Cluster analyses of two acoustic confusion matrices

B. J. T. MORGAN*

M.R.C. Applied Psychology Unit, 15 Chaucer Road, Cambridge CB2 2EF, England

Three methods of cluster analysis are used to illustrate two acoustic confusion matrices. It is shown how the methods complement each other and, together, "explain" the large, unwieldy matrices.

This paper deals with the acoustic confusion matrices of Conrad (1964) and Hull (in press), both derived from recognition tasks. In Conrad's case, the stimulus set comprised the letters of the alphabet; he used 10 untrained speakers, the presentation rate of the stimuli was one letter every 5 sec, and the sound level of the speech signal was continuously monitored and (on average) an equal amount of white noise added. Three hundred Ss were used, and the final overall error rate was 61%. In Hull's case, the stimulus set comprised the letters of the alphabet and all the digits, 1-9; six trained speakers were used, the presentation rate of the stimuli was one every 5 sec, and again equal noise on speech was employed. In this case, 135 Ss were used and the final overall error rate was 34%. In both cases, the stimulus lists were balanced with respect to several desirable characteristics, the particular details of which can be found in the papers cited above.

In each case, the results of the experiments were summarized by a stimulus/response matrix, the (ij) th element of which denotes the number of times that "i" was the response to the presentation of Stimulus j. For ease of reference, the two matrices are given in Tables 1 and 2. Note that the letter "z" has the English, rather than the American, pronunciation.

As they stand, the matrices convey a certain amount of basic information about the confusions; for example, it is immediate and not surprising that B and C confuse more readily than B and R in each case. However, it is clear that the matrices are poor illustrators of all the information they contain.

Shepard (1963) advocated the use of nonparametric multidimensional scaling (henceforth referred to as MDSCAL; for a description of this technique, see Kruskal, 1964a, b) as a means of illustrating such matrices. The most recent advocates of MDSCAL as an

illustrative technique are probably Fillenbaum and Rapoport (1971). This paper presents the results of analyzing both matrices by MDSCAL, single-link cluster analysis (henceforth referred to as SLINK; for a description of this technique, see Johnson, 1967) and also the B(2) cluster analysis of Jardine and Sibson (1971). The basic principles underlying these three methods and the reasons for using these particular methods will now be outlined in the following section.

THE METHODS USED

MDSCAL attempts to place each point Q_i , representing the stimulus i , in n -dimensional space so that if δ_{ij} denotes the dissimilarity between i and j then distance $(Q_i Q_j) \leq \text{distance}(Q_k Q_l)$ if and only if $\delta_{ij} < \delta_{kl}$. A measure of how well we have been able to do this for a given n is provided by a measure known as the "stress." Kruskal (1964a) considered a stress of 20% as "poor" and a stress of 10% as fair, etc., but these are only heuristic guides. Usually, we start with $n = 2$; a two-dimensional representation is the easiest to construct and appreciate ($n = 1$ gives rise to particular problems and is usually never used). If we decide, largely from the stress value, that our attempt at fitting a set of points in n dimensions is not good enough, then we have to increase n by 1 and start again (see also the comments of Kendall, 1971a, in this respect). When ties in the dissimilarities arise, then we have two possibilities: the *primary* approach states that when $\delta_{ij} = \delta_{kl}$ we do not care which of distance $(Q_i Q_j)$ or distance $(Q_k Q_l)$ is the larger, nor whether they are equal or not. The more restrictive *secondary* approach requires distance $(Q_i Q_j) = \text{distance}(Q_k Q_l)$ whenever $\delta_{ij} = \delta_{kl}$. Townsend (1971), whose data had many ties, has compared the two approaches.

The above statement of what MDSCAL tries to do is a statement of *global* scaling. Sometimes we need not be so fastidious and need only require *local* scaling, that is, we only look at pairs of stimuli for which $\delta_{ij} < \delta_{jk}$ and then require distance $(Q_i Q_j) \leq \text{distance}(Q_i Q_k)$. An instance of the use of local scaling in archaeology is given by Sibson (1972).

The method by which MDSCAL works is iterative, and for each run of MDSCAL we must decide upon a

*The author wishes to thank Dr. R. Conrad and Mrs. A. J. Hull for their permission to cite their data, Dr. R. Sibson for making available his efficient computerized algorithms for performing MDSCAL, SLINK, and B(2), and Mr. K. Taylor, who did a lot of the computing work. The computing was performed by the Cambridge University computer, TITAN. Mrs. A. J. Hull, Miss M. Lukianowicz, and Dr. J. Morton were kind enough to read an earlier draft of this paper and made some very useful comments and suggestions.

Table 1
Averaged Confusion Matrix of Conrad (1964) (Total Number of Presentations of Each Item 1440)

Response		Stimulus																										
		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	
A	2	25	9	18	9	1	9	7	19	15	700	164	15	176	-478	33	12	17	17	84	23	79	8	24	32	3	25	A
B	2	171	197	107	2	142	15	13	30	11	3	11	10	4	-94	0	7	84	33	168	4	7	8	24	32	3	25	B
C	2	12	19	24	4	19	7	4	10	4	0	4	5	0	35	40	0	7	42	4	20	4	5	3	3	3	3	C
D	4	128	71	166	11	104	4	12	21	6	3	11	7	0	34	47	5	0	43	13	105	4	7	6	11	10	1	D
E	1	8	195	25	11	708	31	100	9	16	7	59	77	9	111	66	6	16	65	119	3	19	5	6	1	1	1	E
F	1	10	6	4	4	3	35	7	12	30	25	13	8	6	3	0	316	4	3	138	4	4	19	5	81	1	1	F
G	4	40	22	47	14	0	4	9	8	5	0	3	3	1	18	19	0	5	12	33	17	7	7	1	1	1	1	G
H	4	3	43	27	76	3	3	1	15	8	7	9	15	61	6	16	4	43	12	2	5	14	28	1	12	1	1	H
I	1	1	2	5	14	3	1	3	14	6	51	52	10	23	1	5	188	71	6	5	3	4	28	5	1	1	1	I
J	1	10	5	6	3	1	1	0	15	7	158	16	9	4	7	8	7	1	2	3	10	8	3	4	28	5	1	J
K	1	69	6	7	0	6	9	15	7	64	16	9	17	21	6	12	7	6	1	3	10	8	3	4	28	5	1	K
L	1	69	6	7	0	6	9	15	7	64	16	9	17	21	6	12	7	6	1	3	10	8	3	4	28	5	1	L
M	1	106	10	14	5	13	22	9	45	65	47	83	32	28	4	57	18	1	2	3	12	4	61	41	4	1	1	M
N	1	106	10	14	5	13	22	9	45	65	47	83	32	28	4	57	18	1	2	3	12	4	61	41	4	1	1	N
O	1	116	10	15	10	18	22	16	47	104	127	51	512	54	2	7	32	23	3	6	4	17	9	40	41	4	1	O
P	1	116	10	15	10	18	22	16	47	104	127	51	512	54	2	7	32	23	3	6	4	17	9	40	41	4	1	P
Q	1	162	80	20	172	11	167	21	12	18	89	132	15	30	9	8	41	26	46	15	20	26	6	15	50	4	1	Q
R	1	43	90	59	52	0	103	16	8	12	3	31	3	4	211	0	5	505	16	91	2	5	3	31	4	1	1	R
S	1	3	3	13	14	0	103	16	8	12	3	31	3	4	211	0	5	505	16	91	2	5	3	31	4	1	1	S
T	1	20	18	9	17	4888	4	151	3	38	34	123	23	11	15	1	1	15	4	2	5	7	162	28	9	1	1	T
U	1	3	232	158	124	14	116	20	9	11	13	1	12	11	18	2	9	5	7	3	5	191	10	111	5	1	1	U
V	1	10	74	90	128	7	189	26	50	24	1	12	23	11	79	175	6	8	5	29	13	8	5	2	10	1	1	V
W	4	122	61	147	55	1	35	18	0	25	24	1	22	23	11	79	112	3	5	347	13	4	8	15	4	1	1	W
X	4	31	5	23	6	5	11	4	17	5	11	6	8	11	3	6	3	2	2	17	19	2	0	6	15	4	1	X
Y	7	1	6	3	6	5	12	9	2	1	1	6	2	1	1	2	184	3	2	2	2	1	1	14	91	56	1	Y
Z	4	1	5	1	4	40	4	21	2	26	98	25	62	7	1	2	32	5	7	4	6	6	22	4	22	22	1	Z

Acoustic Confusion Matrix of Hull (1972) (Total Number of Presentations of Each Item = 410)

[illegible]

starting configuration. Usually, we try several random initial configurations because the iterative technique is seeking a minimum of the stress function and any particular start could lead to a local, and not a global, minimum. Alternatively, we could start from a specific initial configuration in n dimensions suggested by the properties of the stimuli perhaps or by an MDSCAL fit already obtained in $(n+1)$ dimensions and projected onto the first n principal components (see Kendall, 1971a). The iterative technique has only been feasible by the development of fast computers, and it has been found, using the computerized algorithm developed by Dr. R. Sibson, that 50 iterations usually suffice to find a minimum of the stress function.

The MDSCAL results to be given in this paper were obtained by using global scaling, primary treatment of ties, and are the best of several runs, each starting from a different random initial configuration and continuing for 50 iterations.

SLINK is a hierarchic cluster method giving rise to a tree structure or dendrogram, each level of which gives a system of clusters. As we move up the tree, clusters at lower levels condense to form larger clusters, and this is what is meant by a "hierarchic" cluster method. B(2), which we shall describe later, does not have this property, so that the number of clusters at any level is not necessarily greater than or equal to the number of clusters at a higher level, as is true of SLINK. At the lowest level, we have a set of clusters with just one object in each cluster, while at the highest level we have just one cluster containing all the objects. Frequently, we may decide on just a section of the tree at some intermediate level which seems to result in a "meaningful" set of clusters, and largely discard the information in the rest of the dendrogram. SLINK has certain optimal properties, the most important for us here being its *continuity* (see Jardine & Sibson, 1971); if one is going to compare dendrograms, then it is rather important that the method giving rise to those dendrograms be a continuous method, so that small changes in the original data do not give rise to discontinuous jumps in the form of the dendrogram. There are also shortcomings to SLINK: it forces the data into an ultrametric structure (the dendrogram can be characterized by an ultrametric distance function: see Jardine & Sibson, 1968), and in so doing may drastically distort the data: also, it tends to form long straggly clusters (this is known as "chaining"). At any level, h , the clusters designated by SLINK at that level consist of objects which can be linked together by a chain passing through other objects in the cluster so that consecutive objects on the chain have a dissimilarity less than or equal to h . Clearly, one can then have objects in the same SLINK cluster which are very dissimilar but are chained together by a set of intermediate objects with which they are more similar.

Jardine and Sibson (1971) have generalized SLINK to give a sequence of continuous cluster methods, B(k).

B(1) is SLINK; B(k), for $k > 1$, is a nonhierarchic method permitting *overlapping* clusters with at most $(k-1)$ objects in the overlap. The B(k) methods ($k > 1$) break down any straggly chains formed by SLINK and illustrate the finer structure hidden by SLINK. A simple graph-theoretic description of how they do this is given by Jardine and Sibson (1968). It suffices to say here that B(k) clusters at any level ($k > 1$) require a more stringent connectedness property than the sufficient one for SLINK clusters, namely that a single chain of similarities can be drawn through the elements of the cluster at that level. However, the B(k) dendrograms ($k > 1$) are very complicated and we have only considered the case $k = 2$, and even then have only presented cross sections of the dendrograms at levels which gave a "meaningful" set of resulting clusters. Both the SLINK and B(k) methods have been carried out by means of computerized algorithms, and a further disadvantage of the B(k) methods is that their algorithms can be remarkably time consuming. For p objects, the computation time taken by B(k) is approximately p^{k+2} for $k \leq p/2$.

Much of this general introduction on the three methods will become clearer when we consider the results which follow.

The three methods, MDSCAL, SLINK, and B(2), require as their starting point a symmetric matrix of dissimilarities between all pairs of stimuli, and this raises the initial problem of how to convert the nonsymmetric matrices of Tables 1 and 2 into dissimilarity matrices. To a certain extent, large stimulus-response confusions are associated with large confusions when stimulus and response are interchanged [e.g., 8 to H (329) and H to 8 (159) in Table 2], but there are some drastic asymmetries, e.g., the confusions into and out of A in Table 2, the worst being A to K (8) and K to A (488), and in Table 1, K to N (127) should be compared with N to K (17). We have taken as a measure of dissimilarity between i and j , δ_{ij} , given by,

$$\delta_{ij} = - \left(\frac{r_{ij}}{\sum_{i=1}^m r_{ij}} + \frac{r_{ji}}{\sum_{j=1}^m r_{ji}} \right)$$

where m is the number of stimuli and r_{sk} is the number of times that s is confused for k . This is probably the simplest measure we could propose and the one used by Shepard (1963) in his analysis of Rothkopf's data (1957) on the confusion of Morse code signals. Several studies, for example Morgan, Chambers, and Morton (in press), Townsend (1971), Van der Kamp and Pols (1971), and Wilson (1963), have considered alternative measures to the above, but, generally speaking, the different measures have yielded the same end result. The data of Table 1 were also analyzed by the measure of dissimilarity first suggested by Shepard (1957), which is equivalent to assuming Luce's (1959) choice model for

Fig. 1. SLINK analysis of the data of Table 1. We have taken, as the measure of distortion imposed by each dendrogram on the original dissimilarity matrix, the $\hat{\Delta}_1$ measure suggested by Jardine & Sibson (1971, p. 104). This measure takes values in the range [0, 1], and in this case $\hat{\Delta}_1 = 0.12$.

the elimination of response bias, viz.

$$\delta_{ij} = - \left(\frac{r_{ij} r_{ji}}{r_{ii} r_{jj}} \right).$$

Virtually the same representations obtained as under the simple measure above, and consequently the results are not given here.

Thus, the following analysis loses the information to be found in the *asymmetry* of Tables 1 and 2, but it is felt that this is a reasonable first step: the following results will tell us how the stimuli confuse, "on the average," and will, it is hoped, be useful in helping to design, for example, low-confusing subsets of the stimuli. In particular, however, any similarity in the analyses of the data of Tables 1 and 2 would possibly mask a difference in the asymmetries of the two tables.

Finally, we must justify our use of these three particular methods out of the many that present themselves (see Cormack, 1971, for a general review of the field). MDSCAL, if it works (i.e., if the stress in a small number of dimensions is acceptably small), provides a very useful representation that most people can accept and readily understand. It is, however, not

primarily a cluster method, at best suggesting clusters which may then be identified by means of a bona fide cluster method such as SLINK (cf. Holloway & Jardine, 1968), and this has been attempted in Figs. 3, 4, and 5, by taking "meaningful" sections of the dendrograms. SLINK has been used because it is probably the simplest possible cluster method and it has been assisted by B(2) in order to bring out some of the structure hidden by SLINK. All three methods have the desirable property that they produce the same representations when applied to monotone transformations of the original data (see Sibson, 1972). This approach, of using MDSCAL in conjunction with cluster-analysis methods, is also that advocated by Fillenbaum and Rapoport (1971).

RESULTS

The SLINK analyses of the two tables are given in Figs. 1 and 2.

In each case, the distortion that the hierarchical analysis of SLINK has imposed on the data (see Jardine & Sibson, 1971, Chap. 11, for discussion of this point) is remarkably small, reassuring us that SLINK is an appropriate method for the analysis of this data. (The distortion that the method B(k) imposes on the data decreases with increasing k.)

Fig. 2. SLINK analysis of the data of Table 2. $\hat{\Delta}_1 = 0.07$.

Fig. 4. As for Fig. 3, but now the clusters are given/ at the lower level, 653.

Clarke's constant ratio rule (Clarke, 1957), also known as Luce's choice axiom (Luce, 1959), predicts that the confusions between the letters in each case (with and without the digits present) are the same, and indeed there are distinct similarities between the two dendrograms. Possibly the differences that do occur are due to differences in method and subjects; possibly they are due to a genuine departure from the constant ratio rule. In addition, the associations between the digits are very similar to those observed when the stimulus set comprises the digits alone (see Morgan, Chambers, & Morton, in press). By taking sections of the dendrograms at different levels, we obtain different sets of clusters, and some of these sets of clusters have been entered on the MDSCAL representations to follow.

First of all, MDSCAL representations in two dimensions were sought. The results from the best of several runs are given in Figs. 3 and 5. For Figs. 3 and 4, the MDSCAL configurations are the same.

For both matrices, the stress values are rather high, and it is strongly advisable to proceed into three dimensions. However, the two-dimensional configurations are far easier to construct than the three-dimensional ones and at the very least provide a useful platform on which to examine clusters resulting from sections of dendrograms at various levels. The three-dimensional results are given later.

We can now see how B(2) breaks down the straggly chains of SLINK: From Fig. 3, we see (at this level) that the CQTPDBVEGU cluster is more strongly linked than the AMNJKOLRYI cluster, a fact disguised by the section of the SLINK dendrograms at that level, but revealed by the entire dendrogram, for, of the two clusters, CQTPDBVEGU forms links sooner (at a lower level) and becomes a cluster sooner. The clusters of Fig. 4 throw light on the underlying structure of the

Fig. 3. Two-dimensional MDSCAL analysis of the data of Table 1. Stress = 18.4%. The clusters shown are those identified by SLINK and B(2) at the level 696. Here, and in Figs. 4 and 5, SLINK clusters are enclosed by a broken line and B(2) clusters by an unbroken line. For SLINK, $\Delta_1 = 0.12$; for B(2), $\Delta_1 = 0.06$.

Fig. 7. Three-dimensional MDSCAL analysis of data of Table 1. Projection onto plane of Principal Axes 1 and 2. Stress = 12.50%.

Fig. 6. Graph of stress against dimensionality of MDSCAL representation for data of Table 1 (X) and data of Table 2 (O).

1, 2, and 3, the first containing the most variability, and so on. In Figs. 7-10, the first principal axes are the "x-axes." The similarity of the respective two-dimensional and three-dimensional representations emphasizes the misleadingly poor stress values of the two-dimensional fits from the point of view of giving an *overall* picture, but perhaps the most striking feature of Figs. 7-10 is the strong correspondence of the principal axes of the two configurations.

Although MDSCAL can provide very useful representations, we must be careful not to be misled by them for, after all, the stress values are not zero. Any conclusion to which we might be led by MDSCAL should be checked against the raw data or by comparing the MDSCAL result with the SLINK dendrogram. This is a very cautious strategy, but a sensible one; for example, from Table 3, we have the similarities of F and X as 483 and of J and K as 222, yet in Fig. 3 F and X are further apart than J and K, a situation which is rectified when we move into three dimensions (see Figs. 7 and 8). As a further check, from the dendrogram of Fig. 1, we see that F and X enter the same cluster at a lower level than do J and K; entering SLINK and B(2) clusters on the MDSCAL representation can guard against such MDSCAL errors [see the B(2) clusters in Fig. 3].

DISCUSSION

We have seen how the three methods of cluster analysis can be used to illustrate large confusion

Fig. 8. Three-dimensional MDSCAL analysis of data of Table 1. Projection onto plane of Principal Axes 1 and 3.

matrices. The rationale behind the selection and use of the methods has been given, and ample references allow the interested psychologist to delve more deeply into the methods, if he so requires. MDSCAL and SLINK are, by now, fairly well known, and computerized algorithms for their execution are becoming widely accessible. What is not so common is their use *together*. B(2) is still very new and, while often needing MDSCAL to provide a suitable configuration of points on which to enter the

B(2) clusters at specific levels, it certainly forms a useful complement to SLINK and allows us to examine the extent of chaining.

We are now in a position to use these representations to examine the mechanisms of acoustic confusion for the two alphabets given.

An aspect of MDSCAL analysis, which has received no explanation in the above, is the possibility of obtaining an r -dimensional resulting configuration and then being

Fig. 9. Three-dimensional MDSCAL analysis of data of Table 2. Projection onto plane of Principal Axes 1 and 2. Stress = 12.34%.

Fig. 10. Three-dimensional MDSCAL analysis of data of Table 2. Projection onto plane of Principal Axes 1 and 3.

able to identify the r dimensions in terms of physical or phonetic features of the stimuli (see Carroll & Chang, 1970). An earlier study by Wilson (1963), which used factor analysis, tried to do just this. The data he used was that of Miller and Nicely (1957) on the confusions of 16 consonant phonemes, and an attempt was made to identify the resulting factors with phonetic features (see Jakobson, Fant, & Halle, 1963).

In MDSCAL terms, the value of r corresponding to r underlying physical dimensions might be expected to reveal itself in an "elbow" in the stress vs dimensions graph, such as Fig. 6, after which increase in dimensionality would result in relatively little decrease in stress. This procedure derives mainly from previous experience with factor analysis and is, in general, a dubious procedure, though Shepard (1963) used it to great effect in his analysis of Rothkopf's (1957) Morse code confusion data. Shepard deduced the two dimensions (dots and dashes and number of components to a signal) underlying the confusions. Townsend (1971) used MDSCAL to analyze the confusions of the letters of the alphabet, presented visually, and described his failure to detect an elbow in the stress-dimensions curve as follows: "Perhaps even more unsettling was failure of a set of dimensions clearly related to our intuitions concerning visual similarity to reveal itself. This result points up the difficulty inherent in using scaling techniques as a detective device unless the scaled dimensions are very elementary and obvious or unless they turn out to be equivalent to some previously hypothesized dimensions."

Such obvious underlying dimensions occur when the stimuli are artificially constructed from independent physical dimensions, and Carroll and Chang (1970) and Morgan (in press, a) have both been able to use MDSCAL to identify dimensions of confusion, in such a case. For Carroll and Chang, the dimensions were modulation waveform (sine wave vs square wave), modulation percentage, and modulation frequency; for Morgan, the dimensions were the four binary dimensions: harmonics,

pitch, formant number, and position of formants. An example of this use of MDSCAL is given later in the paper (see Table 6).

As mentioned above, no distinct elbows can be seen in Fig. 6. This, and the fact that both curves in Fig. 6 are so similar, must reflect the fact that acoustic confusion of stimuli such as these must involve many interacting dimensions, with none strongly dominating the others. The facts that when we add the digits to the letter stimulus set the curve of Fig. 6 remains roughly the same and that the principal axes of Figs. 9 and 10 bear such good correspondence with those of Figs. 7 and 8 imply that the digits use the same principles of confusion as the letters and simply slot into position without appreciably increasing the stress. It is in a situation such as this that MDSCAL demonstrates its superiority over a factor-analytic approach, which undoubtedly would have produced a large number of significant factors and a resulting picture which might well have been hard to appreciate. Factor analysis is at its most useful when a small number of significant factors arises.

Before we leave the topic of dimensionality, it is prudent to note a possible pitfall in interpreting MDSCAL results. Sometimes MDSCAL can warp a one-dimensional line to give a two-dimensional curve, the famous "horse-shoe" effect first obtained by Kendall (1971b), who explains how certain types of data can give rise to such an effect. Similarly, a two-dimensional surface can be warped into a three-dimensional curved surface. Such a curved surface has, in fact, been obtained by Levelt, Van de Geer, and Plomp (1966), who interpret the curve as meaningful, psychologically. Perhaps they are correct, but in such an instance one should critically return to the raw data to see if the curve could be merely due to MDSCAL warping.

So we have not been able to identify the MDSCAL dimensions with phonetic features. Can we possibly identify the *clusters* in terms of feature similarities? Johnson (1967) analyzed the Miller and Nicely (1957) confusion data on 16 English consonant phonemes,

using SLINK, and, at the level of five clusters, was able to identify the clusters with the Miller and Nicely distinctive feature description of the phonemes. The clusters were as follows: unvoiced stops (p, t, k), voiced stops (b, d, g), unvoiced fricatives (f, θ, s, ʃ), voiced fricatives (v, ð, ʒ, z), and voiced nasals (m, n). Two alternative distinctive feature descriptions are given by Wickelgren (1966), where he considers the confusions of a number of sets of consonant phonemes. Wickelgren (1965a) has considered the confusions of different vowel phonemes, and from Wickelgren (1965b) we find that confusions in short-term recall of letters and digits tend to have a vowel or consonant phoneme in common with the correct letter or digit. Wickelgren (1965c) remarks on these results as follows: "Thus, 'g' (/jē/) and 'j' (/jā/) were frequently confused in recall as were 's' (/ès/) and 'x' (/èks/), but 'k' (/kā/) and 'x' (/èks/) were not. At the same time 'h' (/ā/) and 'j' (/jā/) were confused even more often than 'j' (/jā/) and 'k' (/kā/) etc." Thus, his results differ somewhat from those above: interchanges of vowel phoneme or consonant phoneme position in a digram (i.e., a stimulus simply comprising a vowel phoneme and a consonant phoneme) stimulus, such as occur in the pairs (8.K), (8.J), (X.K), (H.J), (H.K), do not seem to result in any radical confusions.

Unlike in the study of Johnson (1967), it is not possible to make a simple interpretation of the clusters obtained in these studies. However, if we just consider digram stimuli, the clusters of Figs. 3, 4, and 5 suggest that discrimination operates *more* on the second phoneme than on the first. We see this as follows: Consider the confusions between the letters P, T, K, B, and V. From the Miller and Nicely results quoted above, we expect the consonant phonemes of P and K to be more similar than those of P and V. Yet from Table 3, we have a PK similarity of just 35 and a PV similarity of 125. That the first phoneme is playing a part in the discrimination, however, is reflected, for example, in the PT (similarity 786 from Table 3) and PB (similarity 237 from Table 3) confusions (again we are assuming the results of Miller and Nicely).

Table 5
Confusions of Vowel-Consonant Digrams in Short-Term Memory
(from Wickelgren, 1965c), Neglecting Omissions
and Incorrect Copyings

Response	Stimulus							
	fa	af	fo	of	na	an	no	on
fa	642	62	86	16	44	40	23	17
af	38	567	13	42	23	26	14	9
fo	51	20	622	35	17	19	39	29
of	31	79	75	587	34	27	43	38
na	28	26	15	7	531	32	25	14
an	23	31	16	18	59	586	26	34
no	19	13	35	27	43	18	484	41
on	17	14	40	23	31	44	51	571
Total	849	812	902	755	782	792	705	753

Table 4
Asymmetric Confusions from Tables 1 and 2 that Can
Be "Explained" By an Initial Masking Effect

Letter		Confusions	
Stimulus	Response	Table 1	Table 2
K	A	200	488
A	K	46	8
J	A	151	78
A	J	10	5
V	E	119	66
E	V	55	41
P	E	111	188
E	P	172	33
B	E	271	153
E	B	107	107
C	E	195	39
E	C	24	18
D	E	252	281
E	D	166	66
T	E	163	31
E	T	124	45
G	E	208	97
E	G	34	21

The above effect could be a *general* vowel effect, as P and K differ in the vowel phoneme, but that this is not so can be seen from the HS confusion in Table 3 (similarity, 194), where a vowel difference in the first phoneme still gives rise to a high confusion.

Further, M and N, which share the same terminal nasal consonant, while highly confusing with each other, are far less confusing with F, S, and X, with which they share the same initial vowel.

This apparent effect, of discrimination operating more on the second phoneme, which also explains the low GJ confusion in Fig. 3 (similarity of 20 from Table 3) and Fig. 5, could possibly be due to a masking of the start of each stimulus, caused by the particular levels of background noise employed in the two experiments. In such a discussion, we should also, of course, remember that the similarities of Table 3, for example, arise from a symmetrization of the actual confusions of Table 1. This initial masking effect can, in fact, explain some of the asymmetries in Tables 1 and 2, for were a digram stimulus, such as K, susceptible to initial masking, it would tend to be confused for A, while no initial masking effect could explain the A to K confusion. We might, therefore, expect higher K to A than A to K confusions, and this is exactly what occurs. All such pairs are given in Table 4. It is only in the (P,E) and (T,E) confusions, presented in Table 4, that the data cannot be explained by an initial masking effect.

It would be of interest to conduct a further study on these effects, using new words formed by changing the phoneme order in several of the digram letters of the alphabet to give a stimulus set in which each consonant-vowel pair occurs in each order, CV and VC.

A start in this direction has been made by Wickelgren (1965c), though using just two consonants, "f" and "n" and the two vowels "a" and "o". His confusion matrix is given in Table 5. One way of approaching this aspect of confusion is by using MDSCAL in search of "vowel," "consonant," and "position" dimensions as discussed at the beginning of this section. Thus the data of Table 5 were subjected to an MDSCAL analysis completely analogous to those on Tables 1 and 2. The minimum stress obtained in two dimensions was not very promising, namely, 15.20%, but the minimum stress obtained in three dimensions was the acceptable 6.0%. After rotation of axes to give the principal coordinates of the system, we obtained the configuration given in Table 6. Clearly, Dimensions 1, 2, and 3 can be identified with consonant, vowel, and position, respectively. This is a very simple examination of Wickelgren's data, but it tells us, at a glance, that these three factors play a part in the confusion of the given stimuli, and construction of the three-dimensional configuration would tell us very much more about the confusions, and again, at a glance.

We can also examine Wickelgren's data for an initial masking effect in short-term memory (though, of course, evidence for a masking effect in recognition would not imply that we also expect to find such an effect in short-term memory). We find that three times out of four the average estimated probability of confusion for pairs of stimuli is greater when the common element in a pair is the second element. The exception is (of, af), with average probability of confusion $(79/812) + (42/755) = 0.153$, and (fo, fa), with average probability of confusion $(86/902) + (51/849) = 0.155$.

The high EU confusions imply that discrimination operates more on the first than on the second formant (or on their height rather than place), a result which has already been noted in the high confusion of Digits 2 and 3 and also their confusability in memory (see Morgan, Chambers, & Morton, in press).

Finally, the pattern of confusion for the digits, as shown in Fig. 5, is very similar to that observed in Morgan, Chambers, and Morton (in press), adding weight to their conclusion that there is a strong systematic confusion of digits (this is in contradiction with previous studies on the confusion of digits). The similarity of these two analyses of digit confusions and also the similarity of the letter confusions in Figs. 4 and 5 emphasize the fact that a *rough* constant ratio rule is applicable to acoustic confusion experiments (but see Morgan, in press, b), and we might then feel justified in using the dispositions of stimuli in Figs. 9 and 10, for example, to choose subalphabets with desirable confusion characteristics.

CONCLUSION

It is to be hoped that the above results demonstrate the utility of such cluster analysis methods in illustrating

Table 6
Three-Dimensional MDSCAL Representation of the Confusions Between Wickelgren's (1965c) Digram Stimuli, Referred to the Principal Axes of the Configuration

Stimulus	Principal Axis		
	First (Consonant)	Second (Vowel)	Third (Position)
fa	0.594	-0.015	-0.633
af	0.935	-0.647	0.166
fo	0.420	0.997	-0.131
of	0.526	0.333	0.613
na	-0.440	-0.906	-0.416
an	-0.445	-0.575	0.455
no	-0.719	0.482	-0.487
on	-0.870	0.332	0.433

large, unwieldy banks of data, and that the particular illustrations given will be of use in helping to design subalphabets with specific characteristics. Thus, from Fig. 3, for example, we see that Conrad and Hull's (1964) choice of JKN, CDFHLNQYZ as low-confusing clusters and FSX, BCDGMNPTV as high-confusing clusters was not a bad one (in relation to the entire alphabet) but one which could certainly be improved, particularly with respect to JKN. Similarly, Conrad's (1964) choice of BCPTV and FMNSX as two clusters with high within-cluster confusion and low between-cluster confusion might have been improved by, say, BCPTV and FZHSX.

REFERENCES

- Carroll, J. D., & Chang, J. J. Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckert-Young" decomposition. *Psychometrika*, 1970, 35, 283-319.
- Clarke, F. R. Constant-ratio rule for confusion matrices in speech communication. *Journal of the Acoustical Society of America*, 1957, 29, 715-720.
- Conrad, R. Acoustic confusions in immediate memory. *British Journal of Psychology*, 1964, 55, 75-84.
- Conrad, R., & Hull, A. J. Information, acoustic confusion and memory span. *British Journal of Psychology*, 1964, 55, 429-432.
- Cormack, R. M. A review of classification. *Journal of the Royal Statistical Society A*, 1971, 134, 321-367.
- Fillenbaum, S., & Rapoport, A. *Structures in the subjective lexicon*. New York: Academic Press, 1971.
- Holloway, J. D., & Jardine, N. Two approaches to zoogeography: A study based on the distributions of butterflies, birds and bats in the Indo-Australian area. *Proceedings of the Linnean Society of London*, 1968, 179, 153-188.
- Hull, A. J. A letter-digit listening matrix. *British Journal of Psychology*, in press.
- Jakobson, R., Fant, C. G. M., & Halle, M. Preliminaries to speech analysis: The distinctive features and their correlates. Cambridge: M.I.T. Press, 1963.
- Jardine, N. Towards a general theory of clustering. *Biometrics*, 1969, 25, 609-610.
- Jardine, N., & Sibson, R. The construction of hierarchic and non-hierarchic classifications. *The Computer Journal*, 1968, 11, 177-184.
- Jardine, N., & Sibson, R. *Mathematical taxonomy*. London: Wiley, 1971.

- Johnson, S. C. Hierarchical clustering schemes. *Psychometrika*, 1967, 32, 241-254.
- Kendall, D. G. Construction of maps from "Odd bits of information." *Nature*, 1971a, 231, 158-159.
- Kendall, D. G. Seriation from abundance matrices. In F. R. Hodson, D. G. Kendall, and P. Tautu (Eds.), *Mathematics in the archaeological and historical sciences*. Edinburgh: University Press, 1971b.
- Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 1964a, 29, 1-27.
- Kruskal, J. B. Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 1964b, 29, 115-129.
- Levelt, W. J. M., Van der Geer, J. P., & Plomp, R. Triadic comparisons of musical intervals. *British Journal of Mathematical & Statistical Psychology*, 1966, 19, 163-179.
- Luce, R. D. *Individual choice behavior*. New York: Wiley, 1959.
- Miller, G. A., & Nicely, P. E. An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, 1955, 27, 338-352.
- Morgan, B. J. T. Statistical annex to: Perceiving sounds by their visible spectra by M. Woodhead, A. Carpenter, and J. C. Webster. *British Journal of Psychology*, in press, a.
- Morgan, B. J. T. A statistical test of Luce's choice axiom. *Journal of Mathematical Psychology*, in press, b.
- Morgan, B. J. T., Chambers, S. M., & Morton, J. Acoustic confusion of digits in memory and recall. *Perception & Psychophysics*, in press.
- Rothkopf, E. Z. A measure of stimulus similarity and errors in some paired-associate learning tasks. *Journal of Experimental Psychology*, 1957, 53, 94-101.
- Shepard, R. N. Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, 1957, 22, 325-345.
- Shepard, R. N. Analysis of proximities as a technique for the study of information processing in man. *Human Factors*, 1963, 5, 33-48.
- Sibson, R. Order invariant methods for data analysis. *Journal of the Royal Statistical Society, B*, 1972, in press.
- Townsend, J. T. Theoretical analysis of an alphabetic confusion matrix. *Perception & Psychophysics*, 1971, 9, 40-50.
- Van der Kamp, L. J. Th., & Pols, L. C. W. Perceptual analysis from confusions between vowels. *Acta Psychologica*, 1971, 35, 64-77.
- Wickelgren, W. A. Distinctive features and errors in short-term memory for English vowels. *Journal of the Acoustical Society of America*, 1965a, 38, 583-588.
- Wickelgren, W. A. Acoustic similarity and intrusion errors in short-term memory. *Journal of Experimental Psychology*, 1965b, 70, 102-108.
- Wickelgren, W. A. Similarity and intrusions in short-term memory for consonant-vowel digrams. *Quarterly Journal of Experimental Psychology*, 1965c, 17, 241-246.
- Wickelgren, W. A. Distinctive features and errors in short-term memory for English consonants. *Journal of the Acoustical Society of America*, 1966, 39, 388-398.
- Wilson, K. V. Multidimensional analyses of confusions of English consonants. *American Journal of Psychology*, 1963, 76, 89-95.

(Received for publication June 19, 1972;
revision received September 13, 1972.)