also, the scale factor k between the AE males (F_3mAE) and the Korean females scaled (F_3fKor.sc) was estimated as

$$k = F_3mAE/F_3fKor.sc$$
$$= 2472.5/2594 = 0.95316$$

Table VII lists the Korean female data normalized by the scale factor.

TABLE VII. The first three formants (F_1, F_2, F_3) of the Korean male and female speakers scaled to those of the American males

	Male speakers			Female speakers		
Vowel	F_1	F_2	F_3	F_1	F_2	F_3
a	694	1291	2421	817	1487	2451
ɛ	556	1740	2443	561	1894	2539
e	461	1852	2488	539	1970	2543
i	321	2088	2867	285	2332	2877
o	426	889	2516	414	853	2543
ø	432	1709	2322	499	1820	2497
u	347	923	2413	350	846	2507
y	318	1989	2568	309	2241	2671
ʌ	572	1055	2524	634	1136	2494
ɨ	381	1400	2349	371	1412	2484

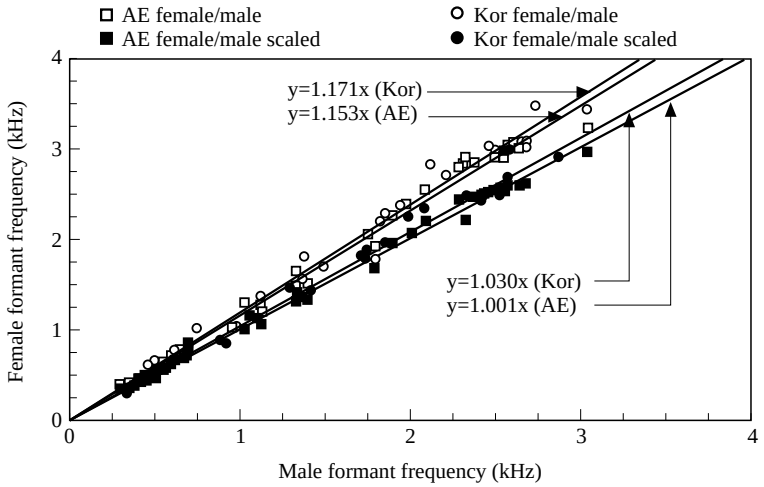

Figure 3. Relationship between female and male formant frequencies for American English and Korean vowels before and after uniform scaling. x axis shows male formant frequencies (in Hz) and y axis indicates those of females. Regression lines and their slopes are given.

How much does the uniform scaling method reduce non-linguistic factors within and across the languages? We compared the original male/female data and those that were scaled within and across the languages. Fig. 3 shows the differences between the raw and scaled data (indicated by the corresponding unfilled and filled symbols). In the original data, both language groups show a trend toward higher formant frequencies for female speakers. The difference between female and male vocal tract lengths seems to be the main determinant of the deviation from the line of identity (i.e., $y = x$). Noting that the uniform scale factors between male and female speakers were 0.86593 for AE and 0.8696 for Korean, we can estimate that

female vocal tracts are about 14% shorter than those of males. Regressional analyses between the male and female data support that estimate. The line superimposed on the scaled data points has a slope of 1.001($r^2 = 0.995$) and that of Korean is 1.03 ($r^2 = 0.993$). The difference in slope between the original data and the scaled data is about 0.14 (i.e., 14%).

Within languages, we find that some female vowels fall on those of the reference male. The average frequency difference between AE males and females scaled is 26 Hz in F_1 (SD = 15 Hz), 64 Hz in F_2 (SD = 35 Hz), 61 Hz in F_3 (SD = 39 Hz). The average difference between Korean males scaled and females scaled is 39 Hz in F_1 (SD = 38 Hz), 118 Hz in F_2 (SD = 77 Hz), 62 Hz in F_3 (SD = 49 Hz). F_2 of the Korean male vowel [ɨ] exactly falls on that of the Korean female. For the Korean front vowels, F_2 and F_3 show some systematic variation. When F_2 goes up, F_3 goes down. This comes from the uniform shift along the reference line.

In addition, we examined the vocal tract length of American and Korean subjects from F_3 of scaled vowel [ʌ]. The scaled values closely fitted the reference AE male ones: the AE male, 2494 Hz; the AE female, 2512 Hz; the Korean male, 2524 Hz; the Korean female, 2494 Hz. This suggests that the uniform scaling was successful in correcting the vocal tract difference among the groups. However, the front and back cavity ratio estimated from the F_2 and F_3 values of the AE and Korean vowel [i] suggests the scaling was not that successful. Here we observe some differences in formant frequencies between and within genders and languages. For example, differences in formant values are 229 Hz (F_2), and 156 Hz (F_3) between the AE and Korean males; 115 Hz (F_2), and 156 Hz (F_3) between the AE and Korean females. Some of the remaining gender differences may be attributable to anatomical differences between male and female vocal tracts or vocal cords; others may come from the less defined spectral envelopes of female speakers. For example, Ryalls & Lieberman (1942: 1632) pointed out that ''fewer 'samples' of the spectrum yielded less acoustic information'' which leads to higher error rates in the perception of vowels with higher pitch. Similarly, formant peaks may be influenced by gender differences in f_0.

3.3. *Cross-language comparison*

Because there still exist gender differences in the vocal tract ratio after normalization, we compared the male and female groups separately. Fig. 4 shows the frequencies of F_1 and F_2 of the AE and Korean male speakers. Fig. 5 represents F_3/F_1 for the male speakers. Figs. 6 and 7 show the corresponding values for the female speakers. For clarity, those vowels cornering on the general vowel space are plotted. The vowel space of each language is shown with adjacent vowel points connected peripherally. A thicker line connects the Korean vowels, while a thinner line connects the AE vowels. The physical frequency scale has been converted to a perceptual dimension, mel (Fant, 1973), in order to better approximate the perceived distances among the vowels.

In Figs. 4 and 6, the Korean vowel space appears wedge-shaped with [i, a, u] at the corners; the American English vowel space looks more rectangular with [i, u, æ, ɑ] at the corners. These general shapes are contracted somewhat in the F_3 dimension but are still retained in Figs. 5 and 7. Korean female [e] and [ɛ] are close together, suggesting that Korean female speakers might not make a distinction

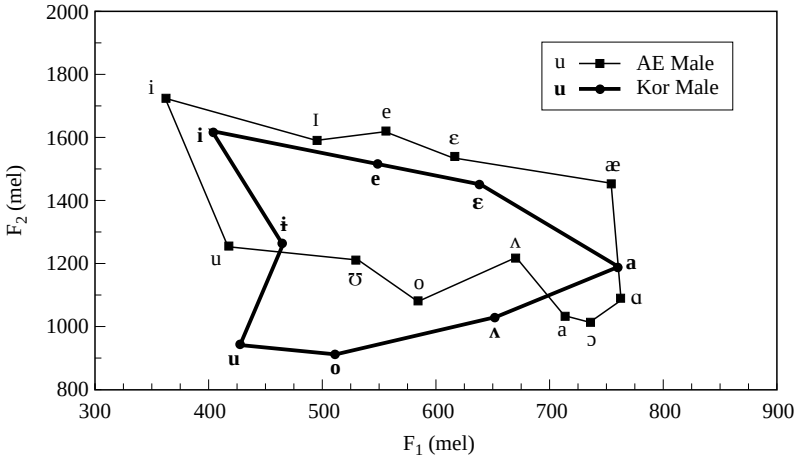

Figure 4. Superimposed F_1/F_2 (in mel) vowel spaces of American English and Korean male speakers normalized. Phonetic symbols are given near the formant position.

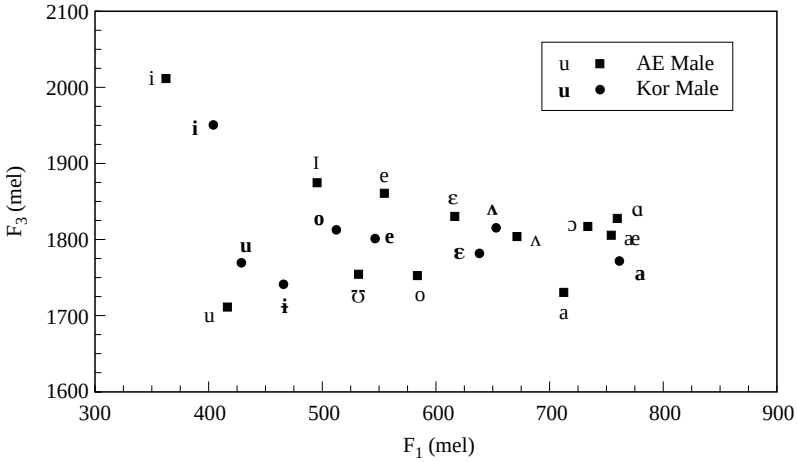

Figure 5. Superimposed F_1/F_3 (in mel) vowel spaces of American English and Korean male speakers normalized.

between these vowels. (This distinction is being lost in Korean. For example, although /e/ and /ɛ/ contrast in the pair [more] (''the day after tomorrow'') and [morɛ] (''sand''), /nɛilmore/ (''tomorrow and the day after tomorrow'') may be realized as [nɛilmorɛ] without any communicative problem.)

The normalized vowel spaces of American English and Korean differ from each other, raising the third research question: What are the theoretical implications of such cross-linguistic differences? We explored the predictions of Lindblom's theory of adaptive dispersion (Lindblom & Engstrand, 1989; Lindblom, 1990) because it might offer an explanation for the differently-shaped vowel systems found here. Lindblom assumes that speakers control, not the acoustic invariance of speech

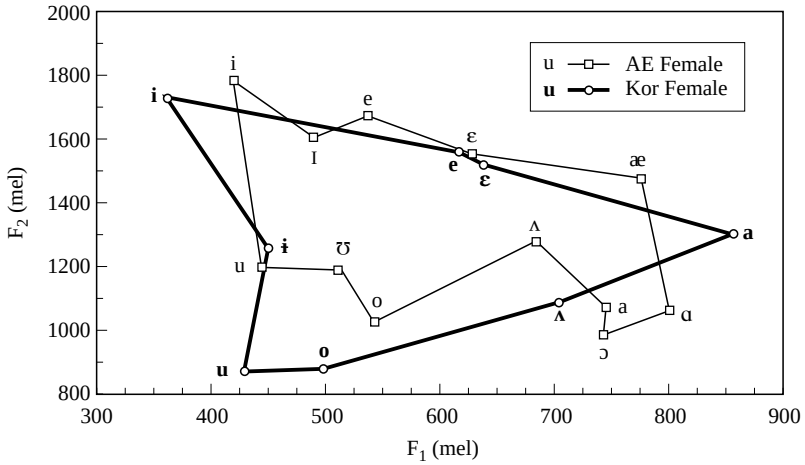

Figure 6. Superimposed F_1/F_2 (in mel) vowel spaces of American English and Korean female speakers normalized.

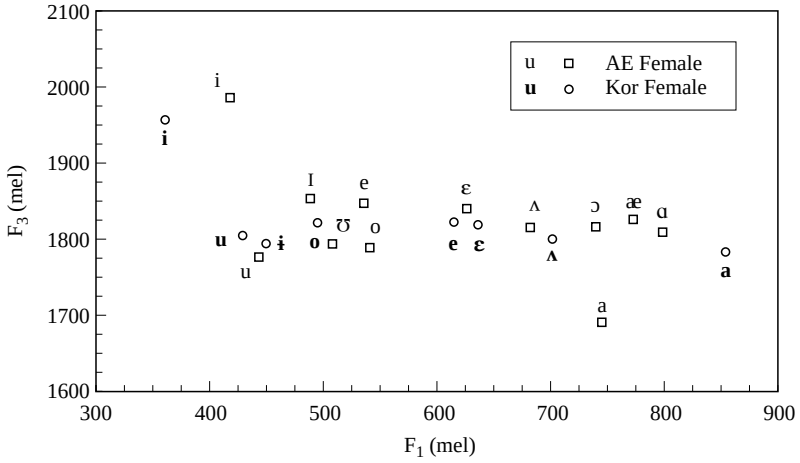

Figure 7. Superimposed F_1/F_3 (in mel) vowel spaces of American English and Korean female speakers normalized.

sounds, but "sufficient perceptual contrast", monitoring a tradeoff between articulatory economy and perceptual distinctiveness. This notion of sufficient contrast can be applied to the vowel systems examined here. To discuss the contrast numerically, we employed Lindblom's (1990: 21) perceptual distance, D_{ij}, defined as the Euclidean distance between two vowel points:

$$D_{ij} = \sqrt{(M1_i - M1_j)^2 + (M2_i - M2_j)^2} \quad (7)$$

in which i and j indicate two different vowels while $M1$ is F_1 frequency in mels.

For example, in Figs. 4 and 6 there are greater cross-language differences for the vowels [u] and [a] than for the others. The Korean vowel inventory has more high tense vowels than does that of American English (Korean /i, y, ɨ, u/ *vs.* English /i, u/) leading to the prediction that the AE vowel [u] can have a somewhat higher

F_2 without crowding into the space of another vowel of similar duration. On the other hand, if Korean [u] were to have a high F_2 it might be confused with Korean [ɨ]. In this respect, sufficient perceptual distance might account for the relatively low F_2 values for Korean [u]. In Fig. 4, for example, American English D_{iu} is 474 mel while that of Korean D_{iu} is 685 mel; however, Korean $D_{i\ddot{i}}$ is 370 mel while Korean $D_{\ddot{i}u}$ is 322 mel. Notice that AE [u] almost overlaps with Korean [ɨ] rather than Korean [u]. Liljencrants & Lindblom's (1972) original work on adaptive dispersion predicted that AE and Korean /u/ would have the same F_2 values because their original theory predicted that languages would tend toward maximal phonetic contrast. However, these data support the elaboration of that theory and the condition of sufficient contrast, advocated in Lindblom's later work.

For the low vowels, AE [a] and [æ] must employ extreme values of F_2 to avoid confusion. The AE speakers separate the two vowels by around 400 mel, as in D_{iu}. On the other hand, Korean [a] is not crowded by other vowels (see Figs. 4 and 6) so that it may be placed at the corner of a regular triangle formed by the acoustically closest vowels [ɛ] and [ʌ]. Similarly, sufficient perceptual distance is also maintained between the AE tense and lax vowels. The greater distance between AE than Korean [i] and [e] may be linked to intervening [ɪ] in AE but not Korean. (For example, for the male speakers, AE $D_{i\imath}$ is 192 mel which almost equals the Korean D_{ie}.) Similar observations hold for AE and Korean [u] and [o], and intervening AE [ʊ]. Interestingly, those lax vowels are pushed "inside" the AE vowel space, increasing the perceptual distance to adjacent vowels. In general, the Korean vowel space shows an expansion of high vowels while the English vowel space shows an expansion of the low vowels.

In addition, we compared perceptual distances of Korean males with those of female speakers. Korean male D_{ia} is 560 mel but that of the Korean female vowels is 654 mel. Korean male D_{ua} is 416 mel while that of the female group is 603 mel. This suggests that the Korean female speakers produced vowels with a wider range of jaw movement than the male speakers since F_1 tends to increase if the jaw lowers.

We calculated the cross-language perceptual distances between AE and Korean vowels typically transcribed with the same IPA symbols as shown in Table VIII. It can be seen that the average perceptual distance between male and female speakers

TABLE VIII. Perceptual distance in mel between AE and Korean vowels denoted by the same IPA symbols

	Male speakers		Female speakers	
Vowel	F_1/F_2	F_1/F_3	F_1/F_2	F_1/F_3
a	167	65	252	143
ɛ	85	54	33	23
e	112	61	138	84
i	110	73	76	65
o	185	92	154	57
u	316	58	327	32
ʌ	183	22	188	25
Average	165	61	167	61

TABLE IX. *t* values for tests of cross-language differences in F_1–F_3 frequencies (in mel) for male and female speakers normalized cross-linguistically. (df = 18, * significant at $p > 0.05$)

	Male speakers			Female speakers		
Vowel	F_1	F_2	F_3	F_1	F_2	F_3
a	1.96	5.83*	1.50	3.72*	10.34*	3.08*
ɛ	0.93	3.95*	2.14	0.39	0.67	0.76
e	0.35	3.82*	2.54*	2.51*	5.29*	1.48
i	3.00*	3.62*	2.26*	3.70*	1.98	1.23
o	3.98*	4.99*	2.30*	1.78	3.02*	1.72
u	5.30*	7.72*	0.48	2.90*	7.86*	0.95
ʌ	0.83	7.28*	0.46	0.72	7.80*	0.37

is almost the same on the F_1/F_2 and F_1/F_3 dimensions. Averaging across male and female speakers, the greatest distances on the F_1/F_2 dimension are found for [u] and [a]. On the F_1/F_3 dimension, the shortest distance is observed in the vowel [ʌ], suggestive of successful normalization.

Finally, we examined the question: Which cross-linguistic comparisons are reliably different? *t*-tests were conducted on the normalized AE and Korean male and female data in which there were 10 values (1 per speaker) for each formant of a given vowel in each language group. For that purpose, we compared only the 7 vowels typically transcribed with the same IPA symbols. Table IX lists the *t* values. Asterisks indicate *t* values that are significant at $p > 0.05$ (df = 18). The significant differences generally conform to the patterns described above (e.g., F_2 values for [u] and [a] differ in AE and Korean; F_1 values for [i] differ in the two languages). *t*-tests were also conducted between the Korean female vowel [ɨ] and the AE female [u] also showed no significant differences in F_1–F_3. This result is consistent with the small perceptual distances between the two vowels (56 mel on the F_1/F_2 dimension; 19 mel on the F_1/F_3 dimension) as well as with the author's impressionistic observations. On the other hand, the author has observed considerable similarities between AE and Korean vowels for which significant differences in formant frequencies were found. More rigorous perceptual studies, possibly using synthetic versions of these vowels, are needed to untangle these issues.

I would like to thank Drs. Björn Lindblom and Randy Diehl for their indispensable support of and enthusiasm for this research. I also thank Dr. Beddor and three reviewers of this paper for their patience and valuable comments.

References

Boothroyd, A. (1986) *Speech acoustics and perception.* Austin: Pro-Ed.
Chiba, T. & Kajiyama, M. (1941) *The vowel—its nature and structure.* Tokyo: Kaiseikan.
Disner, S. F. (1980) Evaluation of vowel normalization procedures, *Journal of the Acoustical Society of America,* **67**, 253–261.
Fant, G. (1968) Analysis and synthesis of speech processes. In *Manual of phonetics* (B. Malmberg, ed.), 243–253. Amsterdam: North Holland.
Fant, G. (1970) *Acoustic theory of speech production.* The Hague: Mouton.
Fant, G. (1973) *Speech sounds and features.* Cambridge, Massachusetts: MIT Press.
Fant, G. (1975) Speech Production, *STL-QPSR,* **2–3**, 1–19.

Fujisaki, H. (1972) Current problems in speech recognition. In *Research on information processing annual report,* **4**, 197–204. Tokyo: Japan.
Ladefoged, P. & Broadbent, D. E. (1957) Information conveyed by vowels, *Journal of the Acoustical Society of America,* **29**, 98–104.
Lehiste, I. (1967) *Readings in acoustic phonetics.* Cambridge, Massachusetts: MIT Press.
Lieberman, P. (1967) *Intonation, perception, and language.* Cambridge, Massachusetts: MIT Press.
Liljencrants, J. & Lindblom, B. (1972) Numerical simulation of vowel quality systems: the role of perceptual contrast, *Language,* **48**, 839–862.
Lindblom, B. & Engstrand, O. (1989) In what sense is speech quantal? *Journal of Phonetics,* **17**, 107–121.
Lindblom, B. (1990) Explaining phonetic variation: a sketch of the H–H theory. In *Speech production and speech modeling* (W. J. Hardcastle & A. Marchal, editors), 403–439. Dordrecht: Kluwer Publishers.
Miller, J. D. (1989) Auditory-perceptual interpretation of the vowel, *Journal of the Acoustical Society of America,* **85**, 2114–2134.
Negus, V. E. (1949) *Comparative anatomy and physiology of the larynx.* New York: Grune & Stratton.
Nordström, P. E. & Lindblom, B. (1975) A normalization procedure for vowel formant data. *Paper* 212 *at the international congress of phonetic sciences in Leeds,* August.
Peterson, G. E. & Barney, H. L. (1952) Control methods used in a study of the vowels, *Journal of the Acoustical Society of America,* **24**, 175–184.
Pols, L. C. W., Tromp, H. R. C. & Plomp, R. (1973) Frequency analysis of Dutch vowels from 50 male speakers, *Journal of the Acoustical Society of America,* **53**, 1093–1101.
Ryalls, J. H. & Lieberman, P. (1982) Fundamental frequency and vowel perception, *Journal of the Acoustical Society of America,* **72**, 1631–1634.
Strange, W. (1989) Dynamic specification of coarticulated vowels spoken in sentence context, *Journal of the Acoustical Society of America,* **85**, 2135–2153.
Syrdal, A. K. & Gopal, H. S. (1986) A perceptual model of vowel recognition based on the auditory representation of American English vowels, *Journal of the Acoustical Society of America,* **79**, 1086–1100.
Traunmüller, H. (1988) Paralinguistic variation and invariance in the characteristic frequencies of vowels, *Phonetica,* **45**, 1–29.
van Nierop, D. J. P. J., Pols, L. C. W. & Plomp, R. (1973) Frequency analysis of Dutch vowels from 25 female speakers, *Acustica,* **29**, 110–118.
Yang, B. (1990) *Development of vowel normalization procedures*: *English and Korean.* Seoul: Hanshin.
Yang, B. (1992) An acoustical study of Korean monophthongs produced by male and female speakers, *Journal of the Acoustical Society of America,* **91**, 2280–2283.