Journal of Phonetics (1985) 13, 205–217

Multidimensional scaling and perceptual features: evidence of stimulus processing or memory prototypes?*

Robert Allen Fox

The Ohio State University, Speech and Hearing Science, 154 N. Oval Mall, Columbus, Ohio 43210, U.S.A.

Received 11th *December* 1984

Abstract: Two experiments were conducted to discover whether the perceptual features obtained in multidimensional scaling studies accurately reflect the acoustic properties of the stimulus vowels or merely aspects of the vowels' long-term memory prototypes. In each experiment subjects were required to make paired-comparison judgments of vowel stimuli on two sets of stimulus pairs. The stimulus sets were composed of the same phonetic qualities but differed in terms of subphonemic (acoustic) characteristics. Multivariant analysis of variance indicated that subjects' perceptual distance judgments were sensitive to the subphonemic as well as phonemic distinctions among the stimuli. The perceptual spaces extracted using INDSCAL analysis also accurately corresponded to these acoustic space variations. The results are argued to demonstrate the use of covert, subphonemic, categories during the paired-comparison task.

Introduction

In developing a model of speech perception one must be concerned with, among other things, determining the number and nature of the perceptual features utilized in the identification and internal representation of vowels. Evidence on the nature of these perceptual features has often been obtained using multidimensional (MD) scaling techniques and the perceptual dimensions extracted have usually been characterized either in terms of the acoustic parameters of the stimuli (e.g. Terbeek, 1977; Fox, 1978, 1982, 1983) or traditional articulatory distinctions (e.g., Singh & Woods, 1970; Shepard, 1972). These perceptual dimensions are viewed as representing the critical factors extracted from the incoming acoustic stream upon which vowel identifications are based. It is our task here to more closely examine the nature of these experimentally obtained perceptual dimensions in terms of how sensitive they are to subphonemic acoustic variations in the stimuli.

At issue is the degree to which perceptual similarity judgments (and the perceptual

---

*Preliminary versions of this paper were presented at the 56th Meeting of the Midwestern Psychological Association, 3–5 May 1984, Chicago, Illinois and at the 107th Meeting of the Acoustical Society of America, 6–10 May, 1984.

0095-4470/85/020205 + 13 $03.00/0

dimensions obtained after such judgments have been analysed using MD-scaling techniques) are sensitive to the acoustic nature of the stimuli being compared. For example, Fox (1983), utilizing a stimulus set containing both naturally produced monophthongs and dipthongs, failed to find any perceptual dimension corresponding to dynamic acoustic information (such as the change in formant patterns over time). If one views perceptual dimensions as reflecting the critical acoustic information utilized during the vowel perception process, it is difficult to reconcile this failure with those studies which have concluded that listeners make critical use of dynamic acoustic information when identifying vowels (e.g. Strange, Verbrugge, Shankweiler & Edman, 1976; Verbrugge, Strange, Shankweiler & Edman, 1976; Strange, Edman & Jenkins, 1979; Verbrugge, Shankweiler & Fowler, 1979; Gottfried & Strange, 1980). Although these studies have been somewhat controversial (cf. Diel, McCusker & Chapman, 1981), even those studies generally critical of such findings have suggested that dynamic information may at least supplement relatively steady-state acoustic cues in the vowel identification process (Assmann, Nearey & Hogan, 1982).

In light of his results, Fox (1983) suggested the possibility that the paired-comparison task allowed subjects to make extensive use of phonetic rather than auditory memory in completing the experimental task. This would mean that the similarity judgments obtained during MD-scaling experiments are done on the basis of the phonetic labels of the categorized vowel stimuli (or the phonetic images associated with those labels) alone. Thus, rather than providing information about the acoustic cues using during the phonetic categorization process, perceptual features may provide information about the nature of representation in phonetic memory. Such a view would be consistent with Rakerd & Verbrugge (1983) and Farrar (1981) who demonstrated, using MD-scaling techniques, that the perceptual structure of imagined stimuli were very similar to those of heard stimuli. In addition, the results obtained by Repp, Healy & Crowder (1979) suggest that subjects make use of phonetic labels even in vowel discrimination, a task not unlike paired-comparisons. A critical issue, then, in interpreting MD-scaling results, *vis-à-vis* on-line perceptual processing models, is to determine whether the similarity judgments obtained (and the perceptual features eventually extracted from such data) are made on the basis of the phonetically uncategorized auditory images (Crowder, 1982), phonetically categorized images (cf. Repp *et al.*, 1978; Massaro, 1975) or both.

At least one study has suggested that subjects *can* make reliable estimates of perceptual distances based upon acoustic information alone, Kewley-Port & Atal (1983) required subjects to scale vowel pairs drawn from an 11-step [i]–[ɪ] continuum and obtained an interpretable three-dimensional perceptual space. However, in this experiment subjects were not making between-category judgments, as is most common in MD-scaling experiments, but were forced to make within-category comparisons. The question remains as to whether these results are generalizable to the normal scaling task which uses many different phonetic categories.

To investigate this question, an MD-scaling experiment was conducted utilizing two different sets of vowel stimuli. Each stimulus set contained the same seven vowel qualities, but differed in terms of the formant structure (i.e. acoustic quality) of a subset of these vowels (producing between-set subphonemic variations). The experiment examined whether or not the acoustic distance variations produced corresponding differences in the perceptual similarity judgments. If such differences are not obtained it will suggest that listeners judge perceptual distances on the basis of internalized phonetic images or labels.

Experiment 1

Method

Stimuli

The experimental stimuli were composed of the seven American English vowels [iɪɛæaʊu] in a [h__d] context. The stimuli were generated using a cascade/parallel speech synthesis program (Klatt, 1980) implemented on a PDP 11/23 computer. One version of the vowels [i a ʊ u] and two different versions of the vowels [ɪ ɛ æ] were constructed. Formant frequencies for each of these 10 tokens are shown in Table I. All vowels were steady-state (no change in formant frequencies over time) and each stimulus was 420 ms in duration. The fundamental frequency for each token began at 130 Hz and fell linearly to 110 Hz after 400 ms where it remained until the end of the token. This produced a very natural-sounding speech token.

Two different stimulus sets were constructed using different subsets of these 10 different tokens, but each set contained only one version of the vowels [ɛ æ]. The first stimulus set (SET A) contained the vowels [i ɪ1 ɛ2 æ1 a ʊ u] while the second stimulus set (SET B) contained the vowels [i ɪ2 ɛ1 æ2 a ʊ u]. Each stimulus set was composed of all possible vowel dyads, excluding vowels paired with themselves (diagonal cells in perceptual distance matrices are presumed to contain zero), in two different random orders. This produced 84 different vowel pairs. The inter-token interval was 500 ms with different stimulus pairs occurring every 5 s. The stimulus were converted into analog form at a digital-to-analog conversion rate of 10 kHz samples/s, low-pass filtered at 5 kHz, and recorded by a Harmon-Kardon 301 cassette recorder.

The phonetic quality of each stimulus tokens was ascertained using a forced-choice identification task with ten subjects who did not participate in the scaling experiment. The results indicated that the stimulus vowels were identified consistently in terms of the indicated phonetic quality. In particular, all stimulus vowels (except [ɛ2] and [æ1]) were identified as having the expected vowel quality at least 96% of the time. The vowels [ɛ] and [æ] were identified as the expected vowel 71 and 81% of the time, respectively.

Subjects

Twenty subjects participated in the perceptual scaling experiment as paid volunteers. None had participated in a speech experiment before. All were native speakers of English from the Central Ohio region with no known hearing loss.

Table I. Formant frequences used in the synthesis of stimulus vowels (all values in Herz)

Vowel	Formant 1	Formant 2	Formant 3
i	310	2100	2960
ɪ1	400	1800	2960
ɪ2	422	1780	2558
ɛ1	487	1720	2523
ɛ2	545	1677	2488
æ1	590	1677	2453
æ2	620	1660	2430
a	700	1220	2600
ʊ	450	1100	2350
u	320	900	2200

Procedure

Subjects were required to listen to two sequentially presented stimulus tokens and to judge the similarity/dissimilarity of the medial vowels of the tokens on a nine-point scale. Subjects read typed instructions describing the experimental task. They were told nothing concerning the differences between the two stimulus sets except that the same token pairs would occur but in different presentation orders. Subjects sat in an IAC sound booth and heard the stimuli at a comfortable listening level over Sennheiser HD 320 headphones as reproduced by a Harmon/Kardon 301 stereo cassette recorder. Subjects practised on an introductory book of 10 stimulus pair at the beginning of each tape. Subjects were randomly assigned to two different groups which heard the stimulus sets either in the order SET A, SET B or SET B, SET A. There was a short (2–3 min) interval between sets.

Each subject's responses were checked for consistency before any further analysis was completed. This was done to ensure that excess "noise" was not included in the perceptual distance data (cf. Terbeek, 1977; Fox 1983). Correlations were obtained between judgments of the two random presentation orders in each stimulus set. A listener's responses were required to exhibit a correlation coefficient (Pearson's r) significant to the 0.01 level of significance on both stimulus sets. Five subjects from the original 20 were eliminated and data were collected from an additional five subjects who met the consistency criterion.

Results

Analysis of the raw distance data

The obtained symmetricized perceptual distance matrices averaged across subjects for Set A and Set B appear in Tables II and III, respectively. Prior to multidimensional scaling analysis, two subsets of these perceptual distance data were analyzed using multivariant analysis of variance (Ray, 1980): (1) the distances among the vowels [i u a] (whose acoustic distances were the same for both stimulus sets) and (2) the distances among the vowels [ɪ ɛ æ] (whose acoustic distances varied across stimulus sets).

For the acoustically stable distances there were no significant effects due to ORDER ($F(3,16) = 1.48, p > 0.25$), SET ($F(3,16) = 0.93, p > 0.45$) or the TAPE × ORDER interaction ($F(3,16) = 2.12, p > 0.14$).[1] However, for the variable acoustic distances there was a main effect of SET ($F(3,16) = 6.73, p < 0.006$) but no significant effect for

Table II. Symmetricized perceptual distance matrix collapsed averaged over all subjects for SET A

	i	ɪ	ɛ	æ	a	ʊ	u
i	0.0	21.2	23.8	24.4	27.8	28.2	29.2
ɪ	21.2	0.0	15.1	17.9	23.3	21.2	26.5
ɛ	23.8	15.1	0.0	8.8	19.6	20.7	26.4
æ	24.4	17.9	8.8	0.0	19.6	20.7	26.4
a	27.8	23.3	19.6	19.6	0.0	20.8	27.5
ʊ	28.2	21.2	20.7	21.5	20.8	0.0	15.1
u	29.2	26.5	26.4	25.9	27.5	15.1	0.0

---

[1] The test statistic reported for all multivariant analyses is the Hotelling-Lawley trace with the appropriate F approximations.

Table III. Symmetricized perceptual distance matrix collapsed averaged across all subjects for SET B

	i	ɪ	ɛ	æ	a	ʊ	u
i	0.0	19.4	19.6	24.6	26.7	28.3	29.8
ɪ	19.4	0.0	10.0	18.2	23.2	20.7	27.3
ɛ	19.6	10.0	0.0	14.6	19.9	21.4	25.1
æ	24.6	18.2	14.6	0.0	15.4	23.0	25.0
a	26.7	23.2	19.9	15.4	0.0	20.5	26.7
ʊ	28.3	20.7	21.4	23.0	20.5	0.0	14.2
u	29.8	27.3	25.1	25.0	26.7	14.2	0.0

either ORDER ($F(3.64) = 0.35$, $p > 0.78$) or the SET × ORDER interaction ($F(3,16) = 0.78$, $p > 0.52$).

The results demonstrate that acoustic distance variations produce corresponding perceptual distance variations and thus suggest that subjects' similarity responses are not made strictly on the bases of categorized phonetic labels. However, to better understand the relationship between the acoustic space of the stimuli and the derived perceptual space of the subjects, as well as to allow these results to generalize to the use of MD-scaling methods as a whole, we must analyze the data using MD-scaling analysis.

INDSCAL Analysis

The obtained similarity judgments were next analyzed using INDSCAL (Carroll & Chang, 1970). Multidimensional-scaling algorithms, in general, view similarity judgments as indicative of the perceptual distance between the objects being compared and accounts for these distances by constructing an *n*-dimensional space such that the distances between the vowels in the space correspond optimally to the obtained similarity responses. No assumptions concerning the number or nature of these dimensions are made prior to analysis (cf. Carroll & Wish, 1974; Wish & Carroll, 1974; Kruskal & Wish, 1978). INDSCAL, the program used here, is a three-way scaling technique and considers perceptual differences across subjects in developing a solution. The orientation of the space is fixed and cannot be rotated without worsening the fit of the solution to the data.

INDSCAL analysis was first done using data obtained with the SET A stimulus sets in one to four dimensions. Four different analyses were done at each dimensionality at different random staring configurations. Determination of the "correct" dimensionality of the solution was based upon three criteria: variance accounted for, uniqueness of solution, and interpretability. In each case the two-dimensional solution was indicated as the correct solution. The cumulative percentage of variance accounted for by the two-dimensional solution was 66.7%, a 32.1% increase from the best one-dimensional solution (34.6%).

The "fit" curve, denoting variance accounted for by each dimensionality, is shown in Fig. 1. Note that there is an elbow in the curve between dimensions 2 and 3, suggesting that all the significant dimensions have been extracted by the two-dimensional level. Since at both three and four dimensions INDSCAL derived multiple, nonequivalent solutions accounting for the same proportion of the variance, the uniqueness-of-solution criterion (Terbeek, 1977) pointed to the two-dimensional solution. In terms of interpretability, the first two dimensions extracted were easily explained in terms of front/back and high/low—the two dimensions almost universally extracted in MD-scaling

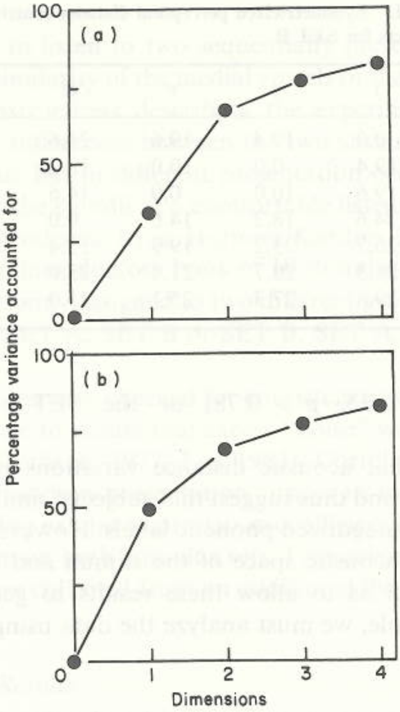

Figure 1 Fit curves for INDSCAL analyses done on the SET A (a) and SET B (b) data at one to four dimensions.

studies utilizing vowels. However, the higher-dimensional solutions (three dimensions and above) obscured these distinctions.

The similarity data for SET B were then analyzed with INDSCAL at one to four dimensions. Again, four different analyses were done at each dimensionality. However, instead of using a different random starting configuration for each analysis, the group space coordinates for each of the SET A solutions were used—a method utilized by Rakerd & Verbrugge (1983). This allows for more direct comparison of the resulting solutions across the two stimulus sets. The fit curve for the SET B analyses is shown in Fig. 1. All three decision criterion again indicated that the two-dimensional solution represented the correct dimensionality of the perceptual space. This solution accounted for 69.1% of the variance, an increase of 19.5% over the best one-dimensional solution (49.6%).

The perceptual spaces for SET A and SET B are shown in Figs 2 and 3, respectively. Each figure is a dimension 1 (D1) by dimension 2 (D2) plot of the stimulus vowels. Note that, in general, these two perceptual spaces are extremely similar—especially in the placement of the vowels [i a ʊ u]. Dimension1 is a front/back dimension which is highly correlated with F_2 (SET B, $r = 0.850, p < 0.001$; SET B, $r = 0.976, p < 0.001$) while D2 is a high/low dimension which correlates highly with F_1 (SET A, $r = 0.917$, $p < 0.001$; SET B, $r = 0.986, p < 0.001$).

Of greatest importance here, however, are the distances among the vowels [ɪ ɛ æ] in the SET A solution compared with the SET B solution. Note that in the INDSCAL solution, just as in the raw perceptual data (as one would expect), the distance between [ɪ]–[ɛ] is

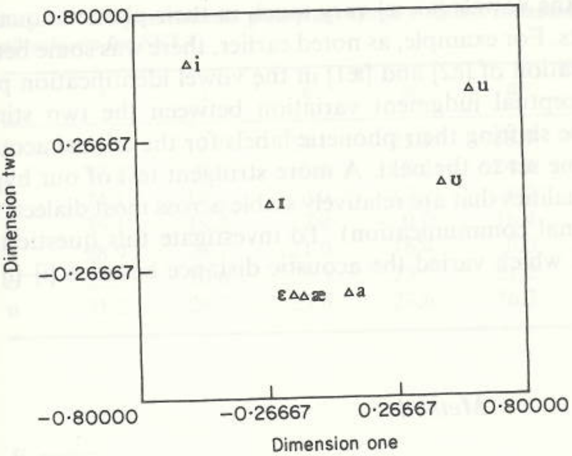

Figure 2 Two-dimensional INDSCAL group stimulus space for the SET A data.

greater in SET A than SET B while the distance between [ε]–[æ] is smaller. Differences in placement of the vowels between these two plots are, in general, consistent with the acoustic space differences.

Discussion

These results support the contention that subjects make their similarity judgments in a scaling task based, at least in part, on an acoustic representation of the stimuli rather than strictly upon phonetic labels (or the phonetic characteristics of the appropriate long-term memory prototype). If subjects based their judgments strictly upon phonetic labels we would expect the perceptual distance to be identical between SET A and B which, of course, they are not. This, in turn, would argue that information about on-line perceptual processing *can* be obtained using MD-scaling procedures.

However, at least one objection can be made that removes the force of our argument.

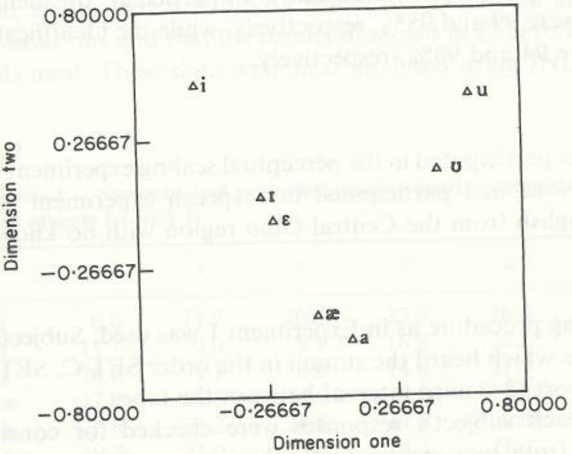

Figure 3 Two-dimensional INDSCAL group stimulus space for the SET B data.

In particular, the vowels [ɪ ε æ] vary much in their phonetic qualities among American English dialects. For example, as noted earlier, there was some between-subject variation in the identification of [ε2] and [æ1] in the vowel identification pretest. This could be a source of perceptual judgment variation between the two stimulus sets since some subjects may be shifting their phonetic labels for the critical acoustically variable vowel stimuli from one set to the next. A more stringent test of our hypothesis would require use of vowel qualities that are relatively stable across most dialects and idiolects (Winifred Strange, personal communication). To investigate this question, a second experiment was conducted which varied the acoustic distance between [i]–[ɪ] and [u]–[ʊ].

Experiment 2

Method

Stimuli
The basic experimental stimuli were the same as in Experiment 1 except that two new tokens, [i2] and [u2], were constructed. The frequencies of F_1, F_2 and F_3 for [i2] were 344, 1988 and 2814 Hz, respectively. The frequencies of F_1, F_2 and F_3 for [u2] were 385, 1000 and 2275 Hz, respectively. Each vowel was steady-state in a [h—d] context and was 420 ms in duration. The same falling F_0 contour was used as in the other 10 tokens. To avoid confusion throughout the remainder of this paper, the [i] and [u] tokens used in Experiment 1 will be referred to as [i1] and [u1], respectively.

Two different stimulus tapes were constructed using a different subset of the available tokens, but each tape contained only one version of each phonetic quality. SET C used the vowels [i1 ɪ1 ε1 æ2 a ʊ u2] while SET D used the vowels [i2 ɪ1 ε1 æ2 a ʊ ü1]. Each tape contained all possible vowel dyads (excluding vowels paired with themselves) in two different random orders. The acoustic vowel spaces in these two different stimulus sets varied only in terms of the acoustic distances involving [i] and [u].

The phonetic quality of each stimulus token was ascertained using a forced-choice identification task with 12 subjects who did not participate in the scaling task. The results indicated that all stimulus vowels were identified very consistently (at least 94% accuracy) in terms of the expected phonetic quality. In particular, the identification scores for [i1] and [i2] as /i/ were 99 and 95%, respectively, while the identification scores for [u1] and [u2] as /u/ were 94 and 98%, respectively.

Subjects
Twenty subjects participated in the perceptual scaling experiment while fulfilling a course requirement. None had participated in a speech experiment before. All were native speakers of English from the Central Ohio region with no known hearing loss.

Procedure
The same scaling procedure as in Experiment 1 was used. Subjects were assigned to two different groups which heard the stimuli in the order SET C, SET D or SET D, SET C. There was a short (2–3 min) interval between the tapes.

As before, each subject's responses were checked for consistency before further analysis. Data from two inconsistent subjects (one from each ORDER group) were eliminated from any further analysis.

Table IV. Symmetricized perceptual distance matrix collapsed averaged across all subjects for SET C

	i	ɪ	ɛ	æ	a	ʊ	u
i	0.0	17.1	24.1	25.1	28.9	22.9	31.2
ɪ	17.1	0.0	10.4	18.2	24.1	20.6	26.2
ɛ	24.1	10.4	0.0	12.9	21.0	22.5	25.6
æ	25.1	18.2	12.9	0.0	16.4	25.2	27.6
a	28.9	24.1	21.0	16.4	0.0	21.2	16.3
ʊ	22.9	20.6	22.5	25.7	21.2	0.0	7.7
u	31.2	26.2	25.6	27.6	26.3	7.7	0.0

Results

Analysis of Raw Distance Data

The obtained symmetricized perceptual distance matrices averaged across subjects for SET C and SET D appear in Tables IV and V, respectively. Prior to multidimensional scaling analysis, selected distances were analysed using multivariant techniques. Multivariant analysis of variance was done on two sets of perceptual distance data: (1) the distances among the vowels [ɪ ʊ a] (similar to the [i a u] subset used in Experiment 1), which represented acoustically stable distances across the two stimulus sets, and (2) the perceptual distances between [i]–[ɪ] and [u]–[ʊ], which represented variable acoustic distances across the stimulus sets.

For the perceptual distances among the vowels [ɪ ʊ a], there were no significant effects due to SET ($F(3,14) = 0.26$, $p > 0.85$), ORDER ($F(3,14) = 2.35$, $p > 0.11$) or the SET × ORDER interaction ($F(3,14) = 0.38$, $p > 0.77$). However, for the perceptual distances between [i]–[ɪ] and [u]–[ʊ] there was a significant main effect due to SET ($F(2,14) = 17.14$, $p < 0.001$). There was no significant effect due to ORDER ($F(2,14) = 1.96$, $p > 0.18$) or the SET × ORDER interaction ($F(2,14) = 1.01$, $p > 0.38$).

These results suggest, again, that a subject's similarity judgments are sensitive to subphonemic acoustic variations and that the results obtained in Experiment 1 were not an artifact of the vowels used. These data were next analysed using INDSCAL.

Table V. Symmetricized perceptual distance matrix collapsed averaged across all subjects for SET D

	i	ɪ	ɛ	æ	a	ʊ	u
i	0.0	11.7	20.8	22.8	26.8	19.7	30.6
ɪ	11.7	0.0	9.1	16.8	23.2	21.0	28.4
ɛ	20.8	9.1	0.0	14.6	21.8	23.8	29.1
æ	22.8	16.8	14.6	0.0	17.3	24.7	30.3
a	26.8	23.2	21.8	17.3	0.0	21.4	28.2
ʊ	19.7	21.0	23.8	24.7	21.4	0.0	12.4
u	30.6	28.4	29.1	30.3	28.2	12.4	0.0

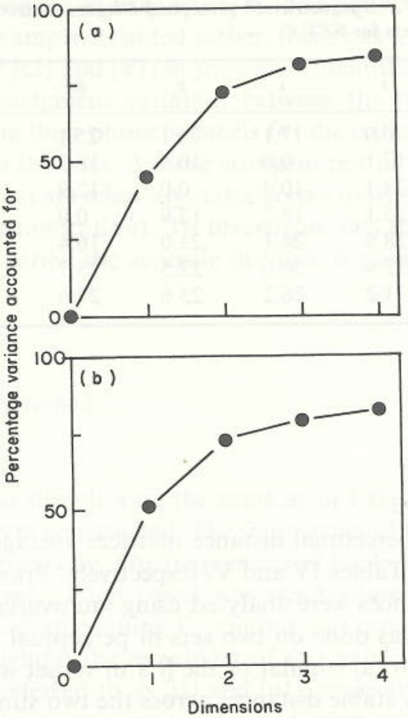

Figure 4 Fit curves for INDSCAL analyses done on the SET C and SET D data at one to four dimensions.

INDSCAL analysis

INDSCAL analysis was first done using the SET C data. Four different analyses were done in one to four dimensions using different random starting configurations. The SET D data were also analysed at one to four dimensions but utilizing the vowel coordinates from each of the SET C solutions as the starting configurations, as in Experiment 1. The resulting fit curves are shown in Fig. 4. For both sets of data the two-dimensional solution was chosen as the correct dimensionality on the basis of variance-accounted-for-, uniqueness-of-solution and interpretability. The two-dimensional solution for the SET C data accounted for 72.4% of the variance, an increase of 27.1% over the best one-dimensional solution (45.3%). Variance accounted for by the two-dimensional solution for the SET D data was an almost identical 72.9%, an increase of 21.5% over the best one-dimensional solution (51.4%).

The solutions for SET C and SET D are shown in Figs 5 and 6, respectively. Again, these two perceptual spaces are very similar. D1 in each solution is a front/back dimension which is highly correlated with F_2 (SET C, $r = 0.968$, $p < 0.001$; SET D, $r = 0.973$, $p < 0.001$). Each D2 corresponds to a high/low dimension which correlates well with F_1 (SET C, $r = 0.842$, $p < 0.005$; SET D, $r = 0.884$, $p < 0.003$). As one would expect, given the raw perceptual data, the perceptual distance in the INDSCAL solutions between [i]–[ɪ] is greater for SET C than SET D while the distance between [u]–[ʊ] is greater for SET D than for SET C. In fact, it is clear from a comparison of Figs 5 and 6 that the configurations of the perceptual spaces, in general, accurately reflects the acoustic difference between the two stimulus sets.

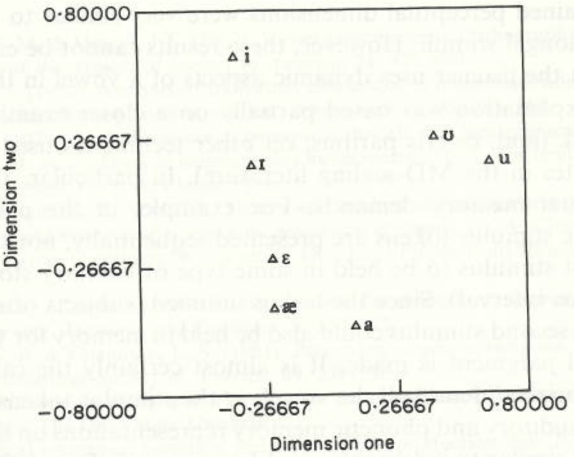

Figure 5 Two-dimensional INDSCAL group stimulus space for the SET C data.

Discussion and conclusions

The results for both Experiments 1 and 2 have demonstrated that subjects are sensitive to subphonemic acoustic variations when making paired-comparison judgments. Thus, these data do not seem to support the conjecture that subjects make such comparisons on the basis of phonetic labels or long-term memory prototypes alone. This, in turn, supports the contention that perceptual features, obtained through MD-scaling of such judgments, will also be sensitive to the acoustic quality of the vowel stimuli and may provide information about relatively low-level acoustic processing. However, before this conclusion is accepted, we should again review the issues raised in Fox (1983) concerning the nature of perceptual features.

One impetus for this study was the failure of Fox (1983) and Terbeek & Fox (1975) to find any evidence, using MD-scaling techniques, that subjects use radically different perceptual features when perceiving diphthongal as opposed to monophthongal vowels;

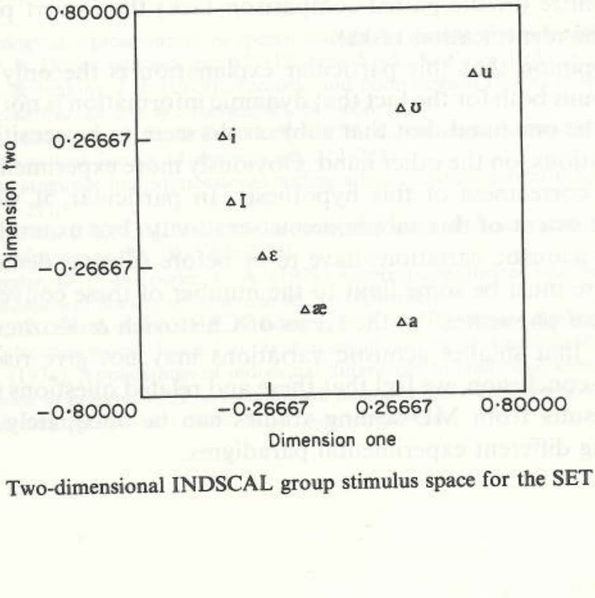

Figure 6 Two-dimensional INDSCAL group stimulus space for the SET D data.

rather, the obtained perceptual dimensions were very similar to those studies utilizing only monophthongal stimuli. However, these results cannot be easily reconciled with a contention that the listener uses dynamic aspects of a vowel in the perceptual process. One possible explanation was based partially on a closer examination of the paired-comparison task (and, ceteris paribus, on other techniques used to obtain perceptual distance estimates in the MD-scaling literature). In particular, this experimental task makes substantial memory demands. For example, in the paired-comparison task utilized here, the stimulus tokens are presented sequentially, not stimultaneously which requires the first stimulus to be held in some type of memory store for at least 500 ms (the interstimulus interval). Since the task is untimed (subjects often take 4–6 s in which to respond), the second stimulus could also be held in memory for several seconds before an experimental judgment is made. It is almost certainly the case that subjects have implicitly categorized (identified) the vowels in the stimulus tokens being compared and thus have both auditory and phonetic memory representations on the basis of which they may make their similarity judgments. As I have argued (Fox, 1983), the failure to find any evidence of dynamic acoustic information in MD-scaling experiments could be explained by suggesting that subjects are making their paired-comparison judgments on the basis of phonetic, rather than auditory memory. Thus, while dynamic properties of the acoustic signal may be critical to the process of phonetic categorization, they may not play an important role in the similarity judgments between two already categorized stimuli.

The data obtained in the present study would seem to contradict this explanation. However, I will argue that the present results still may be best explained in terms of phonetic memory representations. In particular, Repp *et al.* (1979, p. 142) suggested a model of vowel discrimination (which, as suggested above, is similar in many ways to the paired-comparison task) in which discriminations are based entirely upon phonetic distinctions. The surplus discrimination obtained over that predicted by phonetic labeling can be explained by the presence of convert phonetic categories which are used in discrimination (or paired-comparison) but ineligible for the labeling task. Chistovich & Kozhevnikov (1970, p. 38) presented a similar model of discrimination and identification. If one adopts this viewpoint, then we can hypothesize that similarity judgments are based on phonetic labels alone, although there were more convert phonetic labels for subjects to utilize (in the paired-comparison task) than overt phonetic categories (as required in the identification tasks).

It is our opinion that this particular explanation is the only one which will satisfactorily account both for the fact that dynamic information is not reflected in perceptual features, on the one hand, but that subjects do seem to be sensitive to "subphonemic" acoustic variations, on the other hand. Obviously more experimentation is necessarily to establish the correctness of this hypothesis. In particular, it would be important to determine the extent of this subphonemic sensitivity. For example, how small do these subphonemic acoustic variations have to be before subjects demonstrate no sensitivity to them? There must be some limit to the number of these convert phonetic categories ("psychological phonemes" in the terms of Chistovich & Kozhevnikov, 1970), and we would expect that smaller acoustic variations may not give rise to perceptual space variations. In conclusion, we feel that these and related questions still must be answered before the results from MD-scaling studies can be adequately integrated with data obtained using different experimental paradigms.

References

Assmann, P. F., Nearey, T. M. & Hogan, J.T. (1982). Vowel identification: Orthographic, perceptual, and acoustic aspects, *Journal of the Acoustical Society of America*, **71**, 975–989.

Carroll, J. D. & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via *n*-way generalization of "Eckart-Young" decomposition, *Psychometrika*, **35**, 283–319.

Carroll, J. D. & Wish, M. (1974). Multidimensional perceptual models and measurement methods. In *Handbook of perception, II: psychophysical judgment and measurement*. (E. C. Carterette and M. P. Friedman, editors). New York: Academic Press.

Chistovich, L. A. & Kozhenvnikov, V. A. (1970). Theory and method of research and perception of speech signals. In *Theory and methods of research on perception of speech*. (L. A. Chistovich, V. A. Kozhevnikov and V. I. Galunov, editors). Washington, DC: Joint Publications Research Service (JPRS-504233), 1–101.

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

Diehl, R. L., McCusker, S. B. & Chapman, L. S. (1981). Perceiving vowels in isolation and in consonantal context. *Journal of the Acoustical Society of America*, **69**, 239–248.

Farrar, C. O. (1981). *A prototype theory of speech perception*. An unpublished dissertation submitted to the Department of Linguistics, The Ohio State University.

Fox, R. A. (1978). Individual perceptual variation and a perception/production link. *Papers from the 14th Regional Meeting of the Chicago Linguistic Society*, Chicago, Illinois, pp 98–107.

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

Fox, R. A. (1983). Perceptual structure of monophthongs and diphthongs in English, *Language and Speech*, **26**, 21–60.

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

Gottfried, T. L. & Strange, W. (1980). Identification of coarticulation vowels, *Journal of the Acoustical Society of America*, **68**, 1626–1635.

Kewley-Port, D. & Atal, B. S. (1983) Perceptual distances between vowels in a limited phonetic space, *Journal of the Acoustical Society of America*, **73**, S54 (Abstract).

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer, *Journal of the Acoustical Society of America*, **67**, 971–995.

Kruskal, J. B. & Wish, M. (1978). *Multidimensional scaling*. Beverly Hills, California: Sage Publications.

Massaro, D. W. (1975). *Understanding language: an information-processing analysis of speech perception, reading, and psycholinguistics*. New York: Academic Press.

Rakerd, B. & Verbrugge, R. R. (1983) Memory for isolated vowels and for vowels in a consonantal context: An individual differences scaling comparion, *Journal of the Acoustical Society of America*, **73**, S54 (Abstract).

Ray, A. A. (1982). *SAS User's guide: statistics*. Cary, North Carolina.

Repp, B. H., Healy, A. F. & Crowder, R. G. (1978) Categories and context in the perception of isolated steady-state vowels, *Journal of Experimental Psychology: Human Perception and Performance*, **5**, 129–145.

Singh, S. & Woods, G. (1970). Perceptual structure of 12 American vowels, *Journal of the Acoustical Society of America*, **49**, 458–467.

Shepard, R. (1972). Psychological representation of speech sounds. In *Human communication: a unified view* (E. E. David and P. B. Denes, editors), pp. 67–113. New York: McGraw Hill.

Strange, W., Edman, T. R. & Jenkins, J. J. (1979). Acoustic and phonological factors in vowels identification, *Journal of the Acoustical Society of America*, **65**, 643–656.

Strange, W., Verbrugge, R., Shankweiler, D. & Edman, T. (1976). Consonantal context specifies vowel identity, *Journal of the Acoustical Society of America*, **60**, 213–224.

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

Terbeek, D. & Fox, R. A. (1975). INDSCAL study of the perceptual space of American diphthongs, *Journal of the Acoustical Society of America*, **58**, S91 (Abstract).

Verbrugge, R. R., Shankweiler, D. P. & Fowler, C. A. (1979). Context-conditioned specification of vowel identity, *Journal of the Acoustical Society of America*, **65**, S6 (Abstract).

Verbrugge, R. R., Strange, W., Shankweiler, D. P. & Edman, T. R. (1976). What information enables a listener to map a talker's vowels space? *Journal of the Acoustical Society of America*, **60**, 198–212.

Wish, M. & Carroll, J. D. (1974). Applications of individual differences to studies of human perception and judgment. In *Handbook of perception, II: psychophysical judgment and measurement*, (E. C. Carterette and M. P. Friedman, editors), pp. 449–491. New York: Academic Press.