Perceptual and Motor Skills, 1987, 65, 837-838.

RESEARCH NOTE ON PERCEPTUAL FEATURES AND AUDITORY REPRESENTATIONS[1]

ROBERT ALLEN FOX AND JEAN BOOTH

The Ohio State University

Summary.—It has been argued that bark-scale transformed formant frequency values more accurately reflect auditory representations of vowels in the perceptual system than do the absolute physical values (in Hertz). In the present study the perceptual features of 15 monophthongal and diphthongal vowels (obtained using multidimensional scaling) were compared with both absolute and bark-scale transformed acoustic vowel measures. Analyses suggest that bark-transformation of the acoustic data does not necessarily produce better predictions of the vowels' perceptual space.

Multidimensional scaling (MDS) experiments have often been used to determine the salient perceptual features or dimensions used in the perception of vowels (2, 3, 4). The obtained perceptual dimensions (indicating the perceptual space of the vowels) are assumed to reflect the parameters critical to vowel identification. It is most common for a researcher to define each obtained perceptual dimension in terms of the acoustic characteristics of the vowels themselves by correlating the dimensional coordinates of each vowel with a set of relevant acoustic measures (e.g., fundamental frequency, formant frequencies, and duration).

However, in the last decade, several studies have suggested that absolute formant frequencies (in Hz) do not accurately reflect the auditory representation of a vowel in the perceptual system and actually exacerbate problems such as speakers' normalization (1, 5). Syrdal and Gopal (5), in particular, state that a more optimal model of vowel recognition should be based on auditory representations. They argue for a model in which the physical frequency measures (in Hz) are transformed into a more appropriate auditory scale—the bark scale, corresponding to the critical bands of the auditory system (6). The present study evaluated whether or not such transformed acoustic data better account for the *perceptual* features of vowels as obtained in a multidimensional scaling experiment.

The basic data were obtained from a study by the first author (2). This was a multidimensional scaling study of 15 monophthongal and diphthongal vowels which found that a four-dimensional solution best accounted for the obtained perceptual distance data (accounting for 65% of the variance). The data described in this study include the vowel coordinates for each of the four dimensions in the obtained perceptual space (corresponding to front/back, vowel height, low-back onset, and fundamental frequency, respectively) and the acoustic characteristics of the 15 vowels (including F1, F2, and F3 frequencies at four different locations in each vowel, F0, and duration).

In the present study these acoustic measures were transformed into the bark scale using according to the formula

$$b = 13 \arctan(0.76f) + 3.5 \arctan(f/7.5)^2$$

[1]Send reprint requests to Dr. Robert Allen Fox, Speech & Hearing Science, 324 Derby Hall, 154 Oval Mall, Columbus OH 43210-1372. A full list of the dimensional coordinate values, the acoustic measures and all relevant correlations are on file with Microfiche Publications, POB 3513, Grand Central Station, New York, NY 10017. Request Document NAPS-04554 and remit $11.35 for photocopy or $4.00 for fiche.

where b is the critical band value in bark and f is frequency in KHz (6). Pearson correlations were obtained between the vowel coordinates on each separate dimension and both bark-transformed and untransformed acoustic data; some acoustic measures used, such as F1-F0 and F3-F2 suggested by Syrdal and Gopal (5), were not considered in the original study. If bark-transformed measures are a better representation of the perceptual structure of vowels, then the bark-transformed measures should produce higher correlations, all other things being equal.

In general, the correlations between the vowel coordinate values and the untransformed acoustic data, on the one hand, and the bark-transformed data on the other hand, were very similar. In most cases the difference in r was generally less than 0.02. For example, the acoustic measures which produced the highest correlation for dimensions 1, 2, 3, and 4 were F2 offset (.927), F1-F0 offset (.829), F3-F2 onset (.884) and F0 (.467), respectively. The best single bark-transformed data for dimensions 1, 2, 3, and 4 were F2 offset (.917), F1-F0 offset (.831), F3-F2 onset (.827), and F0 (.467). All correlations were significant ($p < .001$), except dimension 4 ($p < .04$).

These results suggest that bark-transformation of acoustic data does not necessarily produce significantly better accounts of perceptually derived vowel features. In addition, the F1-F0 measure was a better representation of vowel height than F1, as suggested by the original study (2), and the proposed F3-F2 measure (for the front-back distinction) was not more salient for the primary front/back dimension (dimension 1) as suggested by Syrdal and Gopal (5), although it did best account for dimension 3 (originally labeled low-back onset).

REFERENCES

1. BLADON, A., HENTON, C., & PICKERING, J. (1984) Towards an auditory theory of speaker normalization. *Language Communication*, 4, 59-69.
2. FOX, R. A. (1983) Perceptual structure of monophthongs and diphthongs in English. *Language and Speech*, 26, 21-60.
3. RAKERD, B., & VERBRUGGE, R. R. (1985) Linguistic and acoustic correlates of the perceptual structure found in an individual differences scaling study of vowels. *Journal of the Acoustical Society of America*, 77, 296-301.
4. SINGH, S., & WOODS, G. (1970) Perceptual structure of 12 American vowels. *Journal of the Acoustical Society of America*, 5, 643-656.
5. SYRDAL, A., & GOPAL, H. (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.
6. 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.

Accepted November 16, 1987.