PERCEPTUAL CONTRAST IN THE KOREAN AND ENGLISH VOWEL SYSTEM NORMALIZED

Byunggon Yang

Dept. of English, *Dong-eui University*, 24 Kayadong, Pusanjingu, Pusan 614-714 KOREA

ABSTRACT

This study applied the uniform scaling method (Nordstroem and Lindblom, 1975) within and across the two languages to the formant data collected equivalently from 40 healthy subjects, which formed four groups of 10 subjects each: Korean males, Korean females, American males, and American females. Then, the formant values were converted to a perceptual unit, mel, and plotted on the first formant against the second formant axes. From the cross-language comparison we observed that the vowels are placed to maintain sufficient perceptual contrast within each vowel system which supports the notion of Lindblom's (1990) "sufficient perceptual contrast". There were greater cross-language differences for the vowels [u] and [a] than for the others. If Korean [u] has a relatively low F2 value it might be confused with Korean [ɨ]. Therefore, Korean [u] has a relatively low F2 value to keep sufficient perceptual distance from [i] and [u]. The AE speakers separate sufficiently the two low vowels [a] and [æ] by around 400 mel. On the other hand, Korean [a] is placed at the corner of a regular triangle formed by acoustically neighboring vowels [ɛ] and [ʌ] simply to maintain sufficient contrast because there're no competitive vowels along the F2 dimension. Similar perceptual contrast is maintained between the AE tense and lax vowels. Lax vowels are pushed inside the AE vowel space to secure sufficient perceptual distance to adjacent vowels. This way each vowel system seems to maintain sufficient perceptual contrast.

1. INTRODUCTION

Several attempts have been made to compare acoustic measurements of vowels from two different language populations. However, most conclusions derived from the comparison were marred greatly by not considering these intervening factors: (1) linguistic factors such as dialectal and sociolectal differences and (2) non-linguistic factors such as physical anatomy, age, gender, and emotional state of the speaker (Ladefoged and Broadbent, 1957; Traunmueller, 1988). Various normalization methods were proposed to reduce these nonlinguistic factors so that any cross-language comparison becomes meaningful. Auditorily-based proposals (Miller, 1989; Syrdal and Gopal, 1986) attempted to transform the raw formant data by using logarithmic scaling, the Bark (critical-band-rate scaling), the mel, or the Koenig scales. Articulatorily-based methods (Nordstroem and Lindblom, 1975; Fant, 1975) were concerned with removing anatomical differences in vocal tract length or in the ratio of pharynx to mouth cavity. In Nordstroem and Lindblom's procedure, the female formant frequencies were uniformly adjusted along a trajectory to a position within the male reference system by multiplying by the scale factor: the ratio of the average male vocal tract to the average female vocal tract. Specifically, their method involved estimating the total length of a subject's vocal tract from an average of F3 in vowels with F1 greater than 600 Hz. Then, the ratio k of the length of the average male vocal tract to the average female vocal tract length is determined from the average of the third male and female formant values. This uniform scale factor k is multiplied to the female formant frequencies.

This study applied the uniform scaling procedures across the languages. In other words, the American English (AE) male vowels were set to a reference system and the uniform scaling factors to the Korean male and female vowel formant values were determined and applied to make the cross-language comparison meaningful. Finally, the formant values were transformed into the auditory-perceptual unit, mel, to observe the perceptual contrast in the Korean and English vowel system normalized.

2. METHOD

2.1. Subjects and Stimuli

59 students took part in recording and listening sessions at the University of Texas at Austin (UT). A total of 40 subjects 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. American subjects spoke Southern or Southwestern dialects. These subjects were selected from the larger pool in two screening procedures. First, information from a questionnaire was employed to group them homogeneously. Second, scores by 8 judges (2 males and 2 females for each language group) were used to exclude those subjects who were perceived as having deviant dialects within each language group.

Each English vowel occurred in an /h(V)d/ context in which the vowel do not exhibit coarticulatory effects of the preceding consonant because /h/ is the voiceless variant of the following vowel. In American English, 12 vowels /æ ɑ ɔ e ɛ i ɪ ɑ o ʊ ʌ u /, as in *had, hard, hawed, hayed, head, heed, 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. The eight Standard Korean vowels investigated were / a ɛ e i o u

ʌ i /, as in *hada, hɛda, heda, hida, hoda, huda, hʌda and hɨda.*

These eight Korean and twelve English vowels appeared five times on the reading list in random order. 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 produced each word from a printed word list at a normal rate. These controlled experimental settings might have caused some unnatural productions from the subjects. However, our main concern was to control the subjects as much as possible so that any non-linguistic factors would not intervene in the data collection procedures.

2.2. Data Collection

The input samples were low pass filtered at 4 kHz and digitized at a 10-kHz sampling rate. 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 spectrum analysis. Furthermore, English vowels were produced much longer than Korean vowels: the average duration of English vowels was 251 ms, with a standard deviation of 61 ms, while that of Korean vowels was 86 ms, with a standard deviation of 32 ms. In view of these temporal differences, this study adopted a proportionate time point for spectral analysis to make a 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. 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.

Formant frequency measures were taken at one-third of the total duration of the vowel portion. Formant values were both automatically computed by a spectrum analysis tool and visually verified using the spectrograms; these methods almost always converged. When formant values of the same vowel and subject showed a wide variation, the author double-checked them by listening to and comparing the spectrograms of the three tokens.

3. DISCUSSION

3.1. Cross-Language Normalization

The Nordstroem and Lindblom model (1975) was adopted for the normalization within and across the languages. The English male vowels were set to the reference for normalization. First, the English and Korean female data were uniformly scaled to those of the male data (within-language normalization). A uniform scale factor k between the American English male and female data was determined to be 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 vowel space. Similarly, another uniform scale factor between the Korean male and female data was calculated to be 0.8696. These two scale factors between the male and female data indicate that there is about 14% gender difference in the vocal tract length within each language group, which can be removed by the uniform normalization.

Second, the Korean male and female data were uniformly scaled to those of the reference American male data (across-language normalization). The uniform scale factor k between the AE males and the equivalent of the Korean males was calculated to be 0.94083. Also, that between the AE males and the Korean females scaled was estimated to be 0.95316. Here again, these factors across the two languages imply that there is a systematic non-linguistic factors involved in the formant data, which might have influenced on mixed results from any simple cross-language comparison.

3.2. Cross-language Comparison

Fig. 1 shows the first and second formant frequencies in mel of the AE and Korean male speakers. Fig 2 shows the corresponding values for the female speakers.

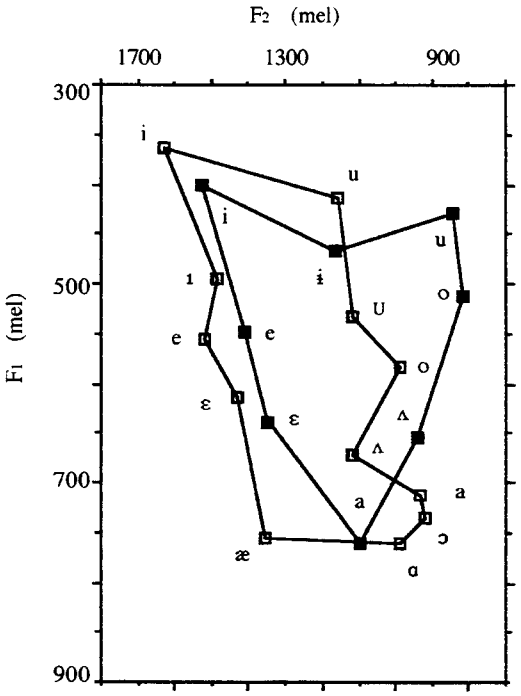

Figure 1: Superimposed F1/F2 (in mel) vowel spaces of American English and Korean male speakers normalized. Filled squares are vowel points of the Korean males.

For clarity, only those vowels cornering on the general vowel

space are plotted along axes shown in the figures. The axes have been arranged in such a way that they could meet traditional articulatory descriptions because formant frequencies are inversely related to the articulatory parameters (Ladefoged, 1982). The vowel space of each language is shown with adjacent vowel points connected peripherally. Each phonetic symbol is placed near the vowel point. Korean vowels are described inside the vowel space while English ones are marked outside the space. The physical frequency scale has been converted to a perceptual dimension, mel (Fant,1973), in order to better approximate the perceived distances in the vowel space. To discuss the perceptual contrast numerically, we determined Lindblom's (1990:21) perceptual distance, Dij, defined as the Euclidean distance between two vowel points in the following equation:

$$Dij = [(M1i - M1j)^2 + (M2i - M2j)^2]^{1/2}$$

in which i and j indicate two different vowels while M1 is F1 frequency in mel.

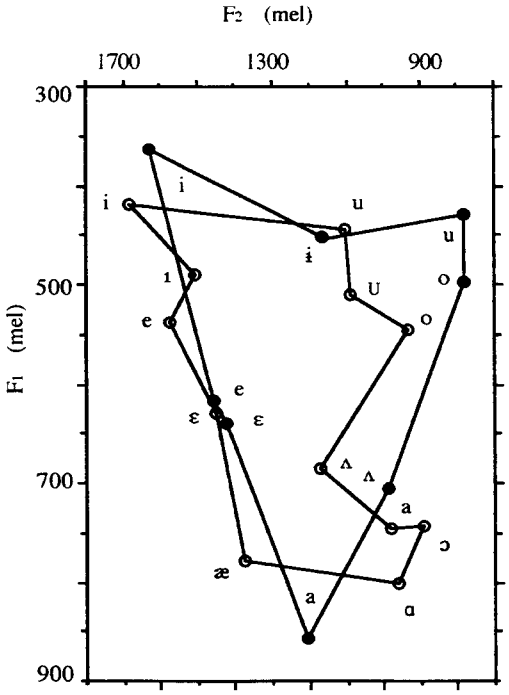

Figure 2: Superimposed F1/F2 (in mel) vowel spaces of American English and Korean female speakers normalized. Filled circles are vowel points of the Korean females.

We compared perceptual distances of Korean males with those of female speakers. Korean male Dia is 560 mel but that of the Korean female vowels is 654 mel. Korean male Dua 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 F1 tends to increase if the jaw moves down. Since there is a great gender difference in the vowel spaces we will compare the male and female groups separately.

The normalized vowel spaces of American English and Korean differ from each other. In Figs. 1 and 2, 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. The Korean vowel space shows an expansion of high vowels while the English vowel space shows an expansion of the low vowels. Korean female [e] and [ɛ] are close together, suggesting that Korean female speakers might not make a distinction between these vowels.

Why do we have such a different vowel space? The notion of sufficient contrast in Lindblom's theory of adaptive dispersion (Lindblom and Engstrand, 1989; Lindblom, 1990) may offer an explanation for the question. Lindblom assumes that speakers control, not the acoustic invariance of speech sounds, but "sufficient perceptual contrast," monitoring a tradeoff between articulatory economy and perceptual distinctiveness. Nooteboom(1983:193) also suggested that the production of words depends partly on "speaker's internalized model of the perceptual and interpretative processes in the listener." In other words, a speaker continuously shapes his production in relation to the process of distinctive perception.

In the figures, greater cross-language differences exist between the vowels [u] and [a] than between the others. The Korean vowel inventory has more high tense vowels than does that of American English, (Korean /i ɨ u/ vs. English /i u/) leading to the prediction that the AE vowel [u] can have a somewhat higher F2 without crowding into the space of another vowel. On the other hand, if Korean [u] were to have a high F2 it might be confused with Korean [ɨ]. In this respect, sufficient perceptual distance might account for the relatively low F2 values for Korean [u]. In Fig. 1, for example, American English Diu is 474 mel while that of Korean Diu is 685 mel; however, Korean Diɨ is 370 mel while Korean Diu is 322 mel. AE [u] almost overlaps with Korean [ɨ] rather than Korean [u]. Liljencrants and Lindblom's (1972) original work on adaptive dispersion predicted that AE and Korean /u/ would have the same F2 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 F2 to avoid confusion. The AE speakers separate the two vowels by around 400 mel, as in Diu. On the other hand, Korean [a] is not crowded by other vowels so that it may be placed at the corner of a regular triangle formed by acoustically closer vowels [ɛ] and [ʌ]. Similarly, sufficient perceptual contrast 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 lax vowel [ɪ] in AE but not Korean. For example, for the male speakers, AE Diɪ is 192 mel which almost equals the Korean Die. Similar observations hold for AE and Korean [u] and [o], and intervening AE [ʊ]. Those lax vowels are pushed "inside" the AE vowel space, securing the sufficient perceptual contrast to adjacent vowels. We observed that when there are more peripheral vowels along the first formant dimension or that of the second formant, the vowels are crowded with a shorter perceptual distance while fewer vowels along the

axes invoke wider perceptual distance. This way each vowel system seems to maintain sufficient perceptual contrast.

This study focused on the controlled acoustic data converted to the perceptual unit, which may be a limitation of the study. Further perceptual studies using a sophisticated speech synthesizer including more language data are desirable to confirm the findings.

4. REFERENCES

1. Fant, G. (1973). *Speech Sounds and Features.* Massachusetts: MIT Press.
2. Fant, G. (1975). Speech Production. *STL-QPSR,* 2-3, 1-19.
3. Ladefoged, P. (1982) *A Course in Phonetics.* New York: Harcourt Brace Jovanovich, Inc.
4. Ladefoged, P. and Broadbent, D.E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America,* 29, 98-104.
5. Liljencrants, J. and Lindblom, B. (1972). Numerical simulation of vowel quality systems: the role of perceptual contrast. *Language,* 48, 839-862.
6. Lindblom, B. and Engstrand, O. (1989). In what sense is speech quantal? *Journal of Phonetics,* 17, 107-121.
7. Lindblom, B. (1990). Explaining phonetic variation: a sketch of the H-H theory. In *Speech Production and Speech Modeling* (W.J. Hardcastle and A. Marchal, editors), Dordrecht:Kluwer Publishers.
8. Miller, J.D. (1989). Auditory-perceptual interpretation of the vowel. *Journal of the Acoustical Society of America,* 85, 2114-2134.
9. Nooteboom, S. G. (1983). Is speech production controlled by Speech Perception? In *Sound Structures* [M. van den Broecke, V. van Heuven, and W. Zonneveld, editors], Dordrecht: Foris Publications.
10. Nordstroem, P.E. and Lindblom, B. (1975). A normalization procedure for vowel formant data. *Paper 212 at the International Congress of Phonetic Sciences in Leeds,* August.
11. Syrdal, A.K. and Gopal, H.S. (1989). 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.
12. Traunmueller, H. (1988). Paralinguistic variation and invariance in the characteristic frequencies of vowels. *Phonetica,* 45, 1-29.