LANGUAGE AND SPEECH, Vol. 26, Part 1, 1983

PERCEPTUAL STRUCTURE OF MONOPHTHONGS AND DIPHTHONGS IN ENGLISH

ROBERT ALLEN FOX
The Ohio State University

This multidimensional scaling investigation seeks to determine the dimensions underlying the perception of diphthongs in American English and whether such dimensions are radically different from those found in studies utilizing monophthongal vowels. Dissimilarity data were obtained from subjects' paired comparison judgments of 15 English vowels [i ɩ eɩ ɛ æ ɑ ʌ ɔ oɷ ɷu ɔɩ aɩ aɷ ju]. INDSCAL analysis of the data revealed that the perceptual structure of the vowels was best modeled in terms of four perceptual dimensions labeled front/back, high/low, low-back onset, and mid/nonmid. Regression analyses demonstrated that the perceptual structure of the vowels was most significantly related to the first two formants of the stimulus vowels. No dimensions relating to dynamic acoustic information (such as direction and/or extent of formant change over time) were found. Separate regression analyses demonstrated that these empirically derived dimensions were significantly related to the distinctive features relating to tongue position and, in particular, that the high/low dimension was better predicted by Ladefoged's multivalued Height feature than by Chomsky and Halle's binary High, Low, and Tense.

INTRODUCTION

In phonetic research much effort has been directed at discovering the perceptual features or dimensions utilized in the perception of American English vowels and comparing them to various distinctive feature systems. Evidence on the nature of these perceptual features has been obtained primarily from multidimensional scaling experiments. The perceptual dimensions found in various studies, although differing slightly, have been remarkably stable in their interpretation and have, for the most part, corresponded to traditional linguistic (articulatory) distinctions. Singh and Woods (1970) found three dimensions which they labeled tongue advancement, tongue height, and retroflexion; Shepard (1972), analyzing Peterson and Barney's (1952) data, found front/back and height dimensions; while Fox (1978) found dimensions he labeled as front/back, high/low, and round/unround. Terbeek (1977), in a study using both English and non-English-like vowel stimuli, extracted dimensions he labeled round/nonround, high/low, retroflex/nonretroflex, back/nonback, and mid/nonmid. Similar results have been obtained in non-scaling studies as well; for example, Wickelgren (1965) analyzed errors in short-term recall of English vowels and found that obtained errors were best predicted in terms of front/back, and open (high/low) differences between the vowels.

However, a problem arises in generalizing these results to the perception of vowel quality that changes over a short course of time. The studies noted above ostensibly examined the perceptual structure of "monophthongs," that is, vowels whose formant (*F*) structure (and therefore phonetic quality) remained constant throughout. Each of the vowel stimuli used in the above experiments was seen to represent an American English monophthong and no attempt was made to include true diphthongs (e.g. [aɪ ɔɪ]) nor to discover what effect even moderate changes in vowel quality over time would have on a subject's perceptual judgments. Note that although Singh and Woods (1970) included two vowels (/e/ and /o/) that are at least potentially (and normally) diphthongal in English, they presented no relevant acoustic information, nor did they comment upon how much of a steady state they had in the productions used for the perceptual testing.

One experiment directed specifically at the question of monophthong vs. diphthong was carried out by Terbeek (1974; Terbeek and Fox, 1975). This was an unpublished MD-scaling experiment examining the perceptual structure of a set of nine diphthongal vowels which included three vowels with [i] off-glides, ([eɪ aɪ ɔɪ]), three vowels with [u] off-glides ([u oʊ aʊ]) and three vowels with level or schwa-like off-glides ([ɪə æə ɔə]). The nine vowels were produced in the phonetic context [# ___t] by a native American English speaker. Twenty-one subjects gave paired-comparison similarity judgments of all different possible vowel pairs using a nine-point rating scale. A different set of 21 listeners performed the same task on a second stimulus set composed of the same vowel phonemes, but recorded in isolation with no surrounding phonetic context (i.e., [#___#]). These two separate sets of perceptual data were then analyzed by INDSCAL (Carroll and Chang, 1970). Both stimulus conditions resulted in basically identical six-dimensional solutions, but represented a perceptual space quite different from what might have been expected given previous results in the literature. No clear-cut front/back dimensions emerged and although a dimension involving high/low distinctions was obtained ($F1$ onset), it was the last to emerge and represented the least salient of the extracted dimensions (i.e. accounted for the smallest amount of variance). The most salient perceptual property of the stimuli seemed to be the presence of rounding anywhere in the vowel vs. complete absence of rounding. Several of the dimensions defied a straightforward acoustic interpretation and were best labeled in terms of how the vowel sounds were commonly written in standard English orthography. The results obtained in this study seem very counterintuitive, especially in light of the fact that a very salient perceptual dimension seen in *all* other studies (i.e. front/back) did not appear at all. Such results would suggest that a set of features completely different from those used to explain monophthong perception is required for nonsteady-state vowels, but this conclusion is both unappealing and unlikely. More realistically, these results reflect some fortuitous skewing factor affecting subject responses and say little about the nature of diphthong perception.

The present study is designed, in part, to repeat this earlier experiment with an expanded stimulus set. If the difficult-to-interpret results of the above study are replicated, then we have a strong basis on which to claim that they have some psychological reality. On the other hand, if skewing factors affecting those previous results have been eliminated, then more interpretable results, compatible with those from monophthongal studies, may be obtained.

A separate problem that emerges in considering the perception of vowels is that of understanding the interaction of vowel duration and perceptual features. It is remarkable that duration has not appeared as a perceptual dimension in any of the MD-scaling experiments given that many studies (e.g., Tiffany, 1953; Ainsworth, 1972; Mermelstein, 1977; Verbrugge and Isenberg, 1978) have shown that durational differences can affect identification of vowel quality – particularly for tense–lax vowel pairs. However, it is not clear whether the lack of a duration dimension is a function of the perceptual structure being modeled or the nature of the stimulus. For example, some studies (e.g., Pols *et al.*, 1969) equalized durations over all vowels eliminating whatever perceptual contribution duration might have made, while other studies (e.g., Singh and Woods, 1970) simply did not report durations and one cannot determine whether such durational differences were actually present. Given that true diphthongs may exhibit durations quite different from monophthongs, it is important to ascertain whether this acoustic cue plays an important role with regard to the perceptual features used.

Aims of the study

This study seeks to discover the perceptual structure of a large set of American English vowels that includes both monophthongs and diphthongs. The perceptual distance data were obtained using dyadic comparisons and were analyzed using INDSCAL. The results obtained here should indicate the most salient features of vowel perception for that set of subjects which should be generalizable to speakers (at least to speakers of that dialect) of American English as a whole. The perceptual features extracted during INDSCAL analysis will then be compared to various sets of distinctive features as well as to physical (acoustic) aspects of the stimulus tokens – these comparisons were done utilizing multiple regression techniques.

METHODS AND PROCEDURE

Stimuli

The experimental stimuli were composed of 15 American English vowels [i ι eι ε æ ɑ ʌ oɷɔ ɷ u ɔι aι aɷ ju] that are representative of the nonrhotic vowels in the speech of a typical Midwestern speaker (Ladefoged, 1975). This set of vowels includes both relatively monophthongal vowels (e.g., [i u]) as well as true diphthongs (e.g., [aι ɔι]). The vowels were produced by a male speaker in the phonetic environment [h___d]. The stimuli were recorded on an Ampex tape recorder using an Ampex microphone, while the speaker sat in an IAC sound chamber with the microphone 12 inches from his mouth. The tokens appeared in pairs, 300 msec apart on the stimulus tape. The interval between separate stimulus pairs was 5 sec.

The syllabic nucleus [ju] included in the stimulus set is here considered to be a diphthong (as classified by Ladefoged, 1975). However, despite its diphthong-like nature (Holbrook and Fairbanks, 1962), it could be argued that [ju] should not be considered a diphthong, but rather a glide + vowel combination (Jones, 1919; Ilse Lehiste, personal communication). Clearly one can offer arguments supporting either side of this issue, but the more important question is an empirical one: do listeners perceive [ju] as a diphthong or as a glide + vowel combination? Inclusion of [ju] in the stimulus set can be defended on the grounds that the resulting perceptual data will yield evidence pertinent to this question. Given the nature of the task (that subjects are to make similarity judgments on the basis of the vowel alone), if [ju] is perceived as a diphthong, then it should pattern in the perceptual space in a manner distinctive from the vowel [u], just as we might expect [ɔɩ] to be distinct from [ɔ] or [ɩ]. If the stimulus is considered to be a glide + vowel combination by subjects (and thus they can respond to the vowel alone), [ju] and [u] should show almost identical coordinate values on all extracted perceptual dimensions. To anticipate the results to be discussed, it will be argued that listeners are not perceiving [ju] as significantly distinct from [u].

With 15 different tokens, there were 105 different vowel pairs possible, disregarding order within a pair and not allowing stimulus tokens to be paired with themselves (diagonal cells in the perceptual distance matrix input to INDSCAL are presumed to contain zero). Each different stimulus pair appeared once in four different trials; several separate similarity judgments were obtained from each subject in order to establish perceptual distance estimates. Presentation order of stimulus pairs was pseudorandomized (i.e., random order except for the requirement that no vowel appear twice in consecutive pairs) across each of these different trials.

In real speech, the vowels used in this study normally appear both with different degrees of vowel quality change over time (true diphthongs exhibiting the greatest amount of change) and differences in intrinsic duration when embedded in the same phonetic context (Peterson and Lehiste, 1960). Since the aim of this study was to determine the features used in perceiving vowels in speech without presupposing the nature or identity of these features, an effort was made to allow the vowel tokens to resemble those that normally appear in fluent speech. Therefore vowel qualities were, for the most part, not steady-state (even for the relatively "monophthongal" tokens) and no effort was made to equalize durations over all vowels (e.g., [i] was consistently longer than [ɩ]).

The steady-state vowel stimuli used in scaling experiments are usually described in terms of mean $F1$ and $F2$; however, since we have used vowels that change in formant structure dynamically over time, it is necessary to capture the acoustic characteristics of such tokens in a more robust manner. To this end each vowel was divided into four equal sections, as shown in Figure 1, yielding five separate reference points. Point I represents the onset of $F1$ and point V represents the onset of the stop closure. Points II, III, and IV are equidistant between points I and V. Acoustic measurements made for each stimulus vowel included frequencies of $F1$, $F2$, $F2–F1$, and $F3$ at points I-IV; change (Δ) in $F1$, $F2$, and $F2–F1$ from point I to point IV (both as signed and unsigned integers), fundamental frequency (F_0), and duration. Acoustic measurements were not made at point V since the formant values here reflect the transition to the following stop. Figure 2 is a graphic display of the stimulus vowels in terms of $F1$ and $F2$ at points I-IV (a complete description of the mean acoustic measurements appears in Appendix I). The mean duration of the vowels can be seen in Figure 3.

Subjects

Twenty-four subjects participated in the experiment. All were students at The Ohio

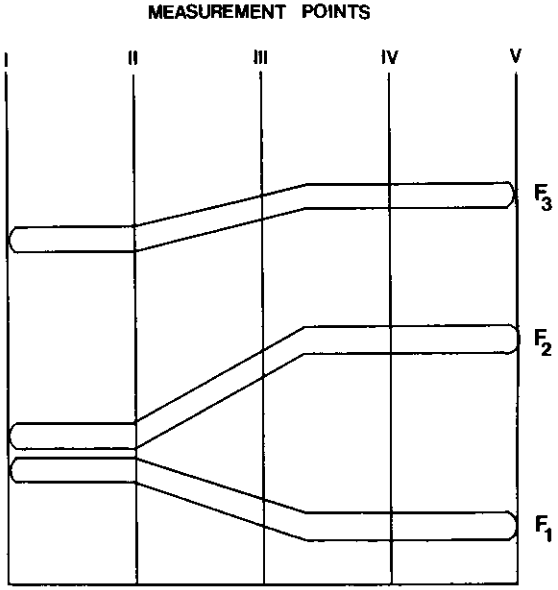

Fig. 1. Schematic of measurements obtained from each stimulus vowel.

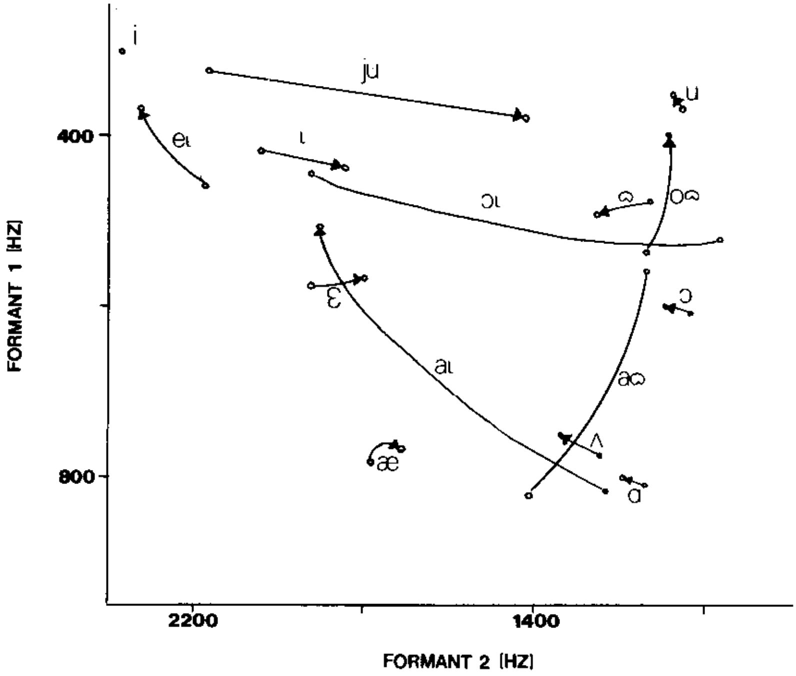

Fig. 2. Mean formant 1 and formant 2 values for the stimulus vowels.

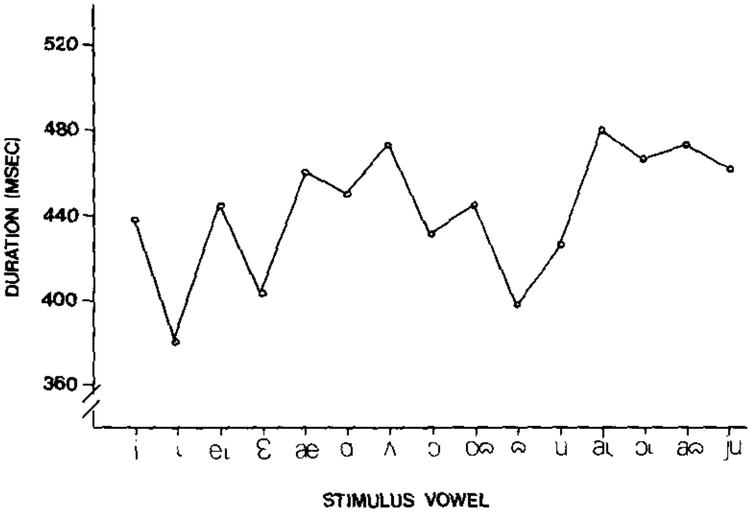

Fig. 3. Mean duration of the stimulus vowels.

State University and were (unpaid) volunteers. All subjects were from central Ohio and spoke a Midwestern dialect of American English. No subject had any known hearing deficiency and all were completely unfamiliar with the experimental technique used and the goals of the study.

Experimental procedure

This study used a scaling task that required subjects to listen to two sequentially presented stimulus tokens (a stimulus pair) and to judge their similarity/dissimilarity on a nine-point scale. Subjects were told that they were going to hear pairs of words on an audio tape and that the words differed from one another only in terms of the medial vowel. Following this description, subjects heard a recorded list of all 15 different stimulus tokens. Subjects next read typed instructions describing the experimental task wherein they were asked to indicate their judgments of similarity between members of any token pair by circling the appropriate point on the rating scale. Subjects were reminded to pay attention only to the vowel difference between the two tokens. Subjects then practiced on an introductory block of 10 stimulus pairs. After assurance that no subject had questions concerning the task (and that subjects were indicating their responses in the correct fashion), the subjects were presented with eight blocks of stimuli, each block consisting of 60 pairs, yielding a total of 480 paired-comparison judgments (four judgments for each different stimulus pair plus buffer pairs at the start and finish of each block). The stimulus pairs for a particular pseudo-random order were spread across two different blocks. The stimulus pairs were presented in one pseudo-random order in Blocks 1 and 2 and 5 and 6, and in a second pseudo-random order in Blocks

3 and 4 and 7 and 8. Vowel stimuli were ordered pairwise in the sequence A B in Blocks 1 and 2 and 3 and 4, and in the sequence B A in Blocks 5 and 6 and 7 and 8. A brief period of rest was allowed after the presentation of four blocks of stimuli. The entire experiment lasted approximately 50 minutes.

Each subject's similarity judgments were checked for consistency before multidimensional scaling analysis was done. This was to ensure that excess "noise" was not included in the perceptual distance information submitted to INDSCAL analysis. As Terbeek (1977) cautioned, excessive noise in the data may cause a solution of too many dimensions to appear to be more formally acceptable than a solution of fewer dimensions. The consistency check involved comparison of a subject's responses on one pseudorandom order of stimulus pairs with his responses on the other order using simple correlation statistics. A subject's data were not used if Pearson's r was not above 0.50, ensuring a significance level of at least $p < 0.001$. On the basis of these consistency statistics, 19 subjects were retained for INDSCAL analysis. The data for each subject were collapsed over all blocks and symmetricized into a 15 × 15 dissimilarity matrix (the raw similarity matrices appear in Appendix II). These symmetrical matrices served as input to the INDSCAL analysis program.

RESULTS

Analysis

The primary analytic tool utilized in this study was multidimensional (MD) scaling, a method used to discover the structure underlying a set of perceptual responses. This procedure involved viewing dissimilarity judgments of every pair of a set of objects (vowels in this case) as indicative of the perceptual distance between those objects. MD-scaling algorithms attempt to account for these perceptual distances by modeling the underlying perceptual space; that is, the stimuli are placed in an n-dimensional space such that the distance between the objects in this space corresponds optimally (in a least-squares sense) to the experimentally obtained perceptual distance estimates between the objects. Following proper rotation of the coordinate axes, these dimensions are seen to represent the perceptual features underlying subjects' similarity judgments and, by extension, the vowel features normally used in speech perception. No assumption of the number or the nature of these dimensions is made prior to MD-scaling analysis. (For a more detailed discussion of multidimensional scaling techniques see Carroll and Wish, 1974; Wish and Carroll, 1974; or Kruskal and Wish, 1978.)

INDSCAL, the MD-scaling procedure used in this study is a particularly powerful scaling program. Not only does INDSCAL develop an n-dimensional solution and assign coordinate values for each stimulus on each resulting dimension, it considers perceptual differences across individual subjects and determines the relative salience or weight of each dimension for each subject. One of the advantages that accrue from consideration of individual differences is that the orientation of the axes of the resulting solution is fixed and cannot be rotated without worsening the overall fit of the solution to the perceptual data. Since the orientation of the axes is fixed, the claim that the dimensions have a psychological reality is thereby strengthened – although it does constrain the researcher to interpret the solution as it appears rather than rotate it into a more interpretable configuration. However, INDSCAL does make certain, strong assumptions about how one should model individual perceptual differences. In particular, individuals are assumed to share the same perceptual dimensions (at least native speakers of a particular language – see Terbeek (1977) for discussion) and are allowed only to differ in terms of the weight, or salience, that they attach to each dimension. In addition, input perceptual distances are analyzed only in terms of an underlying Euclidean space.

It is not certain, however, that violations of these assumptions render INDSCAL solutions questionable. In particular, if there are two or more individuals (or groups of individuals) who do not share one or more perceptual dimensions (as is possible in Tucker and Messick's (1963) concept of "points of view" which assumes that different clusters of subjects can have their own private space with no necessary relation between the spaces for different clusters), the INDSCAL solution should include all different dimensions found in any of these "private spaces" into a composite "group space." Subjects not sharing a dimension in this group space will have a zero weight for that dimension (cf. Carroll and Wish, 1974). In terms of the exclusive use of the Euclidean metric (primarily for reasons of ease of computation, conceptual simplicity, and historical precedence) applications of the algorithm are very robust and usually recover the underlying structure (even if non-Euclidean) quite adequately (cf. Carroll and Wish (1974) and Terbeek and Harshman (1972) for a discussion of these issues). In addition, although the claim of psychological reality for INDSCAL dimensions and validity of its view of individual differences is a theoretical one, numerous applications of INDSCAL support these claims. Studies in several areas of human perception have shown that INDSCAL results have corresponded to previously established perceptual processes (cf., Wish and Carroll, 1974).

INDSCAL analysis was done using data from the entire 15-vowel stimulus sets as well as the 9- and 11-vowel subsets. This analytic design allows one to compare the results obtained here with studies in the literature using similar but smaller stimuli sets (e.g., Fox, 1978; Singh and Woods, 1970) and to see whether, or to what degree, the perceptual space composed of relatively monophthongal vowels is changed by the inclusion of vowels that are increasingly diphthongal in character. INDSCAL analysis of the 15-vowel data was done in 1–8 dimensions. Five different analyses were done at each dimensionality and each separate analysis had a different random starting configuration. The four-dimensional solution was found to provide the best overall representation of the perceptual structure of the stimuli.

Since one of the most crucial decisions in any INDSCAL study involves a judgment as to what the "correct" dimensionality of the solution is, and because several other studies (e.g., Terbeek, 1974, 1977; Terbeek and Fox, 1975; Fox, 1974) have extracted a larger number of dimensions for an equal or smaller number of vowels, some defense of this dimensionality is in order. The "fit" curve shown in Figure 4 (indicating the cumulative proportion of variance accounted for at different dimensionalities – a traditional guide in such a decision) gives relatively little relevant information. One normally looks for

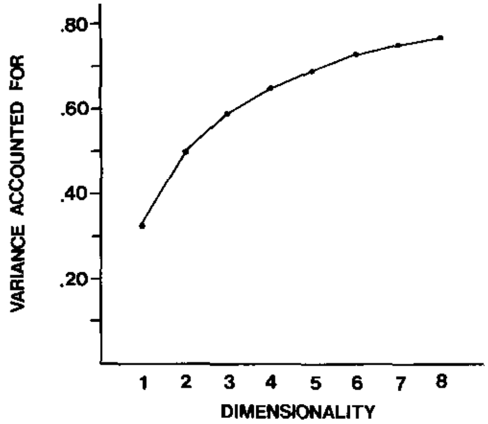

Fig. 4. Fit curve for 15-vowel INDSCAL analysis.

an "elbow" in such a curve indicating a steep decrease in variance accounted for from one dimension to the next, the rationale being that once the significant (and real) dimensions are extracted, dimensions accounting for noise will then be extracted which will account for similar proportions of the variance, and the curve should straighten (i.e., the decrease in variance accounted for will be relatively constant). However, several other criteria were more useful. Interpretability of the solution required that no more than four dimensions be extracted (and perhaps as few as three, since, as we shall see, the fourth dimension was relatively uninterpretable), since at dimensionality five and above the clear perceptual structure indicated at lower dimensionalities degenerated. In particular, the higher dimensionalities obscured the very suggestive front/back and high/low dimensions. Four dimensions also represent the upper limit on the stimuli/dimensions ratio recommended by Kruskal and Wish (1978) for 15 stimuli.

Another method used to ascertain whether or not the four-dimensional solution was appropriate was a version of Gandour and Harshman's (1978) "split-half" procedure. The perceptual data were divided into two separate halves (one containing nine subjects, the other containing 10 subjects) and INDSCAL analysis done on each separate half at a number of dimensionalities, again obtaining several different solutions at each dimensionality using different random starting configurations. Following this procedure, the dimensions obtained from the analysis of each separate half of the perceptual data were compared using correlation statistics. The reasoning behind this method is that those perceptual dimensions common to both halves of the data will include the significant, psychologically real dimensions common to the subject group as a whole. There is much less likelihood that identical error dimensions will replicate across the two separate halves. Perceptual dimensions corresponding to the four obtained in the original INDSCAL analysis replicated in both separate analyses, while the fifth dimensions (and higher) extracted in the separate analyses were different. This supports the selection of the four-dimensional solution.

The final criterion used in making the decision was what has been termed the uniqueness-of-solution property (Terbeek, 1977). INDSCAL theoretically provides unique spatial solutions up to and including the correct number of dimensions and multiple solutions at dimensionalities above this point (each separate analysis starting from a different initial configuration). Since this is so, if one discovers unique solutions at 1 to $n-1$ dimensions and multiple solutions at n dimensions, then one is obliged to view $n-1$ as the correct dimensionality. For these data what were considered to be multiple solutions were found at the five-dimension level. These multiple solutions were not simply a case of finding two or more local minima (as opposed to the global minimum), because each different solution accounted for the same proportion of the variance (within 0.001%). Practically speaking, however, there are potential problems with this particular selection criterion. In obtaining INDSCAL solutions at any given dimensionality from different starting configurations some degree of perturbation in the coordinate values from one solution to the next inevitably occurs. But despite the fact that great importance may be placed upon this single criterion in selecting the proper dimensionality (cf. Terbeek, 1977), there is no generally acceptable method of empirically determining whether two INDSCAL solutions are different enough to be considered multiple solutions rather than insignificantly different versions of a single "unique" solution. Thus although this criterion also points to the four-dimensional solution as the one to interpret, we would hesitate to place heavy reliance upon it if other evidence did not also point to the same conclusion.

All factors considered, we feel that four dimension represents the upper dimensionality limit for the perceptual structure of the 15 stimuli used. However, if interpretability of *each* particular perceptual dimension were the only factor considered, the three-dimensional solution would have been selected (since the fourth dimension extracted received the least meaningful phonetic interpretation). However, the three-dimensional solution is virtually identical to the first three dimensions extracted in the four-dimensional solution.

The 15-vowel INDSCAL solution is presented in Figure 5 (shown are the D1 × D2 and D3 × D4 planes, which account for approximately 65% of the variance in the scaling data, corresponding to a correlation of 0.79 between the perceptual distance data and the INDSCAL spatial model). The most salient dimensions, D1 and D2, account for 32% and 18% of the variance, respectively, while the two remaining dimensions, D3 and D4, account for 8% and 6% of the variance, respectively. The complete INDSCAL solution, including coordinate values and subject weights, appears in Appendix III.

In traditional linguistic terminology one can easily label D1 as front/back and D2 as high/low. The locations of the relatively steady-state vowels in the D1 × D2 plane closely resemble their traditional placement in the vowel quadrilateral, except for the position of [eɩ] relative to [ɩ]. D3 was labeled low-back onset and contrasts vowels having low and back onglides (e.g., [aɩ, ɔɩ, a]) with vowels having either high or front onglides (e.g., [ɩ, ɛ, u, a]). D4 does not readily yield to any standard linguistic interpretation.

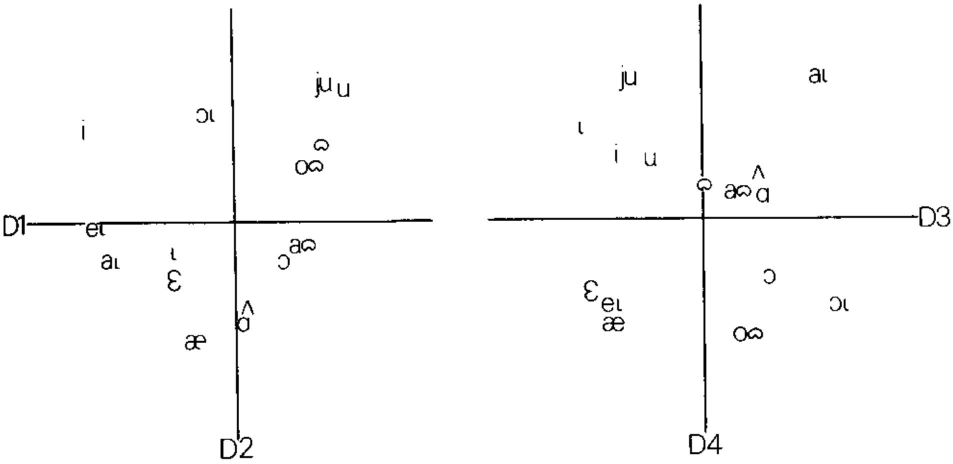

Fig. 5. Four-dimensional INDSCAL solution for 15 stimulus vowels.

The diphthongs [aɪ, ɔɪ, aʊ] are located in the perceptual space in positions distinct from those vowels comprising their respective endpoints. However, this is *not* the case with [ju] which has coordinate values almost indentical with [u] in the first three dimensions. Only in the fourth and least salient dimension does some significant differentiation between these two stimuli appear. The scaling response difference between [aɪ, ɔɪ, aʊ] on the one hand and [ju] on the other would thus support the argument that [ju] is not perceptually processed as a true diphthong, but rather a glide + vowel sequence.

Acoustic interpretation of the solution

While labels and descriptions for each dimension are of interest, a more crucial concern is the degree to which these perceptual dimensions reflect the acoustic characteristics of the speech stimuli themselves, especially since one of the important aims in speech research is the determination of the acoustic cues used in the perception of speech sounds. In terms of this study, what can be said about the relationship between the perceptual and acoustic structure of the stimuli? In particular, can one determine which physical aspects of the stimuli play important roles in the perceptual judgments of the listeners?

A gross indication of this relationship can be determined simply by correlating the vowel coordinates on each perceptual dimension against the various acoustic measures, described earlier (and presented in Appendix I). These correlations will indicate the acoustic property that best matches each perceptual dimension. A representative sample of these correlations (Pearson's r) is shown in Table 1. As can be seen, D1 the front/back dimension, was most significantly related to $F2$IV, accounting for almost 88% of the

TABLE 1

Correlation between perceptual dimensions and selected acoustic measures

Measure	D1	D2	D3	D4
$F1$ I	–0.011	–0.796*	0.519	–0.161
$F1$ II	–0.011	–0.785*	0.529	–0.112
$F1$ III	–0.041	–0.811*	0.497	–0.128
$F1$ IV	–0.110	–0.833*	0.262	–0.225
$F2$ I	0.575	0.017	–0.812*	0.166
$F2$ II	0.672	–0.050	–0.798*	0.100
$F2$ III	0.858*	–0.048	–0.632	0.005
$F2$ IV	0.935*	0.003	–0.331	0.002
$F2–F1$ I	0.477	0.262	–0.831*	0.187
$F2–F1$ II	0.561	0.213	–0.827*	0.124
$F2–F1$ III	0.728	0.225	–0.689	0.045
$F2–F1$ IV	0.843*	0.245	–0.363	0.068
Delta $F1$ u	0.103	–0.104	0.472	0.206
Delta $F1$ s	–0.142	0.196	–0.543	–0.039
Delta $F2$ u	0.092	0.323	0.405	0.264
Delta $F2$ s	0.362	–0.019	0.634	–0.205
Duration	0.078	–0.155	0.461	0.038
F_0	0.018	0.059	0.194	0.521

*$p < 0.001$

variance on that dimension. It is interesting to note that this correlation is higher than any of those between D1 and the $F2–F1$ measures suggested by Ladefoged (1975) as being a better reflection of the front/back distinction. D2, the height dimension, was inversely related to $F1$, with $F1$IV yielding the highest correlation. D3, the low-back onset dimension, was best matched with $F2–F1$ (and $F2$) measures and, in addition, showed suggestive correlations with $F1$I and $F1$II measures – this was not unexpected given the low-back onset label. D4, the least salient dimension and the last one extracted, was not significantly correlated with any acoustic measure at the 0.01 level or better, although the correlation between D4 and F_0 was marginally significant (at the 0.05 level). Given the number of correlations done on the data, this does not necessarily indicate an important relationship.

However, it is not necessarily the case that perceptual features reflect only a single acoustic variable. In a manner analogous to that argued for distinctive features (Parker, 1977), these dimensions may be best interpreted as representing the integration of several acoustic cues in a many-to-one manner. If this is the case, one needs to compare these dimensions to *sets* of acoustic measures. The method used to ascertain these relationships was stepwise multiple linear regression (Edwards, 1979).

Multiple regression as a general technique allows one to analyze the relationship between a dependent variable and a set of independent variables. In the procedure to be described here, each separate perceptual dimension (or rather the vowel coordinates on each of these dimensions) was entered into a series of regressions as the dependent variable with various sets of acoustic measures as the independent variables. The acoustic measures were included into these series of regressions in a stepwise manner; that is, each successive regression included that acoustic variable that brought about the greatest improvement in the prediction of the location of the vowels along a particular perceptual dimension. The acoustic measure that explains the greatest amount of the variance in the dimensional coordinates will be included in the regression equation at the first step; the acoustic variable that, in conjunction with the first variable, will explain the greatest amount of variance will be included at the second step; and so on. Following each step a multiple correlation coefficient (r_m) is calculated which is an estimate of the relationship between a weighted linear combination of the variables included in the equation up to that point and the perceptual dimension. The goodness of the prediction at each step is indicated by the variance-accounted-for ($r_m{}^2$) which reflects the proportion of the variance explained by that set of acoustic variables. An F-test will indicate whether the prediction afforded by these acoustic variables is a significant one.

However, we must be concerned with more than whether or not r_m is significant after the addition of each variable into the regression equation. If one includes a set of random numbers as an independent variable into the regression equation, the resulting r_m may still be very significant, and the amount of variance accounted for may increase slightly, but the *proportionate increase* in the variance accounted for by the addition of such a variable may not be significant. To this end, the proportionate increase in variance accounted for by each variable was noted and the significance of each variable's incremental contribution was calculated (F-test for a partial regression coefficient). In this way one can determine whether the addition of an acoustic variable contributes significantly to the overall prediction of the perceptual dimension.

Before regressions could be done, a decision had to be made concerning which acoustic variables to use in the analyses. Given the large number of acoustic variables (22), it is impossible to include them all into the regression procedure simultaneously. First, one is constrained by the degrees of freedom; the largest number of variables that could possibly be included in these regression equations which would yield even one degree of freedom would be 13. In addition, multicollinearity exists among the independent variables (i.e., many of the acoustic variables are highly correlated with one another), which dictates that they not be included in the same regression equation.

Accordingly, the regression analyses were done using various subsets of the acoustic

TABLE 2

Regression analysis of perceptual dimensions against acoustic measures

Variable name	Beta weight	Multiple correlation	Variance accounted-for	Proportionate increase in variance accounted for	F-test for individual variables
Dimension 1					
$F2$ IV	1.029	0.968	0.932	0.875	$F_{(1,12)} = 164.2, p < 0.001$
$F1$ I	0.303	$F_{(2,12)} = 82.12, p < 0.001$		0.057	$F_{(1,12)} = 10.1, p < 0.05$
Dimension 2					
$F1$ III	1.138			0.658	$F_{(1,11)} = 93.2, p < 0.001$
$F2$ II	0.456	0.947	0.898	0.164	$F_{(1,11)} = 18.9, p < 0.001$
Duration	−0.312	$F_{(3,11)} = 23.1, p < 0.001$		0.076	$F_{(1,11)} = 8.17, p < 0.01$
Dimension 3					
$F2$–$F1$ III	−0.683	0.891		0.684	$F_{(1,12)} = 22.8, p < 0.001$
$\Delta F2$s	0.361	$F_{(2,12)} = 23.1, p < 0.001$	0.794	0.110	$F_{(1,12)} = 6.4, p < 0.05$
Dimension 4					
F_0	0.521	0.521 $F_{(1,13)} = 4.85, p < 0.05$	0.272	0.272	$F_{(1,13)} = 4.85, p < 0.05$

variables. Examples of such subsets included: (1) $F1$I, $F2$I, $F3$II, F_0, duration, Δ$F1$s, Δ$F2$s; (2) $F1$II, $F3$II, $F2$–$F1$II, F_0, duration, Δ$F1$s, Δ$F2$s; etc. Some limitations on set membership had to be followed both in terms of the number of measures and the desire to avoid multicollinearity. To this end, each different subset had no more than a single representation of each different acoustic measure (e.g., $F1$I and $F1$II were never included in the same set).

The most relevant regression results, which appear in Table 2, for the most part represent only the best predictors of each perceptual dimension. The most significant predictors of D1 (front/back) were $F2$IV and $F1$I which together account for 93.2% of the variance on that dimension. However, although significant (at the 0.05 level), the contribution of $F1$I to the explained variance is relatively small (5.7%). Such a result suggests that the first perceptual dimension reflects mainly front/back distinctions among the stimuli, but that some vowel height information is also involved in specifying the location of a particular vowel along this particular dimension.

In terms of D2 (high/low), although the best single predictor was $F1$IV, the best set of acoustic predictors included $F1$III, $F2$II, and duration which together accounted for 65.8%, 16.4%, and 7.6% of the variance on that dimension, respectively. These results indicate that D2, for the most part, reflects mainly $F1$ information, but that inclusion of $F2$ information also explains a relatively large amount of the variance. Duration contributes a small, but significant amount. Thus we can see that the two most salient dimensions are best interpreted in terms of the first two formants, with some significant influence due to duration present.

The third dimension, D3 (low-back onset), was most highly correlated with $F2$–$F1$ ($r = -0.831$) and $F2$–$F1$II ($r = -0.827$), but regression of D3 against various acoustic measures revealed that signed change in $F2$ (Δ$F2$s) from points I to IV made a significant increase in the prediction of placement of vowels along that dimension. The combination of $F2$–$F1$II and Δ$F2$s accounted for 68.4% and 11% of the variance, respectively. This is the first indications that the subject's perceptual responses are sensitive to diphthong movement, although the significance level of Δ$F2$s is only at the 0.05 level.

The fourth dimension, D4, as yet unlabeled, was best predicted by F_0. No other acoustic variable could significantly increase the prediction of the vowels along D4.

Distinctive feature interpretation of the dimensions

Thus far in our discussion of perceptual dimensions we have assumed that they represent phonologically relevant aspects of speech segments extracted during the perceptual process which may reflect more than a single aspect of the acoustic signal. It is clear that so defined they have much in common with the distinctive features utilized in theoretical phonology (particularly generative phonology). It is therefore necessary to determine to what extent one could claim that these empirically derived perceptual and the more abstract, theoretical distinctive features are identical.

Although several different complete distinctive feature systems have been developed (Jakobson, Fant and Halle, 1951; Chomsky and Halle, 1968; Ladefoged, 1971, 1975) and several smaller feature subsets have been suggested to account for certain sets of experimentally derived data (Wickelgren, 1965; Singh and Woods, 1970), the most useful

TABLE 3

Feature values for the vowel stimuli

	i	ɪ	eɪ	ɛ	æ	ɑ	ʌ	ɔ	oʊ	ʊ	u
High	1	1	0	0	0	0	0	0	0	1	1
Low	0	0	0	0	1	1	0	0	0	0	0
Tense	1	0	1	0	0	0	0	0	1	0	1
Round	0	0	0	0	0	0	0	1	1	1	1
Back	0	0	0	0	0	1	1	1	1	1	1
LHigh	4	3	3	2	1	1	1	2	3	3	4

comparisons are between the perceptual features and those distinctive features that have been shown to have the greatest utility and power in explaining phonological data. With these concerns in mind, the perceptual dimensions D1 through D4 were compared to the binary features proposed in Chomsky and Halle (1968): High, Low, Tense, Back and Round; and Ladefoged's four-level Height feature (LHigh). These comparisons were done using stepwise multiple linear regression in a manner similar to that described in the previous section. Each perceptual dimension was regressed against several sets of features. However, since there is no universal agreement on the distinctive feature values for the diphthongs [aɪ, ɔɪ, aʊ], and since we are considering [ju] as a variant of [u], these comparisons were done using a subset of the vowel stimuli, including the vowels [i, ɪ, eɪ, ɛ, æ, ɑ, ʌ, ɔ, oʊ, ʊ, u]. The distinctive feature values assigned for each of these 11 vowels are shown in Table 3. The binary Chomsky/Halle features are marked 1 or 0 (for + or –) while Ladefoged's multivalues Height feature (LHigh) has values ranging from 1 (lowest) to 4 (highest). The results of the regression analyses are summarized in Table 4.

The first perceptual dimension was significantly related to the feature Back, which accounts for over 86% of the variance on that dimension. The inclusion of no other feature in the regression equation produced a significant increase in the amount of variance explained (although the addition of Round into the regression equation produced a nearly significant improvement of 8.3%, $0.05 < p < 0.10$). The positions of [aɪ, ɔɪ, aʊ] along D1 would suggest that they be marked –Back, –Back, and +Back, respectively.

TABLE 4

Regression analysis of perceptual dimensions against distinctive features

Feature	Beta weight	Multiple correlation	Variance accounted-for	Proportionate increase in variance accounted for	*F*-test for individual variables
Dimension 1					
Back	–0.866	0.866 $F_{(1,9)} = 26.98, p < 0.001$	0.750	0.750	$F_{(1,9)} = 26.98, p < 0.001$
Dimension 2					
LHigh	–0.962	0.976		0.851	$F_{(1,8)} = 154.6, p < 0.001$
Back	–0.321	$F_{(2,8)} = 80.79, p < 0.001$	0.953	0.102	$F_{(1,8)} = 17.2, p < 0.005$
Tense	–0.570	0.908	0.823	0.523	$F_{(1,7)} = 10.2, p < 0.01$
High	–0.522	$F_{(3,7)} = 10.9, p < 0.01$		0.295	$F_{(1,7)} = 8.5, p < 0.05$
Low	0.089			0.005	$F_{(1,7)} = 0.21$, n.s.
Dimension 3					
Back	0.868	0.947	0.896	0.793	$F_{(1,8)} = 57.89, p < 0.001$
High	–0.323	$F_{(2,8)} = 34.6, p < 0.001$		0.103	$F_{(1,8)} = 7.99, p < 0.01$
Dimension 4					
High	0.775	0.775 $F_{(1,9)} = 13.6, p < 0.005$	0.601	0.601	$F_{(1,9)} = 13.6, p < 0.005$

D2 is best predicted by a combination of LHigh and Back (variance accounted for = 95.3%, $F_{(2,8)} = 80.79$, $p < 0.001$). The best single predictor of D2 is LHigh which alone accounts for 85.1% of the variance ($r_m = 0.923$, $F_{(1,9)} = 51.5$, $p < 0.001$). Note that the Chomsky/Halle system requires three separate binary features (High, Low, Tense) to account for height distinctions in the English vowel system. Regressing D2 against a combination of these three features, one obtains a multiple correlation of 0.908, accounting for 82.3% of the variance. Not only does this result account for less variance than LHigh alone, but the proportionate contribution of the feature Low is nonsignificant ($F_{(1,7)} = 0.21$, n.s.). One argument that could be made at this point is that the values assigned for the vowels [ʌ, ɔ] on the feature Low (−Low in both cases) are incorrect. Perhaps one or even both of them should be considered low vowels instead of mid vowels. However, although not shown, all possible values for the feature Low were used in additional regression analyses. At no point did the regression results with differently valued vowels produce a significant contribution by Low, nor did the combination of High, Low, and Tense account for more variance that did LHigh alone.

The feature Tense, along with High, is a significant factor in the prediction of vowels along D2. However, note that D2 is not, strictly speaking, a Tense dimension, that is, it does not separate [i, eɪ, oʊ, u] from the other vowels (no perceptual study, to our knowledge, has ever found such a perceptual dimension). Rather, the feature Tense interacts with the feature High to produce a single multivalued height distinction. Viewing D2 as reflecting primarily LHigh, we can assign [ɔɪ, aɪ, aʊ] the values of 4High, 2High, and 2High, respectively.

The location of vowels along D3 (low-back onset) is best predicted by the features Back and High. These two features combine to account for 89.6% of the variance, with the feature Back explaining the majority of the variance (79.3%). Note that unlike at least one previous study (Fox, 1978) there was no indication that the third dimension was significantly related to Round.

The fourth dimension was best predicted by High, which accounted for a little more than 60% of the variance. This is the least amount of variance accounted for on any dimension yet considered, and an examination of the location of vowels along D4 will demonstrate that it is certainly not a particularly good ±High dimension. However, this result provides insight into the nature of the dimension. Except for some degree of skewing (in particular, the locations of [ʌ] and [a]), this dimension separates vowels with mid-onglides from those with high or low onglides. A similar dimension (mid/nonmid) was found by Terbeek (1977, p. 198) for English listeners. If one considers [ʌ] as +Low and [ɔ] and −Low, then the features High and Low will account for 74.4% of the variance in the location of the 11 monophthongal vowels along D4 ($r_m = 0.863$, $F_{(2,8)} = 11.6$, $p < 0.01$), although the contribution of the feature Low does not quite reach significance ($F_{(1,8)} = 4.46$, $0.05 < p < 0.01$).

To summarize, these results suggest (at least for D1, D2 and D3) that perceptual dimensions do, in fact, reflect divisions among vowels predicted by the categorical features used in phonological analyses. In addition, it was suggested that Ladefoged's single multi-valued feature LHigh may have more efficacy in capturing vowel height distinctions than does a combination of the Chomsky/Halle features High, Low, and

TABLE 5

Summary of perceptual dimensions suggested for English vowels

	Study	Stimuli	Context	Task	Analysis	Perceptual dimensions
(a)	Mohr & Wang (1968) (cf. Terbeek's 1977 reanalysis)	[i e æ y ø œ u o ɔ]*	none	paired comparison	COSCAL (2-way)	front/back, high/low
(b)	Singh & Woods (1970)	[i ɪ e ɛ æ ɑ ɔ o ʊ u ʌ ɚ]	none	paired comparison	MD-SCAL	high/low, front/back, retroflexion
(c)	Fox (1974)	[i eɪ ɑ æ aɪ o u]	[___t]	paired comparison	INDSCAL (3-way)	high/low, front onset, back/front glide
(d)	Terbeek & Harshman (1971), Terbeek (1977)	[i ɨ u y ø e æ o ʌ ɑ ɑʳ ɚ]*	[bəb___]	triadic comparison	PARAFAC (3-way)	round/nonround, high/low, retroflex/nonretroflex, back/nonback, mid/non-mid
(e)	Fox (1978, 1982)	[i ɪ ɛ æ ɑ ʌ ɔ ʊ u]	[h___d]	paired	INDSCAL	front/back, high/low
(f)	Shepard (based on Peterson and Barney)	[i ɪ ɛ æ ɑ ʌ ɔ ʊ u ɚ]	[h___d]	identification (confusion data)	Exponential Analysis of Proximities (2-way)	front/back, high/low
(g)	Terbeek (1974) Terbeek & Fox (1975)	[ɪə æə ɔə uw oʊ aʊ eɪ ɔɪ aɪ]	[___t]	paired comparison	INDSCAL	numbers of letters, types of letter groups, "A" to "I", F2 endpoint, glide direction, formant 1 onset

*Stimuli do not represent (necessarily) English vowel qualities.

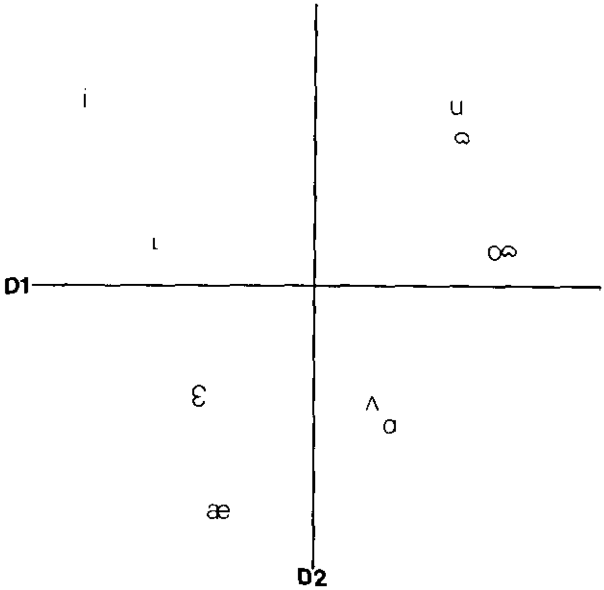

Fig. 6. Two-dimensional INDSCAL solution for 9 stimulus vowels.

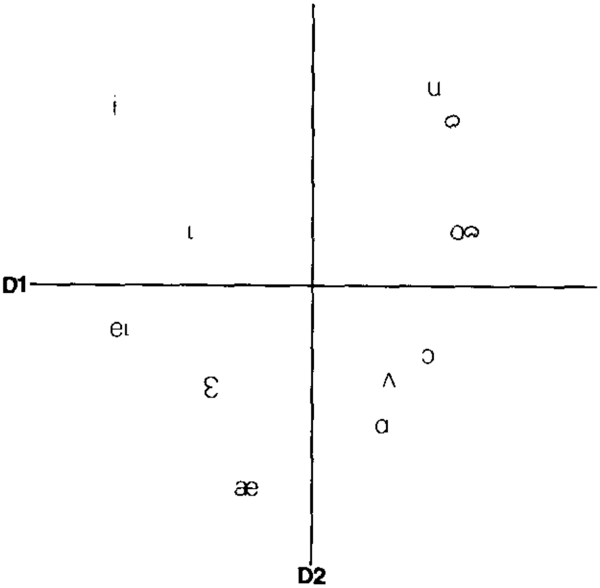

Fig. 7. Two-dimensional INDSCAL solution for 11 stimulus vowels.

Tense. These data would support the viewpoint discussed by Studdert-Kennedy (1976) and Goldstein (1977), among others, that perception involves the analysis of an acoustic syllable into an abstract structure of phonological features (and/or phonemes) by means of its acoustic features, and that the perceptual dimensions obtained through INDSCAL analysis reflect both the important acoustic features of the speech signal relevant to identification and aspects of its internal phonetic (or phonological) representation.

COMPARISON WITH PREVIOUS STUDIES

How do the perceptual dimensions discussed here compare to those extracted in other studies? Table 5 presents a summary of the literature pertinent to this question with regard to the perception of English vowels. Included are the stimulus vowels used, phonetic context, experimental task, analytic technique, and the dimension labels. As can be seen, many of the dimensions seem to show that the most common dimensions found correspond to front/back and high/low distinctions. As suggested in the introduction, the only study that shows a pattern at extreme variance to this is that of Terbeek (1974) and Terbeek and Fox (1975).

There is a possible difficulty in attempting to directly compare the dimensions found here with those in other studies, particularly since none of the studies cited in Table 5 include all 15 vowels used here. This, in fact, brings up the additional problem of determining whether or to what degree the extracted perceptual space was modified by the inclusion of diphthongs in a way that would not be expected if only the relatively monophthongal vowels were involved in the experiment. To address these concerns INDSCAL analyses were done using different subsets of the perceptual distance data. One set of analyses included intervowel perceptual distance data from the nine-vowel subset [i, ɩ, ɛ, æ, ɑ, ʌ, oɷ, ɷ, u] (corresponding to the stimulus set used in Fox, 1978) while the second subset included data from the 11-vowel subset [i, ɩ, eɩ, ɛ, æ, ɑ, ʌ, ɔ, oɷ, ɷ, u] (corresponding to the 11-vowel set used in Singh and Woods, 1970). As with the 15-vowel analyses, several INDSCAL analyses were done at several dimensionalities using different random starting configurations.

Figure 6 shows the two-dimensional INDSCAL solution for the nine-vowel subset, which accounts for 67.5% of the variance (corresponding to a correlation of 0.822 between the data and the spatial model). These dimensions again reflect front/back and high/low distinctions. In terms of acoustic measures, D1 is most highly correlated with $F2\mathrm{II}$ ($r = 0.981$, $p < 0.001$), while D2 is best correlated with $F1\mathrm{III}$ ($r = 0.931$, $p < 0.001$). The two-dimensional solution was chosen because of an elbow in the fit curve at the two-dimension level and because of the somewhat unstable nature of the third dimension extracted at different starting configurations. However, this third dimension, despite some skewing and instability, could be labeled as round/nonround. The perceptual dimensions thus defined are very similar to those found in Fox (1978, 1982) and two of the three dimensions found in Shepard (1972); in the latter study the third dimension did not resemble a clear rounding dimension, but it is unclear whether Shepard had optimally rotated the vowels into congruence with front/back, high/low, and round/nonround distinctions.

Figure 7 shows the 11-vowel INDSCAL solution. Again, the two-dimensional solution was chosen on the basis of the fit curve and extraction of a multiple solution at higher dimensions. This solution accounted for about 64% of the data variance ($r = 0.793$ between data and spatial model). The first dimension (front/back) was highly correlated with $F2$II ($r = 0.987, p < 0.001$), while the second dimension (high/low) was highly correlated with $F1$I ($r = 0.873, p < 0.001$). Comparing this perceptual plane with that found in Singh and Woods (1970), who used the same 11 vowels, we find general agreement in terms of the overall solution, but substantial disagreement in the placement of particular vowels. In Singh and Woods, the distance between [ɑ] and [ʌ] is considerable, whereas here (as in the 15-vowel solution) we find them relatively close. The placement of these two vowels in the perceptual space presented here is completely congruent with their relative positions in the acoustic space (see Fig. 2). However, as mentioned earlier, Singh and Woods do not give acoustic data about their stimulus vowels, and it is conceivable that [ʌ] was acoustically closer to [ɔ]; its placement in the perceptual space would suggest that this was the case.

There was little to suggest that inclusion of the diphthongs in the stimulus set has skewed the most salient perceptual dimensions or that diphthongs as opposed to monophthongs have significantly different perceptual structures. Comparisons among the D1 × D2 planes in the 9-, 11-, and 15-vowel INDSCAL solutions show that there are remarkable similarities between the perceptual dimensions. D1 and D2 are, in each case, very significantly related to $F2$ and $F1$ measures, respectively.

SUMMARY AND DISCUSSION

This was a multidimensional scaling study designed to discover the perceptual structure of a set of American English vowels which were not steady-state but which changed in phonetic quality over time. INDSCAL analysis of paired comparison judgments from 19 Midwestern listeners resulted in four perceptual dimensions: front/back, high/low, low-back onset, and mid/nonmid. Each of these dimensions was interpreted in terms of the acoustic properties of the stimuli and in terms of a set of distinctive features.

There was no evidence that subjects use radically different perceptual features when perceiving diphthongal vowels as opposed to monophthongal vowels (as suggested by Terbeek, 1974; Terbeek and Fox, 1975); rather, the resulting perceptual dimensions were very similar to those obtained from monophthongal studies. The scaling results suggest that both steady-state and diphthongal vowels are perceived primarily on the basis of the first two formants, although there was some suggestion that duration made some small contribution to the front/back decision. The only indication that subjects were attending to dynamic acoustic information (e.g., the direction or extent of formant change over time) in making their similarity judgments was that $\Delta F2$s made a contribution to the prediction of low-back onset (D3).

How does one reconcile these findings with those studies[1] (e.g., Strange, Verbrugge, Shankweiler and Edman, 1976; Strange, Edman and Jenkins, 1979; Gottfried and Strange, 1980; Verbrugge, Strange, Shankweiler and Edman, 1976; Verbrugge, Shankweiler and Fowler 1979) that have concluded, contrary to what is implied by our results, that listeners make critical use of dynamic acoustic information when identifying vowels? For example, Strange *et al.* (1976, 1979) suggest that consonant transitions as well as the vowel formants themselves must carry vowel information utilized by the listener in perceiving vowel quality, since vowels in context (e.g., [#pVp#]) were identified better than were vowels in isolation ([#V#]). These suggestions have been somewhat controversial and attempts have been made to explain such findings by reference to nonperceptual factors such as stimulus-response compatibility and memory load (cf. Diehl, McCusker and Chapman, 1981). However, even studies generally critical of such findings have suggested that dynamic information may play at least some role in vowel perception. For example, Assmann, Nearey and Hogan (1982) found that although gated vowels (i.e., stimuli representing only the center portion of vowels without speaker-controlled onset and offset) in a mixed-speaker condition were identified with a lower error rate than would be expected from results in Strange *et al.* (1976) and Verbrugge *et al.* (1979), nongated isolated full vowels in the same mixed-speaker condition were perceived more accurately. These results would suggest that dynamic properties (such as formant slopes and duration) at least *supplement* relatively steady-state cues (such as formant frequencies) in the vowel identification process.

Another potential difficulty in understanding the MD-scaling results presented in this paper lies in the fact that in the acoustic interpretations given some of the perceptual dimensions are hard to understand vis-a-vis on-line perceptual processing, since some of the dimensions were best predicted from acoustic information that occurs relatively late in the vowel (e.g., D2 is highly predictable from $F1$ at time slice III). Should this be taken to mean that subjects are waiting for that time point in the signal to calculate the placement of the vowel along a particular perceptual dimension? These issues are by no means trivial if only because they concern the role that MD-scaling studies should play in understanding the perception of speech.

To account for the above a closer examination of the MD-scaling paradigm and, in particular, the paired-comparison task must be made. More specifically, we believe that the INDSCAL results reported here accurately reflect some internal representation of the perceptual structure of the vowels used in the experiment but may reflect aspects of the categorized phonetic ideal more than the actual acoustic signal that subjects are hearing.

The paired-comparison task is a paradigm often used to obtain perceptual distance data in MD-scaling studies. A subject is presented two different stimuli with an interstimulus interval of 300–600 msec and is required to judge how similar or dissimilar the vowels of the two stimuli are. Since the tokens are presented sequentially and not simultaneously, this task certainly requires the first stimulus to be held in some type of

---

[1] *I would like to thank an anonymous reviewer for pointing out some of the problematic issues discussed here.*

memory for at least 500 msec. Given that the task is untimed (subjects often have 4–6 sec in which to respond), the second stimulus could also be held in memory for several seconds before an experiment response is made. The task thus represents a paradigm that makes critical use of memorial representations (of some type) and which imposes a significant memory load (more so than, say, an identification task). A critical issue in interpreting such MD-scaling results is in deciding whether the required similarity judgment is being made on the basis of phonetically uncategorized auditory images (stored in auditory memory), phonetically categorized images (stored in phonetic memory), or a combination of the two. The literature on speech discrimination may be of crucial relevance here since in overall form the paired-comparison task would seem to resemble the A–X (same/different) discrimination paradigm. However, there is at least one potentially important difference between the two: whereas in the discrimination task subjects may be faced with making both within-category and between-category discriminations, the paired-comparison task requires only between-category judgments. This difference may allow subjects to use only phonetic memory and therefore make comparisons using phonetic labels (or phonetic images associated with those labels) alone. This possibility, coupled with the reasonable assumption that auditory memory decays faster than phonetic memory (cf. Crowder, 1982a, 1982b) would support the contention that the perceptual distance judgments in the paired-comparison task are being made primarily on the basis of already-categorized (phonetic) elements. It could be, then, that although dynamic properties of the signal may be critical to the process of phonetic categorization (identification) itself, they may not be as important to similarity judgments made on already categorized stimuli.

Another issue relevant to understanding MD-scaling results stems from the nature of the stimuli themselves. Note that in this experiment most of the stimulus tokens correspond to real, high-frequency, English words. In the speech perception literature, especially in the work of psychologists rather than phoneticians, there has been much disagreement on the nature of segment (or phoneme) perception on the one hand, and word perception on the other. In particular, there is a question as to whether individual phonetic segments are identified (1) directly on the basis of the incoming auditory information (as in a bottom-up stage model of perception, e.g., Pisoni and Sawusch, 1975); (2) only after a lexical item containing the segment has been identified (as in a top-down model, e.g., Warren, 1976; Klatt, 1980); or (3) either top-down or bottom-up depending on which is the faster process (as in the dual-code hypothesis, cf. Foss and Blank, 1980). If the stimulus tokens in the experiment described here are being perceived only after lexical access, then the paired comparisons are being made on the basis of post-lexical, phonological codes (in the terminology of Foss and Blank, 1980). These phonological codes may or may not be identical to the phonetic codes involved in direct segment perception.

In conclusion, we feel that these and related questions must be addressed before the results from MD-scaling studies can be adequately integrated with data obtained using other experimental paradigms. In particular, one must lament the fact that the work done on speech perception using the MD-scaling paradigm has often been isolated from most other aspects of the field. For example, little effort has been made to define the status of extracted *perceptual dimensions* with regard to particular models of speech perception (with, perhaps, the exception of Repp, 1977; or Oden and Massaro, 1978) or to discuss such dimensions in terms of real-time processing, memory stores, auditory feature detection, or any of a number of interrelated issues. We feel that MD-scaling results can tell researchers about the nature of internalized phonetic images, but that perhaps such results have little to say directly concerning the real-time processing of acoustic signals. MD-scaling's limitations most likely stem from the characteristic modes of data-collection (e.g., pairwise comparison), which do not tap on-line processing, rather than a problem with the analytic methodology itself (though, as discussed earlier in this paper, a mathematical model like INDSCAL does make certain strong, and perhaps incorrect, assumptions about how individuals may differ in terms of perceptual structure). An important (planned) extension of the work reported here is to generate perceptual distance data using some type of speeded-classification paradigm (e.g., choice reaction-time) which might allow collection of data on the speech comprehension process at a level corresponding to immediate perception. Using these data, one might be able to determine more precisely the relationship between those "perceptual features" utilized in the actual perception process and "perceptual features" representing aspects of an internalized phonetic or phonological code.

REFERENCES

AINSWORTH, W. (1972). Duration as a cue in the recognition of synthetic vowels. *Journal of the Acoustical Society of America,* **51**, 648-651.

ASSMANN, P.F., NEAREY, T.M. and HOGAN, J.T. (1982). Vowel identification: Orthographic, perceptual, and acoustic aspects. *Journal of the Acoustical Society of America,* **71**, 975-989.

CARROLL, J.D. and CHANG, J.J. (1970). Analysis of individual differences in multidimensional scaling via a n-way generalization of "Eckart-Young" decomposition. *Psychometrika,* **35**, 283-319.

CARROLL, J.D. and WISH, M. (1974). Multidimensional perceptual models and measurement methods. In E.C. Carterette and M.P. Friedman (eds.), *Handbook of Perception, II: Psychophysical Judgment and Measurement* (New York), pp. 391-447.

CHOMSKY, N. and HALLE, M. (1968). *The Sound Pattern of English* (New York).

CROWDER, R.G. (1982a). Decay of auditory information in vowel discrimination. *Journal of Experimental Psychology: Human Learning, Memory, and Cognition,* **8**, 153-162.

CROWDER, R.G. (1982b). A common basis for auditory sensory storage in perception and immediate memory. *Perception & Psychophysics,* **31**, 477-483.

DIEHL, R.L., MCCUSKER, S.B. and CHAPMAN, L.S. (1981). Perceiving vowels in isolation and in consonantal context. *Journal of the Acoustical Society of America,* **69**, 239-248.

EDWARDS, A.L. (1979). *Multiple Regression and the Analysis of Variance and Covariance* (San Francisco).

FOSS, D.J. and BLANK, M.A. (1980). Identifying the speech codes. *Cognitive Psychology,* **12**, 1-31.

FOX, R.A. (1974). An experiment in cross-dialect vowel perception. *Papers from the 10th Meeting of the Chicago Linguistic Society* (Chicago), pp. 178-185.

FOX, R.A. (1978). Individual perceptual variation and a perception/production link in vowels. *Papers from the 14th Meeting of the Chicago Linguistic Society* (Chicago), pp. 98-107.

FOX, R.A. (1982). Individual variation in the perception of vowels: Implications for a perception-production link. *Phonetica,* **39**, 1-22.

GOLDSTEIN, L. (1977). Three studies in speech perception: Features, relative salience and bias. *UCLA Working Papers in Phonetics*, **39**, 1-87.

GANDOUR, J.T. and HARSHMAN, R.A. (1978). Crosslanguage differences in tone perception: A multidimensional scaling investigation. *Language and Speech*, **21**, 1-33.

GOTTFRIED, T.L. and STRANGE, W. (1980). Identification of coarticulated vowels. *Journal of the Acoustical Society of America*, **68**, 1626-1635.

HOLBROOK, A. and FAIRBANKS, G. (1962). Diphthong formants and their movements. *Journal of Speech and Hearing Research*, **5**, 38-58.

JAKOBSON, R., FANT, G. and HALLE, M. (1951). *Preliminaries to Speech Analysis* (Cambridge, Mass).

JONES, D. (1919). *An Outline of English Phonetics* (Cambridge).

KRUSKAL, J.B. and WISH, M. (1978). *Multidimensional Scaling* (Beverly Hills, Cal.).

KLATT, D.H. (1980). Speech perception: A model of acoustic-phonetic analysis and lexical access. In R. Cole (ed.), *Perception and Production of Fluent Speech* (Hillsdale, NJ), pp. 243-288.

LADEFOGED, P. (1971). *Preliminaries to Linguistic Phonetics* (Chicago).

LADEFOGED, P. (1975). *A Course in Phonetics* (New York).

MERMELSTEIN, P. (1977). On the relationship between vowel and consonant identification when cued by the same information. *Haskins Laboratories Status Report on Speech Research*, SR-51/52, 201-212.

MOHR, B. and WANG, W. (1968). Perceptual distances and the specification of phonological features. *Phonetica*, **18**, 31-45.

ODEN, G.C. and MASSARO, D.W. (1978). Integration of featural information in speech perception. *Psychological Review*, **85**, 172-191.

PARKER, F. (1977). Distinctive features and acoustic cues. *Journal of the Acoustical Society of America*, **62**, 1051-1054.

PETERSON, G.E. and BARNEY, H.L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, **24**, 175-184.

PETERSON, G.E. and LEHISTE, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, **32**, 693-703.

PISONI, D.B. and SAWUSCH, J.R. (1975). Some stages of processing in speech perception. In Cohen and Nooteboom (eds.), *Structure and Process in Speech Perception*, pp. 16-35.

POLS, L., VAN DER KAMP, L. and PLOMP, R. (1969). Perceptual and physical space of vowel sounds. *Journal of the Acoustical Society of America*, **46**, 458-467.

REPP, B.H. (1977). Dichotic competition of speech sounds: The role of acoustic stimulus structure. *Journal of Experimental Psychology: Human Perception and Performance*, **3**, 37-50.

SHEPARD, R. (1972). Psychological representation of speech sounds. In E.E. David and P.B. Denes (eds.), *Human Communication: A Unified View* (New York), pp. 67-113.

SINGH, S. and WOODS, G. (1970). Perceptual structure of 12 American vowels. *Journal of the Acoustical Society of America*, **49**, 1861-1866.

STRANGE, W., EDMAN, T.R. and JENKINS, J.J. (1979). Acoustic and phonological factors in vowel identification. *Journal of the Acoustical Society of America*, **5**, 643-656.

STRANGE, W., VERBRUGGE, R., SHANKWEILER, D. and EDMAN, T. (1976). Consonant context specifies vowel identity. *Journal of the Acoustical Society of America*, **60**, 213-224.

STUDDERT-KENNEDY, M. (1976). Speech perception. In N.J. Lass (ed.), *Contemporary Issues in Experimental Phonetics* (New York), pp. 243-293.

TERBEEK, D. (1974). Multidimensional scaling of the perception of diphthongs. Presented to the Classification Society, April 29-30.

TERBEEK, D. (1977). A cross-language multidimensional scaling study of vowel perception. *UCLA Working Papers in Phonetics*, **37**, 1-271.

TERBEEK, D. and FOX, R. (1975). INDSCAL study of the perceptual space of American diphthongs. Presented at the 90th Meeting, Acoustical Society of America. [*Journal of the Acoustical Society of America*, **58**, 91 (Abstract).]

TERBEEK, D. and HARSHMAN, R. (1971). Cross-language differences in perception of natural vowel sounds. *UCLA Working Papers in Phonetics,* **19,** 26-38.

TERBEEK, D. and HARSHMAN, R. (1972). Is vowel perception non-Euclidean? *UCLA Working Papers in Phonetics,* **22,** 13-29.

TIFFANY, W. (1953). Vowel recognition as a function of duration, frequency modulation, and phonetic context. *Journal of Speech and Hearing Research,* **18,** 289-301.

TUCKER, L.R. and MESSICK, S.J. (1963). Individual difference model for multidimensional scaling. *Psychometrika,* **37,** 3-27.

VERBRUGGE, R. and ISENBERG, D. (1978). Syllable timing and vowel perception. *Haskins Laboratories Status Report on Speech Research,* SR-55/56, 113-122.

VERBRUGGE, R.R., SHANKWEILER, D.P. and FOWLER, C.A. (1979). Context-conditioned specification of vowel identity. *Journal of the Acoustical Society of America,* **65,** S6 (Abstract); preprinted in J.J. Wolf and D.H. Klatt (eds.), *Speech Communication Papers.*

VERBRUGGE, R.R., STRANGE, W., SHANKWEILER, D.P. and EDMAN, T.R. (1976). What information enables a listener to map a talker's vowel space? *Journal of the Acoustical Society of America,* **60,** 198-212.

WARREN, R.M. (1976). Auditory illusions and perceptual processes. In N.J. Lass (ed.), *Contemporary Issues in Experimental Phonetics* (New York), pp. 389-418.

WICKELGREN, W. (1965). Distinctive features and errors in short-term for English vowels. *Journal of the Acoustical Society of America,* **38,** 583-588.

WISH, M. and CARROLL, J.D. (1974). Applications of individual differences to studies of human perception and judgment. In E.C. Carterette and M.P. Friedman (eds.), *Handbook of Perception II: Psychophysical Judgment and Measurement* (New York), pp. 449-491.

[Appendix over

APPENDIX I

Mean Acoustic Measurement for Stimulus Vowels

Acoustic Measure*	Vowels														
	i	ɪ	ɛ	æ	ɑ	ʌ	oʊ	ʊ	u	eɪ	ɔ	ɔɪ	aɪ	aʊ	ju
*F*1 I	302	413	577	785	810	772	533	477	362	458	607	520	813	821	326
*F*1 II	302	411	577	766	803	754	501	471	350	434	601	528	817	791	372
*F*1 III	305	421	584	769	811	744	474	471	348	399	619	517	755	744	378
*F*1 IV	305	437	569	771	801	696	391	482	348	367	595	446	504	557	378
*F*2 I	2371	2041	1921	1780	1139	1241	1128	1119	1054	2171	1035	969	1246	1399	2164
*F*2 II	2371	1995	1850	1750	1136	1241	1081	1116	1031	2219	1030	964	1246	1352	1814
*F*2 III	2371	1948	1817	1732	1148	1252	1067	1145	1042	2306	1041	1419	1509	1244	1510
*F*2 IV	2371	1867	1803	1714	1189	1317	1082	1250	1074	2326	1089	1940	1896	1125	1418
*F*3 II	3000	2714	2726	2716	2871	2864	2678	2678	2678	2882	2850	2857	2871	2857	2571
*F*3 IV	2943	2770	2723	2736	2821	2779	2678	2678	2678	2908	2714	2929	2857	2700	2500
Duration	439	378	410	460	445	474	445	378	427	443	430	455	480	474	463
F_0	145	146	143	143	149	145	144	145	146	146	144	145	146	145	146
*F*2–*F*1 I	2069	1628	1344	995	329	469	595	642	692	1713	428	449	433	578	1838
*F*2–*F*1 II	2069	1584	1226	984	333	487	580	645	681	1785	429	436	429	561	1442
*F*2–*F*1 III	2066	1527	1233	963	336	508	593	674	694	1907	422	902	754	500	1132
*F*2–*F*1 IV	2066	1430	1234	943	388	621	691	768	726	1959	494	1494	1492	568	1040

Acoustic Measure*	i	ι	ε	æ	ɑ	ʌ	oɷ	ɷ	u	eι	ɔ	ɔι	aι	aɷ	ju
	Vowels														
Δ$F1$ U	3	26	8	14	9	76	142	5	14	91	12	74	309	264	52
Δ$F1$ S	3	26	–8	–14	–9	–76	–142	5	–14	–91	–12	–74	–309	–264	52
Δ$F2$ U	0	174	118	66	50	76	46	131	20	155	54	971	750	274	746
Δ$F2$ S	0	–174	–118	–66	50	76	–46	131	20	155	54	971	750	–274	–746
Δ$F2$–$F1$ U	3	198	110	52	59	152	96	126	34	246	66	1045	1059	10	798
Δ$F2$–$F1$ S	3	–198	–110	–52	59	152	96	126	34	246	66	1045	1059	–10	–798

*Roman numerals refer to point in vowel where measurement taken (see Fig. 1). Absolute, unsigned, value for change in formant measures is indicated by U, signed values indicated by S.

APPENDIX II

Raw Input-Dissimilarity Matrices (symmetricized)*

Subject 1

0	14	20	23	27	23	24	23	28	12	23	29	20	24	21
14	0	15	13	24	19	22	21	20	18	26	26	18	18	25
20	15	0	10	18	15	14	23	24	10	17	30	14	25	19
23	13	10	0	11	12	22	25	30	14	19	28	16	15	24
27	24	18	11	0	7	24	23	30	12	16	22	11	15	31
23	19	15	12	7	0	13	22	28	16	13	21	13	12	22
24	22	24	22	24	13	0	14	11	22	12	9	20	22	13
23	21	23	26	23	22	14	0	11	25	18	11	31	24	12
28	20	24	30	30	28	11	11	0	23	17	18	28	27	6
12	18	10	14	12	16	22	25	23	0	24	28	19	21	26
23	26	17	19	16	13	12	18	17	24	0	20	14	14	24
29	26	30	28	22	21	9	11	18	28	20	0	19	20	18
20	18	14	16	11	13	20	31	28	19	14	19	0	11	26
24	18	25	15	15	12	22	24	27	21	14	20	11	0	25
21	25	19	24	31	22	13	12	6	26	24	18	26	25	0

Subject 2

0	22	25	29	27	32	32	24	25	13	26	27	25	28	21
22	0	9	17	15	15	24	22	28	15	13	31	23	26	27
25	9	0	10	20	16	23	23	23	20	18	30	21	29	34
29	17	10	0	22	16	29	27	28	26	22	31	25	30	27
27	15	20	22	0	7	23	16	23	30	8	20	14	24	29
32	15	16	16	7	0	20	9	18	30	8	24	15	20	31
32	24	23	29	23	20	0	14	11	31	14	24	24	18	26
24	22	23	27	16	9	14	0	16	32	12	28	16	18	25
25	28	23	28	23	18	11	16	0	30	19	23	28	22	10
13	15	20	26	30	30	31	32	30	0	27	21	20	23	30
26	13	18	22	8	8	14	12	19	27	0	21	21	16	26
27	31	30	31	20	24	24	28	23	21	21	0	13	27	23
25	23	21	25	14	15	24	26	28	20	21	13	0	24	26
28	26	29	30	24	20	18	18	22	25	16	27	24	0	21
21	27	34	27	29	31	26	25	10	30	26	23	26	21	0

Subject 3

0	27	17	31	28	23	24	27	26	19	33	30	26	31	28
27	0	15	23	26	24	25	19	28	15	24	29	26	30	29
17	15	0	9	23	18	25	26	26	17	22	24	28	24	28
31	23	9	0	14	13	26	26	28	19	13	20	34	21	25
28	26	23	14	0	5	24	14	20	19	10	16	12	14	23
23	24	18	13	5	0	10	14	15	23	9	20	14	15	25
24	25	25	26	24	10	0	17	16	27	27	19	31	20	19
27	19	26	26	14	14	17	0	8	25	19	19	32	16	20
26	28	26	28	20	15	16	8	0	29	18	27	32	20	7
19	15	17	19	18	23	27	25	29	0	22	22	24	27	24
33	24	22	13	10	9	27	19	18	22	0	8	19	16	28
30	29	24	20	16	20	19	19	27	22	8	0	19	22	24
26	26	28	34	12	14	31	32	32	24	19	19	0	26	26
31	30	24	21	14	15	20	16	20	27	16	22	26	0	25
28	29	28	25	23	25	19	20	7	24	28	24	26	25	0

Subject 4

0	16	16	21	23	21	24	21	21	22	24	25	23	22	26
16	0	13	21	23	20	21	20	20	18	23	24	25	24	24
16	13	0	11	21	21	21	21	25	14	22	14	22	23	23
21	21	11	0	21	20	20	23	23	18	20	24	21	20	22
23	23	21	21	0	16	18	19	20	20	16	21	17	21	26
21	20	21	20	16	0	17	17	22	25	22	23	18	23	24
24	21	21	20	18	17	0	16	19	22	17	18	23	20	23
21	20	21	23	19	17	16	0	15	24	23	23	25	22	22
21	20	25	23	20	22	19	15	0	25	20	24	27	21	17
22	18	14	18	20	25	22	24	25	0	22	24	21	23	26
24	23	22	20	16	22	17	23	20	22	0	21	20	16	25
25	24	24	24	21	23	18	23	24	24	21	0	22	24	27
23	25	22	21	17	18	23	25	27	21	20	22	0	20	23
22	24	23	20	21	23	20	22	21	23	16	24	20	0	23
26	24	23	22	26	24	23	22	17	26	25	27	23	23	0

Subject 5

0	24	18	22	20	20	29	23	26	23	28	21	28	21	24
24	0	16	18	20	18	24	22	21	22	22	28	15	20	19
18	16	0	15	19	18	14	18	21	18	18	22	18	19	20
22	18	15	0	18	20	25	23	24	14	22	23	26	26	24
20	20	19	18	0	10	16	12	14	22	10	10	12	14	16
20	18	18	20	10	0	15	9	16	24	12	11	16	17	15
29	24	14	25	16	15	0	10	10	24	10	14	25	14	22
23	22	18	23	12	9	10	0	13	25	14	16	19	18	19
26	21	21	24	14	16	10	13	0	26	15	19	21	14	14
23	22	18	14	22	24	24	25	26	0	21	21	24	27	31
28	22	18	22	10	12	10	14	15	21	0	11	18	14	21
21	28	22	23	10	11	14	16	19	21	11	0	15	14	18
28	15	18	26	12	16	25	19	21	24	18	15	0	22	20
21	20	19	26	14	17	14	18	14	27	14	14	22	0	19
24	19	20	24	16	15	22	19	14	31	21	18	20	19	0

Subject 6

0	24	29	30	27	23	30	29	31	25	31	25	31	24	23
24	0	14	18	18	15	20	16	25	25	18	28	24	23	29
29	14	0	4	17	8	15	23	25	13	10	28	26	23	28
30	18	4	0	9	13	17	27	28	12	9	30	26	11	28
27	18	17	9	0	7	14	14	27	14	8	25	23	12	29
23	15	8	13	7	0	8	12	22	21	10	18	15	10	31
30	20	15	17	14	8	0	9	13	31	11	12	26	8	27
29	16	23	27	14	12	9	0	11	25	15	19	28	14	20
31	25	25	28	27	22	13	11	0	30	27	12	30	27	8
25	25	13	12	14	21	31	25	30	0	18	27	23	25	30
31	18	10	9	8	10	11	15	27	18	0	27	29	8	28
25	28	28	30	25	18	12	19	12	27	27	0	18	21	25
31	24	26	26	23	15	26	28	30	23	29	18	0	28	31
24	23	23	11	12	10	8	14	27	25	8	21	28	0	24
23	29	28	28	29	31	27	20	8	30	28	25	31	24	0

Subject 7

0	28	24	31	32	31	30	30	30	27	28	32	23	31	29
28	0	23	26	26	27	27	31	30	23	26	31	19	30	28
24	23	0	16	29	25	32	30	31	27	18	31	26	29	26
31	26	16	0	21	17	29	32	32	28	27	28	33	22	32
32	26	29	21	0	9	29	30	27	27	16	31	21	24	31
31	27	25	17	9	0	21	25	29	29	21	24	17	22	27
30	27	32	29	29	21	0	20	22	25	23	23	30	24	25
30	31	30	32	30	25	20	0	24	33	26	26	31	30	27
30	30	31	32	27	29	22	24	0	31	29	28	32	23	18
27	23	27	28	27	29	25	33	31	0	22	31	25	27	31
28	26	18	27	16	21	23	26	29	22	0	25	30	27	28
32	31	31	28	31	24	23	26	28	31	25	0	25	25	29
23	19	26	33	21	17	30	31	32	25	30	25	0	26	27
31	30	29	22	24	22	24	30	23	27	27	25	26	0	29
29	28	26	32	31	27	25	27	18	31	28	29	27	29	0

Subject 8

0	20	20	22	25	29	33	27	32	16	31	24	25	22	23
20	0	14	18	22	18	29	17	21	24	22	21	29	29	22
20	14	0	8	18	16	19	23	22	13	18	21	26	19	21
22	18	8	0	11	13	17	25	24	14	17	19	28	22	33
25	22	18	11	0	8	21	19	18	16	13	22	13	16	26
29	18	16	13	8	0	23	13	25	25	13	21	14	21	22
33	29	19	17	21	23	0	15	24	20	19	19	25	16	30
27	17	23	25	19	13	15	0	19	26	27	19	20	23	18
32	21	22	24	18	25	24	19	0	28	21	21	30	29	10
16	24	13	14	16	25	20	26	28	0	25	29	22	23	33
31	22	18	17	13	13	19	17	21	25	0	12	20	21	23
24	21	21	19	22	21	19	19	21	29	12	0	18	28	27
25	29	26	28	13	14	25	20	30	22	20	18	0	23	30
22	29	19	22	16	21	16	23	29	23	21	28	23	0	23
23	22	21	33	26	22	30	18	10	33	23	27	30	23	0

Subject 9

0	26	27	31	33	31	36	33	35	20	33	32	32	31	31
26	0	22	26	31	26	34	26	32	25	29	31	32	31	30
27	22	0	12	20	21	30	25	28	25	24	29	31	24	32
31	26	12	0	21	23	33	31	32	26	25	28	30	29	34
33	31	20	21	0	13	25	23	30	29	20	27	24	25	31
31	26	21	23	13	0	28	22	31	30	18	28	16	27	34
36	34	30	33	25	28	0	21	27	34	21	24	34	24	35
33	26	25	31	23	22	21	0	20	35	17	28	32	28	26
35	32	28	32	30	31	27	20	0	35	28	31	36	31	8
20	25	25	26	29	30	34	35	35	0	34	33	31	34	36
33	29	24	25	20	18	21	17	28	34	0	25	31	33	33
32	31	29	28	27	28	24	28	31	33	25	0	28	29	32
32	32	31	30	24	16	34	32	36	31	31	28	0	32	36
31	31	24	29	25	27	24	28	31	34	27	29	32	0	27
31	30	32	34	31	34	35	26	8	36	33	32	36	27	0

Subject 10

0	18	18	20	20	24	24	21	23	16	24	27	19	26	25
18	0	20	16	17	17	20	20	22	18	23	29	15	22	21
18	20	0	13	17	16	16	24	22	16	18	25	17	20	30
20	16	13	0	16	20	23	27	23	15	15	22	19	20	25
20	17	17	16	0	14	20	18	19	15	16	24	15	16	23
24	17	16	20	14	0	21	17	22	22	16	21	15	19	24
24	20	16	23	20	21	0	18	20	24	18	21	25	20	20
21	20	24	27	18	17	18	0	15	26	20	23	26	17	14
23	22	22	23	19	22	20	15	0	20	20	20	14	19	7
16	18	16	15	15	22	24	26	20	0	20	23	16	24	23
24	23	18	15	16	16	18	20	20	20	0	17	23	20	25
27	29	25	22	24	21	21	23	20	23	17	0	25	21	22
19	15	17	19	15	15	25	26	24	15	23	25	0	23	21
26	22	20	20	16	19	20	17	19	24	20	21	23	0	20
25	21	30	25	23	24	20	14	7	23	25	22	21	20	0

Subject 11

0	28	18	35	31	29	35	31	27	12	29	27	27	27	23
28	0	26	34	29	30	31	25	29	28	32	32	29	31	27
18	26	0	25	22	31	29	31	24	27	32	28	29	23	27
35	34	25	0	15	29	32	25	29	23	26	31	31	25	29
31	29	22	15	0	10	33	22	28	31	13	31	30	27	27
29	30	31	29	10	0	31	28	25	24	17	25	19	28	28
35	31	29	32	33	31	0	22	28	31	28	15	27	20	25
31	25	31	25	22	28	22	0	14	28	22	27	31	25	20
27	29	24	29	28	25	28	14	0	29	23	20	32	21	12
12	28	27	23	31	24	31	28	29	0	31	32	30	33	29
29	32	32	26	13	17	28	22	23	31	0	23	31	27	24
27	32	28	31	31	25	15	27	20	32	23	0	33	27	26
27	29	29	31	30	19	27	31	32	30	31	33	0	31	31
27	31	23	25	27	28	20	25	21	33	27	27	31	0	25
23	27	27	29	27	28	25	20	12	29	24	26	31	25	0

Subject 12

0	36	29	33	34	27	33	33	33	36	33	32	36	30	36
36	0	33	28	31	30	32	28	29	29	33	34	28	34	32
29	33	0	13	19	18	31	31	28	14	17	20	33	30	32
33	28	13	0	18	11	32	29	31	14	20	29	32	20	34
34	31	19	18	0	5	27	28	29	22	23	24	25	30	31
27	30	18	11	5	0	28	29	28	27	11	23	17	31	30
33	32	31	32	27	28	0	22	25	32	23	22	33	25	33
33	28	31	29	28	29	22	0	15	32	27	27	32	30	22
33	29	28	31	29	28	25	15	0	33	24	25	32	25	20
36	29	14	14	22	27	32	32	33	0	29	34	34	34	33
33	33	17	20	23	11	23	27	24	29	0	19	29	28	30
32	34	20	29	24	23	22	27	25	34	19	0	32	28	33
36	28	33	32	25	17	33	32	32	34	29	32	0	32	34
30	34	30	30	30	31	25	30	25	34	28	28	32	0	29
36	32	32	34	31	30	33	22	20	33	30	33	34	29	0

Subject 13

0	23	27	30	31	25	30	28	29	24	32	33	25	32	27
23	0	10	26	28	23	30	27	32	23	30	34	22	31	32
27	10	0	11	26	22	26	25	30	29	27	28	28	29	30
30	26	11	0	31	17	28	31	31	21	28	33	32	24	32
31	28	26	31	0	5	24	23	28	26	11	26	18	25	32
25	23	22	17	5	0	15	18	30	30	12	27	16	22	31
30	30	26	28	24	15	0	14	30	28	24	27	30	31	29
28	27	25	31	23	18	14	0	17	32	13	23	30	27	26
29	32	30	31	28	30	30	17	0	31	26	30	33	31	5
24	23	29	21	26	30	28	32	31	0	31	33	28	31	31
32	30	27	28	11	12	24	13	26	31	0	16	26	21	29
33	34	28	33	26	27	27	23	30	33	16	0	30	29	29
25	22	28	32	18	16	30	30	33	28	26	30	0	30	32
32	31	29	24	25	22	31	27	31	31	21	29	30	0	35
27	32	30	32	32	31	29	26	5	31	29	29	32	35	0

Subject 14

0	25	21	31	13	26	24	10	20	20	34	24	19	22	23
25	0	16	22	16	21	23	18	16	19	14	27	19	21	19
21	16	0	16	12	16	21	28	27	18	22	24	26	21	23
31	22	16	0	12	19	21	31	25	17	22	30	24	25	23
13	16	12	12	0	7	22	25	21	19	15	17	24	13	29
26	21	16	19	7	0	20	19	25	23	13	11	24	10	20
24	23	21	21	22	20	0	28	15	18	16	16	26	24	29
10	18	28	31	25	19	28	0	7	20	15	25	32	28	18
20	16	27	25	21	25	15	7	0	18	15	23	31	17	6
20	19	18	17	19	23	18	20	18	0	24	28	26	31	25
34	14	22	22	15	13	16	15	15	24	0	18	20	25	20
24	27	24	30	17	11	16	25	23	28	18	0	16	27	20
19	19	26	24	24	24	26	32	31	26	20	16	0	19	27
22	21	21	25	13	10	24	28	17	31	25	27	19	0	32
23	19	23	23	29	20	29	18	6	25	20	20	27	32	0

Subject 15

0	19	20	32	30	27	23	23	27	19	30	28	24	27	19
19	0	17	24	23	28	33	31	20	32	19	34	20	30	30
20	17	0	7	16	23	12	26	27	11	9	22	24	27	25
32	24	7	0	13	18	30	31	32	10	18	23	20	26	33
30	23	16	13	0	17	19	21	21	11	8	21	12	28	23
27	28	23	18	17	0	17	17	13	28	18	23	20	22	13
23	33	12	30	19	17	0	12	21	20	18	15	25	29	23
23	31	26	31	21	17	12	0	14	31	22	24	25	24	23
27	20	27	32	21	13	21	14	0	28	18	19	32	27	4
19	32	11	10	11	28	20	31	28	0	14	27	14	30	25
30	19	9	18	8	18	18	22	18	14	0	15	16	19	27
28	34	22	23	21	23	15	24	19	27	15	0	23	13	28
24	20	24	20	12	20	25	25	32	14	16	23	0	20	26
27	30	27	26	28	22	29	24	27	30	19	13	20	0	25
19	30	25	33	23	13	23	23	4	25	27	28	26	25	0

Subject 16

0	18	23	29	27	26	29	24	24	25	25	27	27	27	30
18	0	21	18	28	21	22	20	25	22	27	27	25	24	24
23	21	0	12	21	15	24	24	24	17	20	24	24	20	25
29	18	12	0	20	14	20	21	25	19	21	21	26	22	22
27	28	21	20	0	8	22	18	24	25	17	20	17	17	28
26	21	15	14	8	0	16	16	21	23	17	17	11	18	24
29	22	24	20	22	16	0	14	22	28	16	14	26	22	23
24	20	24	21	18	16	14	0	14	24	19	16	24	22	15
24	25	24	25	24	21	24	14	0	25	21	22	27	21	9
25	22	17	19	25	23	28	24	25	0	24	26	26	21	25
25	27	20	21	17	17	16	19	21	24	0	13	17	15	26
27	27	24	21	20	17	14	16	22	26	13	0	24	25	24
27	25	24	26	17	11	16	24	27	26	17	24	0	20	21
27	24	20	22	17	18	22	22	21	21	15	25	20	0	25
30	24	25	22	28	24	23	15	9	25	26	24	21	25	0

Subject 17

0	19	19	27	30	27	29	24	25	18	26	29	23	31	31
19	0	14	19	28	21	29	21	25	23	23	30	21	26	30
19	14	0	8	16	18	22	22	25	21	14	20	22	28	31
27	19	8	0	16	15	29	24	31	22	15	27	24	30	32
30	18	16	16	0	4	15	12	14	31	8	10	15	12	24
27	21	18	15	4	0	14	8	13	26	8	12	14	12	19
29	29	22	29	15	14	0	11	14	30	10	13	27	16	20
24	21	22	24	12	8	11	0	10	28	10	11	23	13	16
25	25	25	31	14	13	14	10	0	26	9	14	27	14	10
18	23	21	21	31	26	30	28	26	0	30	29	20	31	32
26	23	14	15	8	8	10	10	9	30	0	10	20	15	22
29	30	20	27	10	12	13	11	14	29	10	0	20	17	24
23	21	22	24	15	14	27	23	27	20	20	20	0	22	32
31	26	28	30	12	12	16	13	14	31	15	17	22	0	22
31	30	31	32	24	19	20	16	10	32	22	24	32	22	0

Subject 18

0	9	9	12	20	10	29	31	29	6	26	11	10	30	27
9	0	4	8	20	14	28	22	24	11	22	16	19	25	29
9	4	0	5	11	10	25	20	19	7	15	18	12	24	28
12	8	5	0	5	4	22	20	27	11	10	25	12	18	30
20	20	11	5	0	4	13	12	17	17	7	17	12	10	16
10	14	10	4	4	0	9	13	24	20	5	18	12	14	26
29	28	25	22	13	9	0	10	6	32	8	15	30	6	15
31	22	20	20	12	13	10	0	4	31	7	21	33	12	9
29	24	19	27	17	24	6	4	0	33	12	30	27	11	4
6	11	7	11	17	20	32	31	33	0	29	15	11	32	36
26	22	15	10	7	5	8	7	12	29	0	16	29	8	14
11	16	18	25	17	18	15	21	30	15	16	0	13	27	28
10	19	12	12	12	12	30	33	27	11	29	13	0	26	32
30	25	24	18	10	14	6	12	11	32	8	27	26	0	7
27	29	28	30	16	26	15	9	4	36	14	28	32	7	0

Subject 19

0	15	30	26	27	32	28	27	33	30	32	31	27	33	31
15	0	14	16	21	25	30	23	29	21	24	31	13	23	31
30	14	0	5	21	18	26	24	29	13	16	29	21	28	33
26	16	5	0	19	16	21	27	32	11	25	29	27	16	33
27	21	21	19	0	4	23	25	25	18	6	22	16	14	35
32	25	18	16	4	0	12	18	28	26	4	24	13	13	34
28	30	26	21	23	12	0	12	28	24	21	11	26	21	34
27	23	24	27	25	18	12	0	13	29	14	21	22	26	27
33	29	29	32	25	28	28	13	0	28	25	19	29	16	9
30	21	13	11	18	26	24	29	28	0	30	24	19	25	34
32	24	16	25	6	4	21	14	25	30	0	19	20	9	36
31	31	29	29	22	24	11	21	19	24	19	0	20	21	31
27	13	21	27	16	13	26	22	29	19	20	20	0	19	35
33	23	28	16	14	13	21	26	16	25	9	21	19	0	33
31	31	33	33	35	34	34	27	9	34	36	31	35	33	0

*These are Vowel x Vowel matrices with the order of column and row stimuli being:

[i, ɩ, ɛ, æ, ɑ, ʌ, oɷ, ɷ, u, eɩ, ɔ, ɔɩ, aɩ, aɷ, ju]

APPENDIX III

INDSCAL Solution in Four Dimensions

	Dimension 1	Dimension 2	Dimension 3	Dimension 4
		Vowel Coordinates		
[i]	0.442	–0.257	–0.243	0.177
[ɩ]	0.179	0.115	–0.358	0.252
[eɩ]	0.427	0.024	–0.248	–0.274
[ɛ]	0.155	0.217	–0.314	–0.258
[æ]	0.078	0.377	–0.266	–0.333
[ɑ]	–0.050	0.327	0.219	0.036
[ʌ]	–0.061	0.281	0.218	0.106
[ɔ]	–0.189	0.144	0.223	–0.206

[oʊ]	–0.248	–0.184	0.167	–0.363
[ʊ]	–0.300	–0.255	0.013	0.088
[u]	–0.354	–0.382	–0.131	0.170
[ɔɪ]	0.064	–0.312	0.442	–0.284
[aɪ]	0.343	0.155	0.370	0.408
[aʊ]	–0.222	0.119	0.121	0.056
[ju]	–0.265	–0.398	–0.211	0.423

Subject Weights

S1	0.606	0.790	0.628	0.167
S2	0.701	0.652	0.618	0.349
S3	0.691	0.594	0.653	0.486
S4	0.652	0.539	0.703	0.521
S5	0.706	0.367	0.654	0.596
S6	0.624	0.744	0.533	0.484
S7	0.636	0.688	0.557	0.515
S8	0.643	0.560	0.628	0.596
S9	0.728	0.636	0.645	0.527
S10	0.709	0.666	0.484	0.580
S11	0.646	0.613	0.570	0.471
S12	0.576	0.643	0.554	0.612
S13	0.624	0.652	0.666	0.509
S14	0.489	0.652	0.582	0.485
S15	0.597	0.630	0.542	0.594
S16	0.584	0.632	0.697	0.559
S17	0.782	0.581	0.643	0.443
S18	0.904	0.496	0.407	0.335
S19	0.613	0.710	0.646	0.539