A THREE FEATURE SYSTEM FOR ENGLISH VOWELS

RICHARD C. BERRY

1. INTRODUCTION

In 1860, Fechner proposed that an analysis of confusions would provide a meaningful approach to the study of complex stimuli whose discriminative attributes are unknown. Peterson and Barney (1952) and Fairbanks and Grubb (1961) have analyzed the perceptual confusions among vowels. The intelligibility scores in these studies were high since the vowels were presented under ideal conditions. The few resulting confusions were between contiguous vowels in an $F1$-$F2$ plot. Miller (1956) analyzed the resulting confusions when English vowels and diphthongs were subjected to low-pass filtering. He found that among vowels with similar $F1$ frequencies, long vowels were confused with long vowels, short vowels with short vowels, but long and short vowels were not confused with each other. This study raised the fundamental question regarding the role of duration in the perception of phonological features and its relationship with formant frequency. The purpose of this study was to further investigate the relationship between duration and formant frequency.

2. METHOD

Four undergraduates served as speakers and twelve students selected from the same population served as listeners. The listeners were divided into two groups: Group I consisting of four subjects; and Group II of eight subjects. All subjects (speakers and listeners) demonstrated normal hearing, and were judged by a phonetician to be speakers of the same General American Dialect. The four speakers produced each of the nine English cardinal vowels (i, ɪ, ɛ, æ, a, ɔ, ʊ, u, ʌ) in an h-d context six consecutive times. Prior to producing the vowel, they listened to a model production on a tape loop. Each speaker selected his three better productions of each vowel. These speaker-preferred vowels were evaluated by a phonetician who selected the better two productions of each vowel for each speaker. A master tape was then constructed containing each phonetician-preferred vowel ten times in random order. This tape was played via speaker to Group I, and the one production of each vowel

for each speaker with the lowest error rate was selected. No selected vowel had an error rate greater than 20 %. An experimental tape containing each of these 36 listener-preferred vowels ten times in random order was then constructed. The tape was then copied utilizing four cascaded low-pass filters set at 670 Hz. The filter function of this system was determined to be 72 dB/octave. The 36 vowels were then copied once in random order in the same manner except they were high-pass filtered at 700 Hz. These high-pass vowels followed the low-pass vowels on the experimental tape. A five-second silent interval separated the vowels.

The eight listeners in Group II were seated (four at a time) in a sound-treated room at a distance of five feet from the speaker. The experimental tape containing the filtered vowels preceded by unfiltered directions was presented at approximately 75 dB, SPL.

3. RESULTS AND DISCUSSION

Table 1 is a confusion matrix for the 2880 low-pass vowels. Under the low-pass condition, it can be assumed that the second formants have been eliminated or greatly reduced (Peterson and Barney 1952). Under this condition, 44.5 % of the vowels were correctly identified. Miller (1956) concluded that under low-pass filtering vowel confusions cluster into four groups: [i-u], [ɪ-ʊ], [ɛ-ʌ], and [æ-a-ɔ]. The present results confirm Miller's findings. Figure 1 is a graphic representation of the data in Table 1.

TABLE 1

Confusion Matrix for 2880, 670-Cycle Low-Pass Filtered Vowels in an /h–d/ Context

	heed	who'd	hid	hood	head	hud	hawed	had	hod	Total
heed (i)	286	32	1	—	1	—	—	—	—	320
who'd (u)	211	92	12	2	1	—	1	1	—	320
hid (ɪ)	1	1	207	103	7	—	—	1	—	320
hood (ʊ)	—	1	118	163	30	2	1	1	4	320
head (ɛ)	3	1	18	31	237	22	1	7	—	320
hud (ʌ)	8	—	2	3	219	71	1	16	—	320
hawed (ɔ)	—	—	—	—	47	4	41	212	16	320
had (æ)	3	—	—	—	100	24	22	163	8	320
hod (ɑ)	—	—	—	1	48	4	23	223	21	320

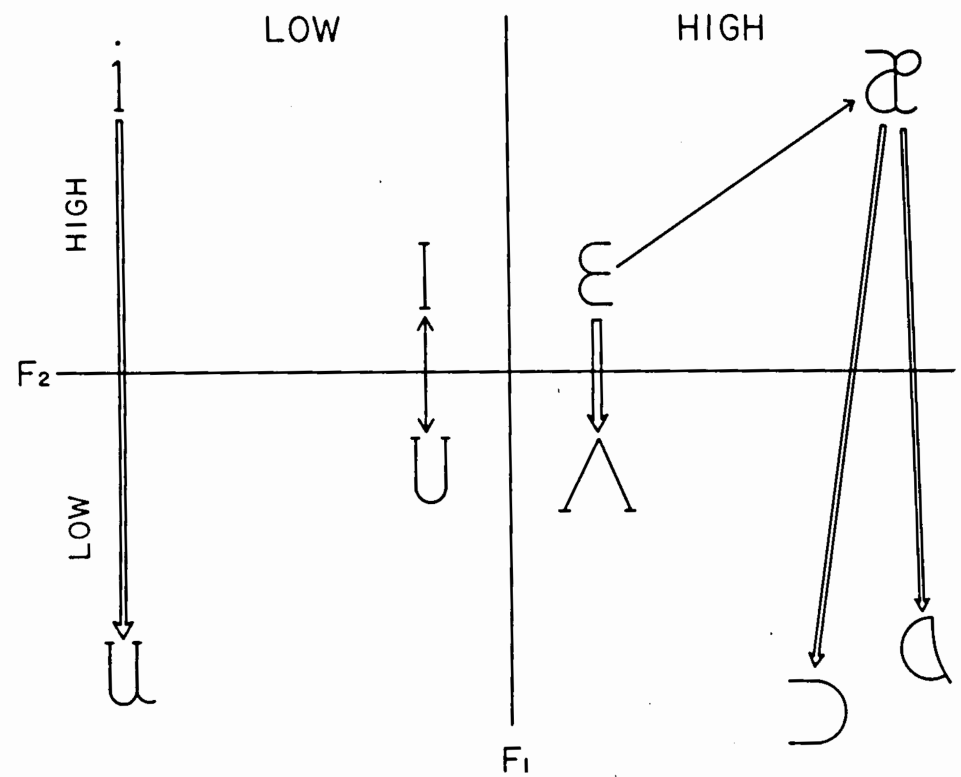

Fig. 1. A graphic representation of the confusions under the low-pass filtered condition.

When all the confusions are considered as a whole, as in Figure 1, they cluster into two major groups: (i, u, ɪ, ʊ), and (ε, ɔ, a, ʌ, æ). The confusions that fall outside these two groups represent only 7.4% of the total confusions. It is also interesting to note that these are confusions based on placement of the tongue; i.e., high-back vowels confused with high-front vowels, and low-back vowels with low-front vowels. Figure 2 is a graph of frequency of $F1$ for males, females, and children (Peterson and Barney 1952). It can be seen that the low-pass confusions fall into two groups based on frequency of $F1$: high $F1$ (ε, ɔ, a, ʌ, æ) and low $F1$ (i, u, ɪ, ʊ).

Table 2 is the resulting confusion matrix for the 288 high-pass vowels. Inspection of this table reveals that the confusions for the high-pass condition also cluster into two major groups: (i, ɪ, ε, æ) and (ʌ, a, ʊ, u, ɔ), with only 6.8 % falling outside these groups. Furthermore, only 49.3 % of the vowels were correctly identified. Figure 3 is a graphic representation of the data from Table 2. It can be seen from this figure that the confusions are almost exclusively height confusions; i.e., low-back vowels confused with high-back vowels and low-front vowels confused with high-front vowels. Figure 4 is an $F2$ plot based on Peterson and Barney (1952). It is interesting to note that a rather drastic shift in frequency of $F2$ between (æ) and (ʌ) divides the vowels into the above-mentioned groups.

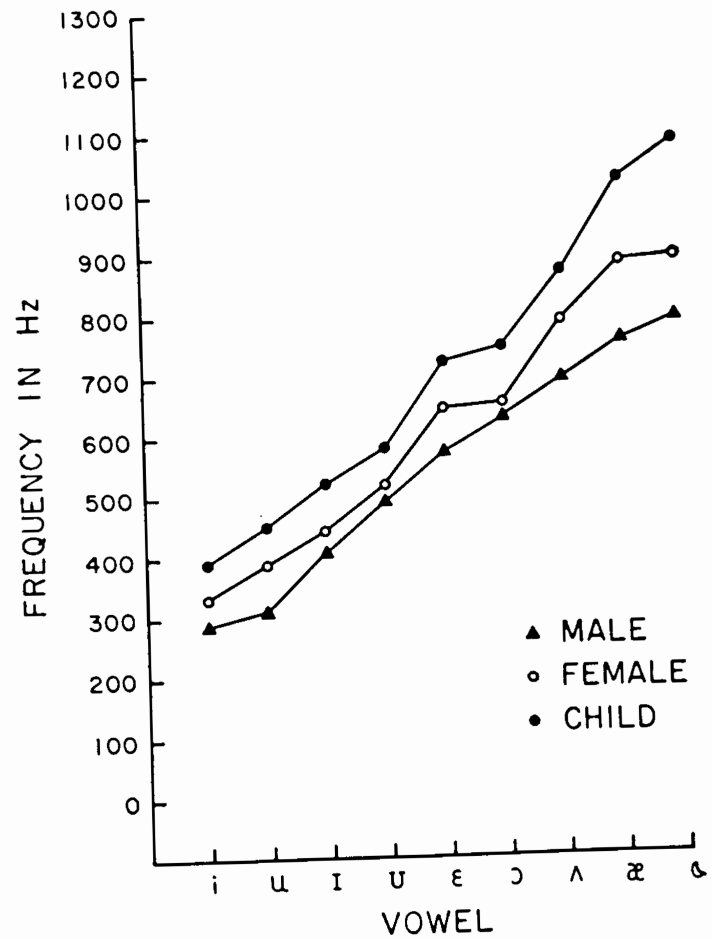

Fig. 2. Graph of frequency of $F1$ for males, females, children (from Peterson and Barney 1952).

Miller (1956) found only 6% of the confusions were durational, the present study found 17% for the low-pass condition and 26.7% for the high-pass condition. Table 3 summarizes the type of durational confusions obtained for each condition. It can be seen from this table that the type of durational confusion, short for long, or long for short, depends largely on which formant had been removed. When $F1$ is removed (high-pass) 100% of the durational confusions are long for short, while with $F2$ removed (low-pass), 88% are short for long. Thus, a strong relationship between formant frequency and perceived duration is evident. However, the present study can only describe this relationship, not explain it.

4. CONCLUSIONS

Identification of English vowels seems to be based on three binary distinctions: (1) frequency of $F1$ (high-low), (2) frequency of $F2$ (high-low), and (3) duration

TABLE 2

Confusion Matrix for 288, 700-Cycle High-Pass Filtered Vowels in an /d_h/ Context

	heed	hid	head	had	hud	hod	hood	who'd	hawed	Total
heed (i)	3	—	—	28	—	1	—	—	—	32
hid (i)	—	6	16	10	—	—	—	—	—	32
head (ɛ)	—	1	—	10	—	—	—	—	—	32
had (æ)	—	—	—	29	—	2	—	—	1	32
hud (ʌ)	—	—	2	5	20	1	2	—	2	32
hod (ɑ)	—	—	—	—	—	27	—	—	5	32
hood (u)	—	—	—	—	9	6	13	—	4	32
who'd (u)	—	—	—	—	—	15	—	3	14	32
hawed (ɔ)	—	—	—	—	—	13	—	—	19	32

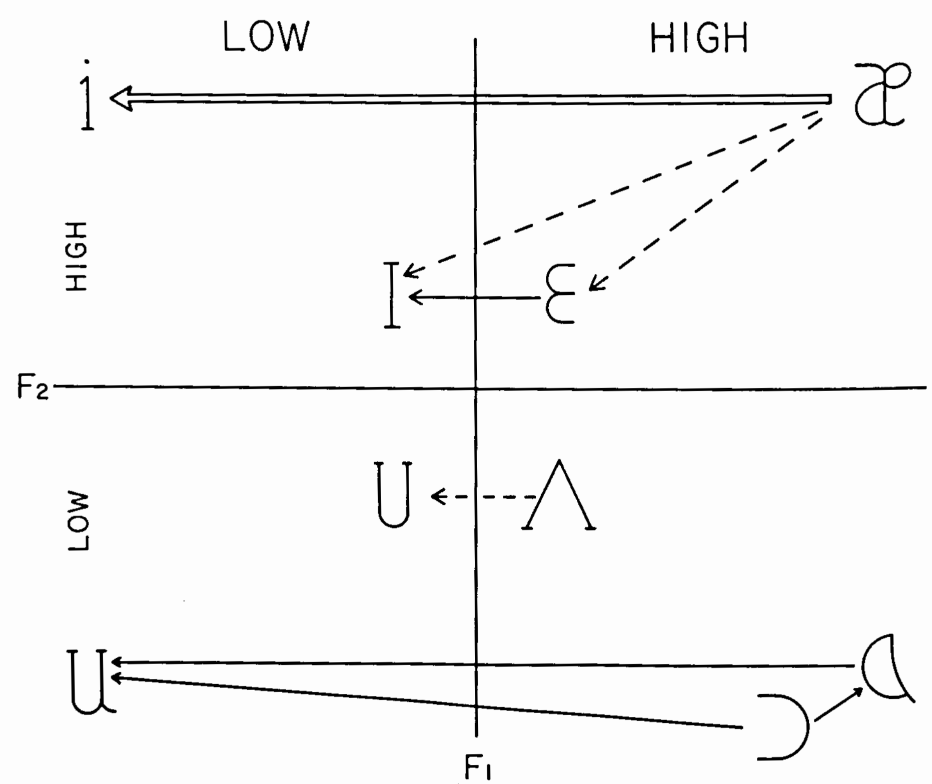

Fig. 3. A graphic representation of the confusions under the high-pass filtered condition.

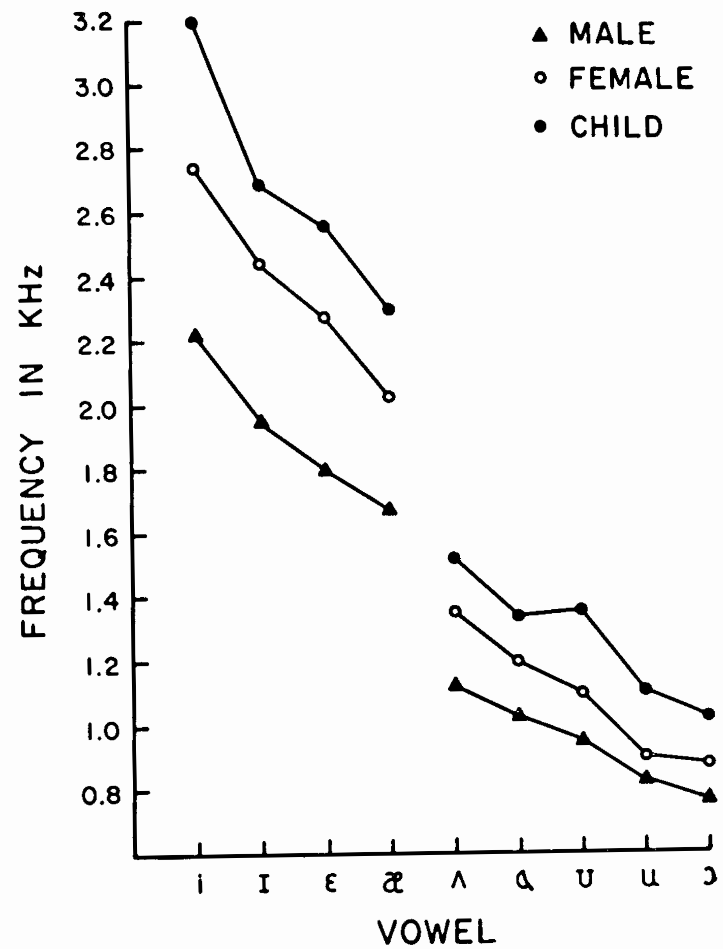

Fig. 4. Graph of frequency of *F*2 for males, females, and children (from Peterson and Barney 1952).

TABLE 3

Percentage of Confusions as a Function of Vowel Duration and Filtering Effects

Low Pass Filtering (670 Hz)	88%	12%
High Pass Filtering (700 Hz)	0%	100%
	Short for Long	Long for Short

(short-long). Perceived duration seems to be related to formant information in some way not yet specified. Finally, $F1$ carries information as to tongue height (high-low), while $F2$ carries information as to tongue placement (front-back).

Speech Department
Northeastern University
Boston, Massachusetts

REFERENCES

Fairbanks, C. and P. Grubb
1961 *Journal of Speech and Hearing Research* 4:203-219.
Miller, G.A.
1956 "The Reception of Speech", in *For Roman Jakobson* Morris Halle, ed. (The Hague, Mouton) pp. 353-360.
Peterson, G. and H. Barney
1952 "Control Methods Used in a Study of the Vowels", *Journal of the Acoustic Society of America* 24:175-184.

DISCUSSION

SINGH (Washington)
As I understand, your figures do not indicate the perceptual distances of these vowels. You do not have a measure to indicate a point (vowel) in space and its relation to all other points (vowels). I would suggest that you subject your matrices to a multi-dimensional analysis technique of Shepard-Kruskal or of Carroll and Chang. These analyses will probably give you better estimation of dimensionality, space, and also of the differences in condition (i.e., you can treat your three conditions, unfiltered, low-pass filtered, and high-pass filtered as three different subjects for the IND-SCAL analysis).

BERRY
I am in complete agreement with Prof. Singh that a multi-dimensional analysis would yield more information as to the perceptual distance between the vowels. I am quite sure, however, that such an analysis would yield the same three features. I feel Prof. Ladefoged's findings employing such an analysis based on formant frequency information, reported at the Acoustical Society of America, in Washington, D.C., April, 1971, supports this contention. He found three features; place, height, and lip rounding. Rounding was attributed primarily to $F3$. I am quite sure that if Mr. Ladefoged had utilized durational information a short-long distinction would have been evident. In the present study the effect of $F3$ per se was not evaluated. However, in the high-pass condition both $F2$ and $F3$ were present. It must be pointed out that for the nine vowels investigated, rounding is not distinctive. In regard to

perceptual distance, it seems evident that confusibility and similarity can be assumed to be synonymous. Thus, the more frequently two vowels are confused the more similar they are, i.e., the closer they are in perceptual space. Thus, based on $F1$ /i/ and /u/ are very close perceptually. However, based on $F2$ /i/ is much closer to /æ/. Of course, the degree of confusibility gives only a relative approximation of the perceptual distance between vowels. Therefore, I would greatly appreciate the opportunity to obtain a greater degree of accuracy by subjecting my confusion matrices to a multi-dimensional analysis. However, I do not have such a program at my disposal. If Prof. Singh could offer some assistance in obtaining such an analysis, I would be, indeed, grateful.