Perception & Psychophysics
1979, Vol. 26 (1), 1-19

On the prediction of confusion matrices from similarity judgments

DAVID J. GETTY, JOHN A. SWETS, JOEL B. SWETS, and DAVID M. GREEN
Bolt Beranek and Newman Inc., 50 Moulton Street, Cambridge, Massachusetts 02138

Three observers viewed visual representations of eight complex sounds in both a pairwise similarity-judgment task and an identification task. A multidimensional scaling procedure applied to the similarity judgments yielded a three-dimensional perceptual space and the relative positions of the stimuli in that space. A probabilistic decision model based on weighted interstimulus distances served to predict well the confusion matrices of the identification task. Three conditions of the identification task, calling for identification of different subsets of the eight stimuli, led the observers to vary the weights they placed on the dimensions; they apparently adjusted the weights to maximize the percent correct identification. An additional group of 14 subjects, participating only in the similarity-judgment task, manifested the same three dimensions as the observers (corresponding to the locus of low-frequency energy, the locus of midfrequency energy, and visual contrast), and also a fourth dimension (corresponding to the periodicity, or waxing and waning, of the sound). Although not evident in the scaling analysis for the three observers, our utilization of the additional dimension increased significantly the variance accounted for in their identification responses. The overall accuracy of the predictions from a perceptual space to identification responses supplies a substantial validation of the use of multidimensional scaling procedures to reveal perceptual structure in demonstrating the ability of that structure to account for behavior in an independent task. The empirical success of this approach, furthermore, suggests a relatively simple and practical means of predicting, and possibly enhancing, identification performance for a given set of visual or auditory stimuli.

A reasonably complete account of the process by which humans are able to identify complex auditory or visual stimuli must address two issues: (1) the nature of the psychological representation of complex stimuli, and (2) the nature of the decision processes that act upon the internal representations to yield an identification response. While there has been substantial research directed at the representation and decision processes separately, there has been much less effort directed towards understanding the integration of these processes in the identification of complex stimuli. Our major concern in this paper is with the relationship between perceptual representation and identification performance. The question is, given specific assumptions about the structure of the perceptual space, how well can we account for the pattern of responses observed in an identification task?

Our approach to the problem involves two parts: (1) the derivation of a multidimensional perceptual space for a set of complex stimuli from the application of a multidimensional scaling (MDS) procedure to judgments of stimulus similarity, and (2) the use of a probabilistic decision model to predict the matrix of identification confusions from the geometric structure of the derived perceptual space. We are interested in the validation of the MDS procedure that would be supplied by a demonstration that the MDS-derived perceptual space can be used to predict behavior in an independent task. And we are interested in the possibility that the MDS procedure can be a substantial aid in understanding, and predicting behavior in, the fundamental task of stimulus identification.

We discuss first the rationale for, and assumptions made in, inferring a psychological space using MDS techniques. Then, in the next section, we describe our decision model for predicting identification confusions.

THE MULTIDIMENSIONAL PERCEPTUAL SPACE

Much of the work on the representation of complex auditory or visual stimuli suggests that perception is based on an analysis of the stimulus patterns along a number of psychological dimensions or features. In information-processing models, this processing is often referred to as "feature extraction" and is thought to reflect a selective reduction of

This research was supported by a contract with the Engineering Psychology Programs, Office of Naval Research. We thank Barbara Freeman for her able programming assistance.

0031-5117/79/070001-19$02.15/0

information whereby perceptually important features are extracted from the pattern while other information is lost.

We may conceive of these dimensions as forming a multidimensional perceptual space in which each stimulus is represented as a point. This space, of course, is not directly observable. Both the set of dimensions comprising the space and the loci of the stimuli within the space must be inferred by indirect methods. Extensive development in recent years has led to the emergence of multidimensional scaling as an important method for deriving a representation of the perceptual space (e.g., Romney, Shepard, & Nerlove, 1972; Shepard, Romney, & Nerlove, 1972). MDS procedures are designed to decompose a matrix of pairwise similarity judgments on a set of complex stimuli into a metric space of some (investigator-specified) number of orthogonal dimensions. Each stimulus is defined as a point in the space such that, ideally, the distances between pairs of stimuli in the space are monotonically related (inversely) to the degrees of judged similarity of the pairs.

The set of abstracted dimensions and the relative loci of the stimuli within the space may be interpreted to reflect the structure of the psychological space. This interpretation involves several assumptions. Most MDS procedures assume that the measure of the underlying psychological space is a member of the family of power metrics.[1] This family includes the Euclidian and city-block metrics. The power metrics carry with them three important properties: (1) decomposability, (2) interdimensional additivity, and (3) intradimensional subtractivity (Tversky & Krantz, 1970). Decomposability means simply that the distance between any two points in the space is a function of dimensionwise contributions. Interdimensional additivity asserts that distance is a function of the sum of the dimensionwise contributions. Intradimensional subtractivity asserts that each dimensionwise contribution is the absolute value of the difference between the two points on that dimension. An intuitive implication of these properties is that, for a given pair of stimuli, the contribution to distance made by the values of the two stimuli on one dimension is independent of the values on all other dimensions.

Having obtained an abstract multidimensional solution, an investigator may attempt to relate the derived psychological dimensions to the known physical structure of the stimuli. Success in identifying the psychophysical functions relating psychological to physical dimensions is typically measured by a high correlation between values on a psychological dimension and values on the candidate physical measure, across stimuli.

MDS procedures have been used successfully to identify psychological dimensions underlying the perception of speech sounds (e.g., Klein, Plomp, & Pols, 1970; Pols, Van der Kamp, & Plomp, 1969; Shaw, 1975; Shepard, 1972), complex, nonspeech sounds (e.g., Howard, 1977; Howard & Silverman, 1976; Miller & Carterette, 1975; Morgan, Woodhead, & Webster, 1976; Plomp & Steeneken, 1969), and complex visual patterns (e.g., Pachella & Somers, in press; Shepard & Chipman, 1970; Stenson, 1968; Hardzinski & Pachella, Note 1). In this context, success has usually meant that the derived multidimensional space accounts for a large proportion of the variability in the similarity judgments and that the revealed identity of the psychological dimensions is intuitively reasonable.

We suggest here that meeting either or both of these criteria does not provide strong evidence to support the validity of the derived representation. What is desirable is a demonstration that the perceptual space derived from similarity judgments in one task can then be used to predict behavior in some other, independent task. Our present experiments provide an example of one such test, in that we use the MDS-derived space to predict performance in various identification tasks.

THE IDENTIFICATION MODEL

We present here a decision model intended to predict the confusion matrix for a set of m stimuli in an identification task, on the basis of a multidimensional perceptual space. In the simplest case, the perceptual space is that revealed by application of an MDS procedure to judgments of similarity, and we shall confine our presentation in this section to that case. We note, for later reference, that the decision model can accept additional dimensions, as suggested by other evidence.

We take as our starting point the set of spatial coordinates, $\psi_{i,k}$, for each of m stimuli, s_i $(1 \leq i \leq m)$, on each of n dimensions, d_k $(1 \leq k \leq n)$, as provided by the MDS procedure. Though the model can be used with any MDS procedure that yields such a set of spatial coordinates, the data analyses presented in this paper are based on the INDSCAL procedure (Carroll, 1972; Carroll & Chang, 1970; Carroll & Wish, 1973). INDSCAL assumes that the judged similarity between any pair of stimuli is a (decreasing) linear function of the Euclidean interstimulus distance in the underlying perceptual space. INDSCAL differs from most other MDS procedures in that it yields not only the spatial configuration of the set of m stimuli in the n-dimensional "group stimulus space," but also a vector of weights for each observer that reflects the relative importance or salience of each dimension for that observer. The effect of these salience weights for a given observer is to weight differentially the contribution of each dimension in determining interstimulus distance. Specifically, the distance between stimulus S_i and stimulus S_j for an observer with salience weights w_k ($1 \leqslant k \leqslant n$) is given by

$$D_{i,j} = \left[\sum_k w_k(\psi_{i,k} - \psi_{j,k})^2\right]^{1/2}. \quad (1)$$

Our model will also assume the weighted Euclidean distance metric given in Equation 1. We will not assume, however, that the particular set of salience weights determined for each observer by INDSCAL in the similarity-judgment task necessarily applies to the identification tasks. In fact, we will show later that, within observers, the set of salience weights changes in predictable ways across different conditions of the identification task. Accordingly, the salience weights for each observer are treated in the model as parameters whose values are to be estimated from the confusion data. We assume further that the salience weights are all positive ($w_k \geqslant 0$) and sum to 1,

$$\sum_k w_k = 1.$$

The latter constraint is simply a normalizing convention which reflects the fact that only the relative magnitudes of the weights are meaningful in the model.

Having defined the set of interstimulus distances $D_{i,j}$, the next step is to relate these distances to interstimulus confusability in the identification task. On intuitive grounds, confusability should be some monotone decreasing function of interstimulus distance. We define a set of confusion weights $C_{i,j}$, assuming that confusability between stimulus S_i and stimulus S_j is given by

$$C_{i,j} = \exp(-aD_{i,j}), \quad (2)$$

where a is a sensitivity parameter, greater than 0. As a decreases towards 0, overall stimulus confusability increases; as a becomes larger, overall confusability decreases.

Several consequences of this relation are worth noting: (1) with both a and $D_{i,j}$ bounded below by 0, $C_{i,j}$ is bounded between 0 and 1; (2) since $D_{i,j}$ is a distance measure, $D_{i,j} = D_{j,i}$, and therefore $C_{i,j} = C_{j,i}$; and (3) since $D_{ii} = 0$ for all i, $C_{ii} = 1$ for all i. The choice of this particular function, from several considered, was dictated by its clear superiority in accounting for our confusion data in preliminary analyses. It is the same assumption used successfully by Shepard (1957, 1958a, 1958b) in his work on stimulus and response generalization.

Finally, the conditional probability of giving the response assigned to stimulus S_j when stimulus S_i was presented is assumed to be the confusability of S_j with S_i relative to the summed confusability of all stimuli with S_i:

$$\Pr(R_j \mid S_i) = \frac{C_{i,j}}{\sum_k C_{i,k}}. \quad (3)$$

Equation 3 is essentially Luce's (1963) choice model, with the added assumption that there are no differential response biases. While it would be a simple matter to include measures of response bias in the model, we have chosen to exclude them here for reasons of simplicity (fewer parameters to estimate) and because we have no reason to expect strong response biases. In our tasks, the a priori presentation probabilities (known to the observers) were equal across stimuli, the response set was homogeneous, and there were no differential payoffs.

We are also implicitly asserting in Equation 3 that the set of responses assigned to stimuli are sufficiently distinguishable that response confusions are negligible, and they are therefore not incorporated into the model.

It is relatively common, in identification tasks, to find asymmetries in the confusion matrix about the main diagonal; that is, it is often true that $\Pr(R_j \mid S_i) \neq \Pr(R_i \mid S_j)$. One well-known source of this asymmetry is response bias. Of interest here is the observation that there is a second possible source of confusion asymmetries, one which arises from the decision rule itself. This can most easily be seen in an example, shown in Figure 1, in which three stimuli are embedded in a two-dimensional space. As drawn,

$$D_{1,3} > D_{1,2} > D_{2,3} > D_{i,i} = 0 \text{ (for all i)},$$

and therefore,

$$C_{1,3} < C_{1,2} < C_{2,3} < C_{i,i} = 1 \text{ (for all i)}.$$

Calculating $\Pr(R_1 \mid S_2)$ and $\Pr(R_2 \mid S_1)$, we find

$$\Pr(R_1 \mid S_2) = \frac{C_{2,1}}{C_{2,1} + C_{2,2} + C_{2,3}} = \frac{C_{1,2}}{1 + C_{1,2} + C_{2,3}}$$

and

$$\Pr(R_2 \mid S_1) = \frac{C_{1,2}}{C_{1,1} + C_{1,2} + C_{1,3}} = \frac{C_{1,2}}{1 + C_{1,2} + C_{1,3}}.$$

Since $C_{1,3} < C_{2,3}$, we may conclude that $\Pr(R_2 \mid S_1 > \Pr(R_1 \mid S_2)$, thus demonstrating asymmetry in the confusion matrix about the main diagonal.

From this example, we can see that the confusion asymmetry between two stimuli, S_i and S_j, arises

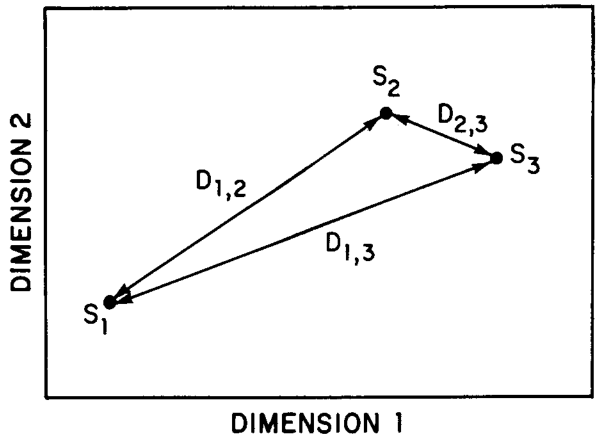


Figure 1. Example of a stimulus configuration for which an asymmetric confusion matrix would be obtained (see text).

because the two response probabilities, $\Pr(R_j \mid S_i)$ and $\Pr(R_i \mid S_j)$, are determined by the magnitude of the shared confusion weight, C_{ij}, relative to a sum of confusion weights, where the sum is determined from the "point of view" of the stimulus, S_i or S_j. These two sums may differ substantially in magnitude, depending upon the geometric configuration of the stimuli in the perceptual space.

TWO TYPES OF IDENTIFICATION TASK

The primary test of the ability of the MDS-derived perceptual space and the associated decision model to predict identification behavior involves a straightforward identification task in which m responses are paired (one-to-one) with the m stimuli presented. We call this a "complete" identification task.

We also consider a "partial" identification task in which fewer than m responses are available for use with the m stimuli presented. This task corresponds to a situation in which some subset of the total number of stimuli present is of special interest—the members of this subset constituting "signals"—while the remaining stimuli are regarded as "noise" and as not requiring identification. We have referred to this class of task elsewhere as a "detection-and-identification" task (Swets, Green, Getty, & Swets, 1978). The observer is asked first to make a detection response ("signal" or "noise") and then to choose one of the available identification responses—the one corresponding to the one of the signals most likely to be present.

The partial identification task employed here had three conditions, with a different subset of the m stimuli defined as signals in each condition. This variation across conditions reflects the practical fact that the subset of stimuli that is of special interest varies from one situation to another. The listener to degraded speech may want to concentrate at some time on distinguishing between just 2 particular phonemes, rather than among 30 or so. The sonar observer may wish to distinguish among ships and ignore variations in ocean depth, or vice versa. In the present instance, the variation across conditions provides a test of our decision model's ability to deal with changes in the salience weights, which may reflect changes in the relative usefulness of the various perceptual dimensions as the set of signals is changed.

METHOD

Stimuli

Our stimuli consisted of visual representations of a set of eight underwater sounds, originally selected by Howard (1977) to represent a range of confusable natural and mechanically produced sounds. They were referred to by Howard as (1) sheet cavitation (SC), (2) biologics (BI), (3) compressed cavitation (CC), (4) torpedo (TO), (5) diesel engine (DE), (6) rain squall (RS), (7) steam noise (SN), and (8) flutter (FL). Their long-term energy spectra are shown in Figure 2.

Our visual representations displayed the spectra as frequency (horizontal axis) vs. time (vertical axis) vs. energy (darkness—the greater the energy, the darker the trace). We introduced periodicity as an additional physical dimension by sinusoidally varying the average darkness of the signal profile in the temporal (vertical) direction. We grouped (1) SC and (2) BI with (7) SN and (8) FL, giving all four relatively low-frequency periodicity, and gave the remaining four relatively high-frequency periodicity. Specifically, the cycles per stimulus were 15, 16, 17, and 18 for Stimuli 1, 2, 7, and 8, respectively, and 21, 22, 23, and 24 for Stimuli 3, 4, 5, and 6, respectively. The resulting visual patterns are shown in Figure 3.

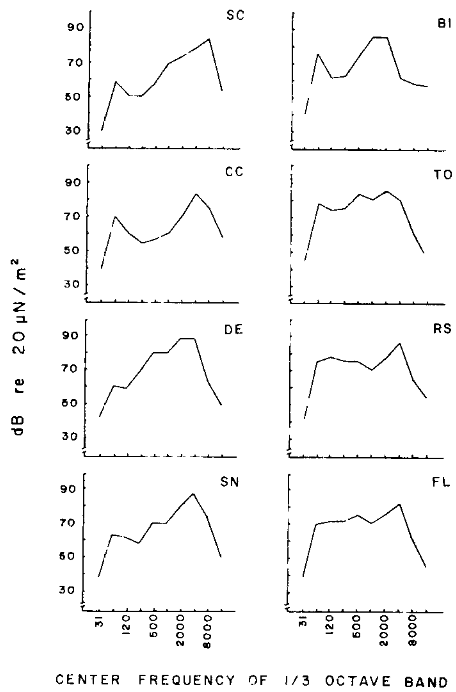


Figure 2. Long-term spectra of eight underwater sounds (from Howard, 1977).

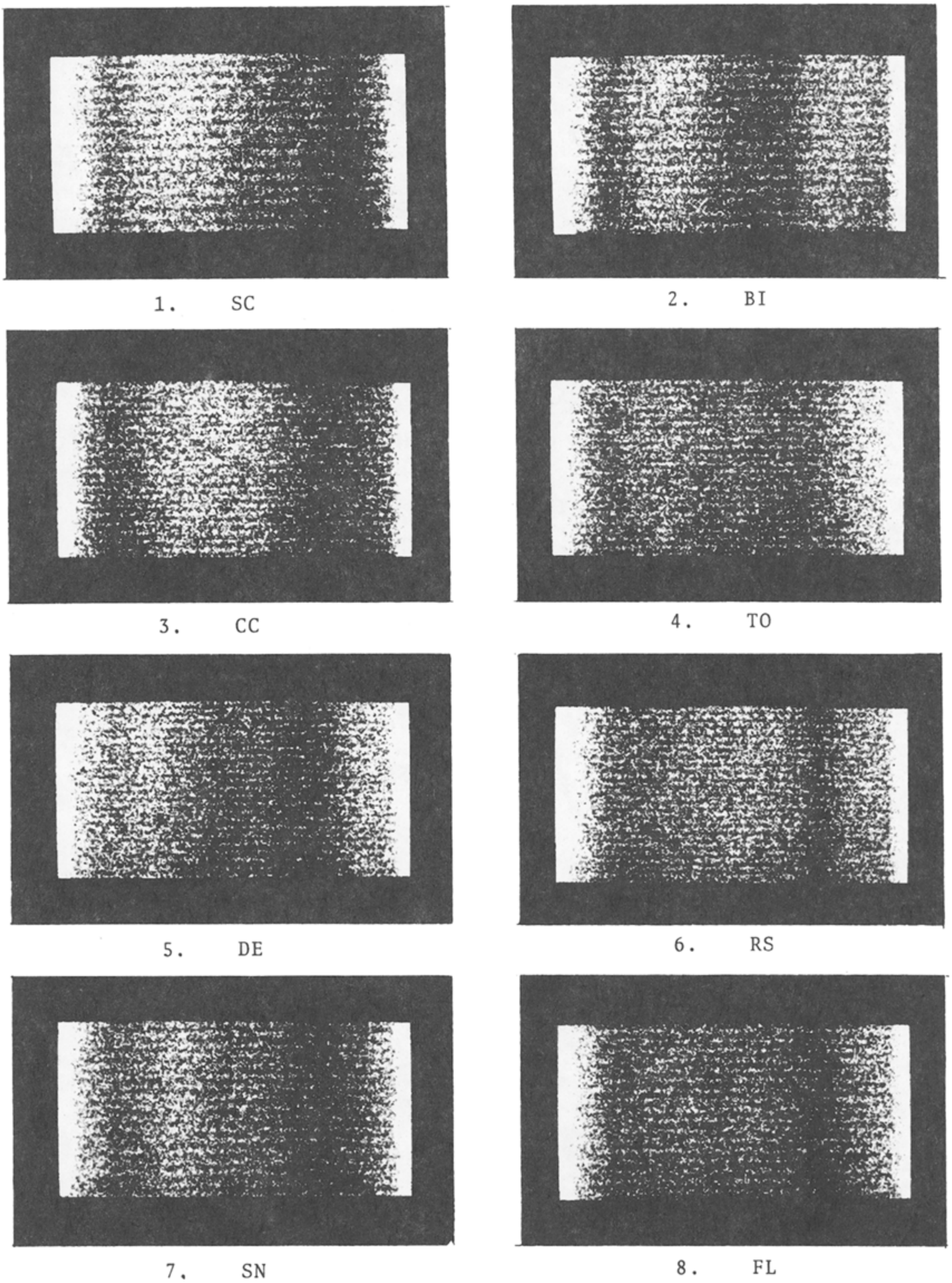


Figure 3. Visual representations of the eight underwater sounds, photographed from the display monitor.

The stimuli were constructed on a COMTAL Model 8000-SA image-processing system, driven by a DEC PDP-11/34 minicomputer, and displayed in an area 24 cm wide × 12 cm high on a CONRAC 17-in. (43-cm) SNA television monitor. As is apparent in Figure 3, we added a background of random noise to each stimulus pattern. The noise consisted of a 256 by 128 matrix of elements, each having an independent gray value sampled from a Gaussian distribution with mean 128 units and standard deviation 15 units on the 256-unit gray scale of the COMTAL.

Each stimulus pattern was constructed by subtracting from the noise background, sampled anew on each trial, the horizontal brightness profile corresponding to its long-term spectrum. Thus, increasing energy in the spectrum resulted in a darker trace. The spectral profiles of all eight signals were scaled to have the same space-average darkness of 20 gray units below the mean gray value of the noise background. In addition, the darkness of each point in a signal's profile was sinusoidally varied in the vertical dimension. For the similarity-judgment task, the peak-to-peak brightness variation was 200% about the steady-state value, resulting in high-contrast images similar to those shown in Figure 3. For the various identification tasks, the peak-to-peak brightness variation was reduced to 60% about the steady-state value, thereby reducing the image contrast considerably.

For all tasks, the brightness and contrast controls on the CONRAC monitor were adjusted such that the middle gray (128 units) had a luminance of about 62 cd/m^2, and full white (255 units) had a luminance of about 308 cd/m^2.

Apparatus

In the various identification tasks, three observers sat at individual video computer terminals (Lear Siegler ADM-3A) approximately 2 m from the stimulus-display screen, whose center was about 1.1 m above the floor. Ambient room lighting was maintained at a dim level.

All experimental events in the identification tasks (stimulus display, response recording, and trial timing) were controlled by the PDP-11 computer.

For the similarity judgments, the stimuli displayed on the television monitor were photographed and presented to groups of four judges by means of 35-mm slides. The projection screen was approximately 5 m from the judges; each stimulus subtended a visual angle of about 3.5°

Subjects

The three observers in the identification tasks were research assistants on BBN's staff, each with a bachelor's degree in psychology, including one of the experimenters (J.B.S.). These three observers served in the similarity-judgment task after they had completed the partial identification task and before they began the complete identification task.

Additionally, serving as judges in the similarity task were 20 members of BBN's technical staff: 3 females and 17 males, about one-half with a master's degree and one-half with a doctor's degree in psychology or information science, ranging in age from 30 to 50 years.

Procedure

Similarity-judgment task. After seeing the eight stimuli presented successively, twice, for 15 sec each, the subjects rated similarity on a 10-point scale for the 28 possible pairs, with members of each pair presented side by side for 15 sec; the following response interval was also 15 sec. To assess response consistency, the 28 pairs were then presented a second time, with left-right positions reversed, and in a different random order—to the 20 additional judges—and second and third times to the three primary observers.

Complete identification task. Each trial began by blanking of the COMTAL screen, followed 2.5 sec later by a 2.0-sec low-contrast display of one of the eight visual stimuli. Each observer then made a self-paced identification response on the terminal keyboard, pressing one of eight keys labeled with the digits 1 to 8 (typing errors could be corrected with an "erase" key). Observers could make reference to a sheet on which were arranged labeled, high-contrast Polaroid photographs of the eight stimuli—again, similar to those shown in Figure 3. As each observer responded, the number of the presented signal was displayed on his/her terminal's screen. When all three had responded, the stimulus image was redisplayed along with the correct answer for about 2 sec.

The sequence of stimuli was determined completely at random. Fifty trials were presented in a block, and three blocks were presented in a 1-h session.

Partial identification task. In this task, four of the eight stimuli were designated as "signal" and allowed as responses, while the remaining four were designated as "noise" and not allowed as responses. A different set of four stimuli was designated as the "signal" set in each of three conditions. The procedure for the partial identification task was generally the same as that described above for the complete identification task except that, for reasons not germane to the present interest, on each trial a stimulus pattern was presented in five sequential stages. At each stage, a successive fifth of the pattern was revealed, pushing down the display of earlier stages, and the observer made both a detection and an identification response. Our present analyses use only the identification responses from the last (fifth) stage of each trial when the complete display of the stimulus was in view. Observers were provided feedback—either the signal number (when one of the four "signals" was presented) or the word "noise" (when one of the four "noise" stimuli was presented)—at the end of each trial.

DERIVATION OF THE PERCEPTUAL SPACE: THE SIMILARITY-JUDGMENT TASK

The third (last) set of 28 similarity judgments for each of the three observers, and the second (last) set of 28 similarity judgments for 14 of the 20 additional judges, were submitted to INDSCAL analysis. For each of the three observers, the correlation between the first and second sets, and between the second and third sets, was greater than .50; but the correlation of the first and third sets was less than .50 for two of the three observers. The 14 of the 20 additional judges whose data were used in the INDSCAL analysis were those whose judgments on Trials 11 to 28 of the first set correlated higher than .50 with the same pairs in the second set. We therefore used a set of judgments based on as much experience as was available to the various subjects, and used only the data of the additional judges who were behaving rather consistently.

The Perceptual Space of the Three Primary Observers

For the three observers, the proportion of variance accounted for by the INDSCAL solutions with one to four dimensions is given by the open circles in Figure 4. The psychological coordinates of the eight stimuli in the three-dimensional perceptual space for those observers are listed as ψ_1 to ψ_3 in Table 1. Also listed in the table are measures along three physical dimensions of the stimuli, ϕ_1 to ϕ_3, that were found to correlate highly with the three psychological dimensions.

The first physical dimension, ϕ_1, is termed "low-frequency energy," and represents the amount of energy in the second and third filter bands as shown in Figure 2 (the 1/3-octave bands with center frequencies of 63 and 125 Hz). The second physical dimension, ϕ_2, is termed "midfrequency energy"—the energy in the fifth and sixth bands of Figure 2,

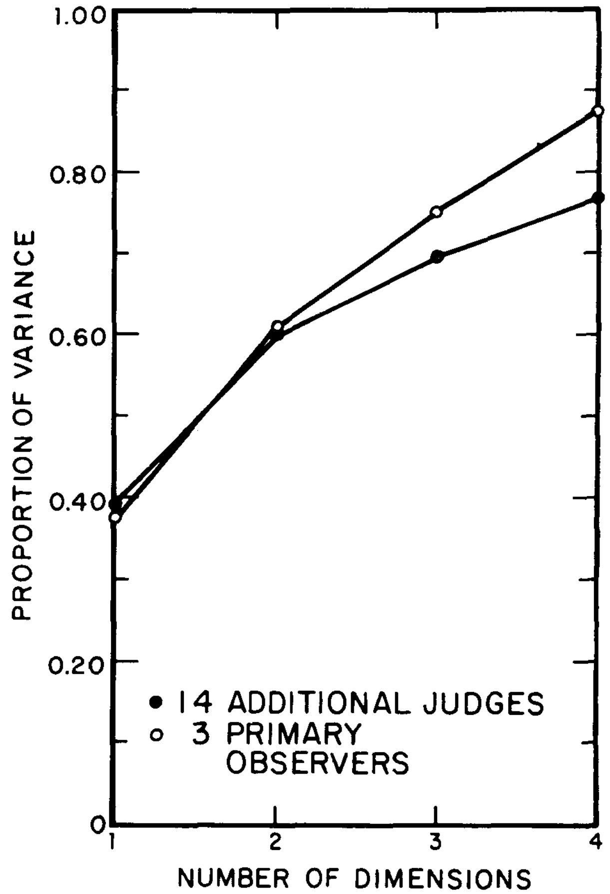


Figure 4. The proportion of variance accounted for by the INDSCAL solution with varying numbers of dimensions, for both the three primary observers and the 14 additional judges.

i.e., the bands with center frequencies of 500 and 1,000 Hz. The physical dimension denoted ϕ_3 gives a measure of "contrast" and, more specifically, of the depth of the "primary white trough"—defined as the average deviation between the points on the energy profile (Figure 2) and a line connecting the two local maxima bounding the primary trough. (The stimuli are not ranked in the table on the physical dimensions in exactly the way they would have been if based directly on Figure 2, because, as mentioned earlier, the signal profiles of Figure 2 were normalized in our visual transformations to have the same space-average darkness.)

The physical dimensions were selected to represent the dimensions we experimenters saw in the stimuli, and, as can be seen in Table 2, to correlate highly with the psychological dimensions revealed by the INDSCAL perceptual scaling analysis as applied to the three observers. There is a clear physical correlate—a product-moment r in the vicinity of .90—for each of the three perceptual dimensions. Correlations greater than .83 have a probability less than .01 for a two-sided test. The remaining correlations fall well below .71, which has a corresponding probability of .05 for a two-sided test.

We point out that "periodicity"—the physical dimension that we specifically introduced in the stimuli in a controlled way—did not emerge as a psychological dimension for the three observers; specifically, periodicity showed a low correlation with the fourth dimension examined for them. This result occurred despite the fact that periodicity was salient for the three observers, as we shall see, in the identification tasks. Indeed, these observers mentioned that while they were aware of the periodicity variable in the judgment of similarity, they did not find it salient in that task.

The Perceptual Space of the 14 Additional Judges

For the 14 additional judges whose data were submitted to INDSCAL analysis, the proportion of variance accounted for by the solutions with one to four dimensions is given by the closed circles in Figure 4. The weights of the eight stimuli in the four-dimensional solution are listed as ψ_1 to ψ_4 in Table 3. Also listed in the table are the three physical dimensions repeated from Table 1, and the physical dimension of periodicity, ϕ_4.

Table 4 shows the product-moment correlations of the four physical measures, ϕ_1 to ϕ_4, with the four

Table 1
Physical Coordinates for Each of the Eight Stimulus Patterns and Psychological Coordinates Yielded by the Three Experienced Observers

	Psychological Dimensions			Physical Dimensions		
Stimulus	ψ_1	ψ_2	ψ_3	ϕ_1	ϕ_2	ϕ_3
1. SC	−.405	.427	.229	90	104	8.8
2. BI	.441	−.143	.675	103	121	10.5
3. CC	−.552	−.209	.243	104	93	12.8
4. TO	.380	−.083	−.344	107	116	4.5
5. DE	.317	.597	−.213	87	118	4.5
6. RS	−.154	−.521	−.233	110	105	5.5
7. SN	−.198	.207	.097	96	108	6.0
8. FL	.171	−.276	−.454	107	110	4.5

Table 2
Product-Moment Correlations Between Each of the Psychological and Physical Dimensions for the Three Experienced Observers

	Psychological Dimensions			Physical Dimensions		
	ψ_1	ψ_2	ψ_3	ϕ_1	ϕ_2	ϕ_3
ψ_1		.030	−.189	.095	.947*	−.491
ψ_2			.118	−.971*	.246	−.113
ψ_3				−.215	−.037	.827*
ϕ_1					−.142	.007
ϕ_2						−.475
ϕ_3						

Note—Each r based on n = 8 stimuli. *$p < .01$.

Table 3
Physical Coordinates for Each of the Eight Stimulus Patterns and Psychological Coordinates Yielded by 14 Additional Judges

	Psychological Dimensions				Physical Dimensions			
Stimulus	ψ_1	ψ_2	ψ_3	ψ_4	ϕ_1	ϕ_2	ϕ_3	ϕ_3
1. SC	−.546	.261	−.369	.385	90	104	8.8	15
2. BI	−.083	.403	.616	.575	103	121	10.5	16
3. CC	.262	.595	−.470	−.182	104	93	12.8	21
4. TO	.004	−.264	.358	−.440	107	116	4.5	22
5. DE	−.514	−.098	.276	−.227	87	118	4.5	23
6. RS	.515	−.177	−.130	−.392	110	105	5.5	24
7. SN	.058	−.210	−.170	.300	96	108	6.0	17
8. FL	.305	−.511	−.112	−.019	107	110	4.5	18

Table 4
Product-Moment Correlations Between Each of the Psychological and Physical Dimensions for the 14 Additional Judges

	Psychological Dimensions				Physical Dimensions			
	ψ_1	ψ_2	ψ_3	ψ_4	ϕ_1	ϕ_2	ϕ_3	ϕ_4
ψ_1		−.230	−.224	−.381	.890*	−.371	−.010	.339
ψ_2			−.133	.378	−.182	−.347	.948*	−.242
ψ_3				.080	.052	.934*	−.260	.074
ψ_4					−.366	.168	.434	−.927*
ϕ_1						−.142	.007	.280
ϕ_2							−.475	−.053
ϕ_3								−.358
ϕ_4								

Note—Each r based on n = 8 stimuli. $^*p < .01$.

psychological dimensions, ψ_1 to ψ_4. There is a clear physical correlate (an r in the vicinity of .90) for each of the four perceptual dimensions. Again, correlations greater than .83 have a probability less than .01 for a two-sided test. And again, the remaining correlations fall below .71, which has a corresponding probability of .05 for a two-sided test.

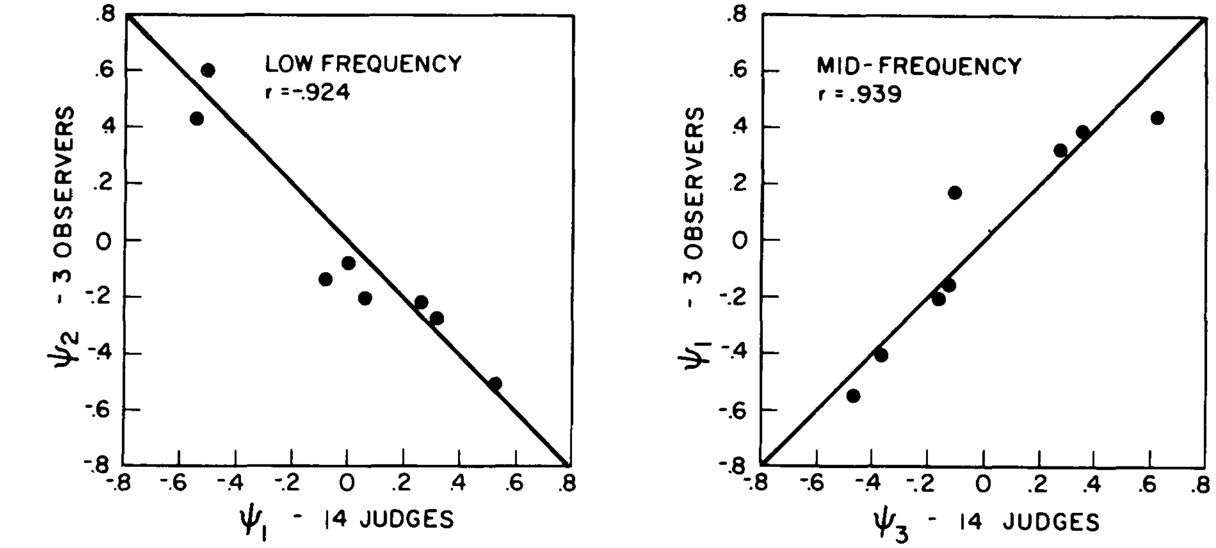
Comparison of the INDSCAL Solutions for the Two Groups of Subjects

Immediately apparent in Tables 2 and 4 is that the physical measures ϕ_1 to ϕ_3, which were selected to correlate highly with the three psychological dimensions of the three primary observers, correlate highly as well with the independent set of data obtained from the 14 additional judges. Moreover, the additional judges yield a fourth dimension that corresponds to the physical variable of periodicity.

We asked the judges, after the similarity ratings, to write down the dimensions they were using in those ratings. According to our translation, 12 of the 14 judges listed something related to our two frequency-related physical dimensions—having to do with the relative darkness of vertical bands in different left-mid-right positions. And 13 of the 14 judges listed a variable corresponding to our physical dimension of "contrast." Six of the 14 judges mentioned a variable related to the physical dimensions of "periodicity." Though "periodicity" was remarked upon by a minority of the judges, it appeared to be quite salient for that minority. These 6 judges had an average relative weighting of .29 on ψ_4, while the remaining 8 judges had an average relative weighting of .08 on that dimension. For that minority, the correlation between original data and computed scores increased an average of .067 in moving from three to four dimensions; for the others, the average increase was .014.

Returning to the matter of the correspondence of the first three psychological dimensions yielded by the two groups of subjects, we can observe that the three dimensions do not appear in the same order for the two groups. Tables 2 and 4 show that the three ordered psychological dimensions for the three observers correlated, respectively, with the physical dimensions of (1) midfrequency energy, (2) low-frequency energy, and (3) contrast, while the three ordered dimensions for the 14 additional judges correlated, respectively, with (1) low-frequency energy, (2) contrast, and (3) midfrequency energy. We submit, however, that the ordering is of relatively little importance; the fact that two independent sets of subjects yielded essentially the same three dimensions as their first three dimensions constitutes fairly strong support for the validity and generality of the (INDSCAL) technique of multidimensional scaling.

This kind of support for the validity and generality of the MDS technique can be obtained at a more fundamental level in our data, that is, at a level not dependent upon our having correctly identified the



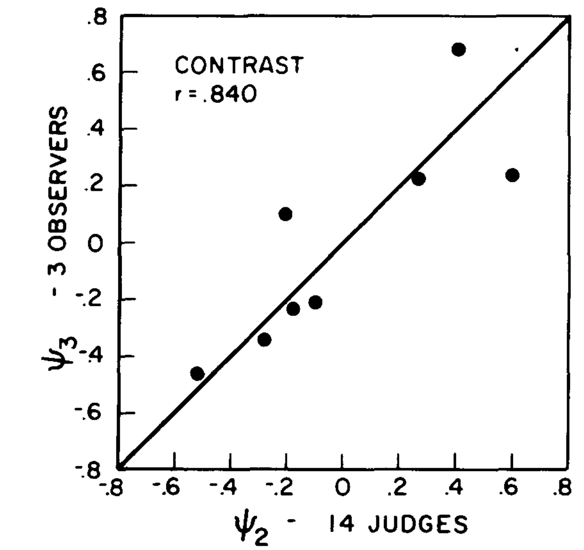


Figure 5. Relative loci of stimuli along pairs of psychological dimensions yielded by the two groups of subjects.

physical dimensions corresponding to the psychological dimensions. The question is thus better framed as one of the degree of correspondence, for the two groups of subjects, between the relative loci of the stimuli along separate dimensions of the psychological space. The answer is given in Figure 5, which shows the loci of stimuli for the pairs of psychological dimensions that correspond most fully between groups of subjects. The product-moment correlations for the three pairs of dimensions are −.924, .939, and .840, for the three panels of the figure, as shown in the insets of the figure. Each of these coefficients has an associated probability less than .01.

A Procedure to Relate Periodicity to Identification Performance

In order to determine the gain, if any, in predicting confusion matrices when periodicity was included as a fourth dimension for the three observers, we created a periodicity coordinate for each stimulus. The assigned values, given in Table 5, are linearly related to our physical measure of periodicity, subject to the constraints—used by INDSCAL in assigning coordinates on a psychological dimension—that (1) the mean across stimuli is 0, and (2) the variance across stimuli is 1. We realize that the true psychological coordinates for periodicity are probably not linear with the physical measure; however, lacking a rationale for any other specific relationship, we may suppose that linearity is a reasonable first approximation to the true relationship.

THE 8 BY 8 COMPLETE IDENTIFICATION TASK

There were eight response alternatives in the complete identification task, each corresponding to identification of one of the eight stimulus patterns. Of the 54 blocks of trials run, the first 3 were regarded

Table 5
Values Assigned to the Eight Stimuli for a Periodicity Coordinate for the Three Observers

	Stimulus							
	1. SC	2. BI	3. CC	4. TO	5. DE	6. RS	7. SN	8. FL
Coordinate Value	−.540	−.386	.077	.231	.386	.540	−.231	−.077

as practice and omitted from analysis. In addition, the data of 2 other blocks, and part of a third, were lost due to equipment failures. The remaining 2,421 trials for each observer were included in the analyses that follow.

Results

Error probability. A plot of error probability against stimulus number shown for individual observers in Figure 6, reveals that the stimuli were not equally confusable. In fact, Stimulus 2 (BI) was never, or almost never, confused with any other stimulus. Individual observers showed similar patterns of errors across stimuli, as seen in Figure 6. They also showed similar overall error rates (13%, 20%, and 21% errors for observers B.F., J.K., and J.S., respectively). The overall probability of a confusion error, averaged across stimuli and observers, was 18% (1,304 errors in 7,263 trials). That error rate was somewhat lower than intended but is adequate for our purposes.

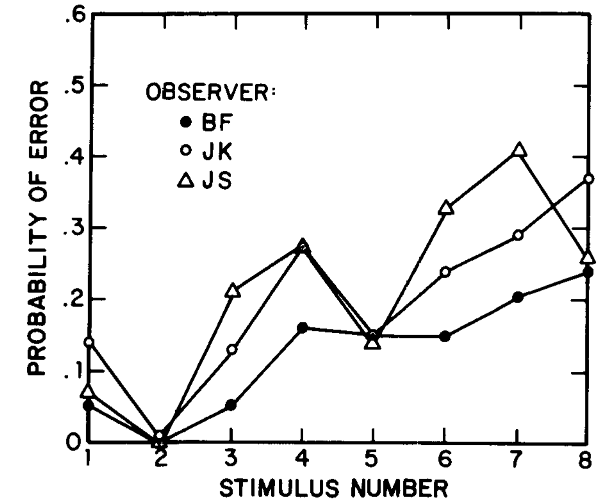


Figure 6. 8 by 8 experiment: Probability of a confusion error for each of the eight stimuli for each observer.

Confusion matrix. The matrices of raw confusion frequencies are given for each observer in Table 6. The patterns of confusions embedded in these numbers are most readily apparent when response probability distributions are plotted for each stimulus, shown separately for each observer by the solid lines and filled circles in Figure 7. A prominent feature of these data is the high degree of similarity across the three observers in the confusion pattern for each stimulus. It is this matrix of confusion distributions that we seek to predict by the model, as described in the following section.

Table 6
8 by 8 Experiment: Matrix of Confusion Frequencies for Each Observer

	Response								
Stimulus	1	2	3	4	5	6	7	8	Total
				Observer B.F.					
1	273	0	1	1	0	0	12	0	287
2	0	325	0	0	0	0	0	0	325
3	2	0	271	0	0	3	8	2	286
4	1	2	0	238	18	7	8	10	284
5	0	0	3	26	249	8	3	3	292
6	0	1	7	20	0	283	0	21	332
7	23	1	5	1	3	1	245	29	308
8	1	0	2	21	2	25	23	233	307
				Observer J.K.					
1	248	0	1	0	1	0	36	1	287
2	0	323	0	0	1	1	0	0	325
3	0	0	250	2	0	26	3	5	286
4	0	2	4	206	31	28	2	11	284
5	0	0	0	17	247	24	0	4	292
6	0	1	25	32	2	251	1	20	332
7	12	5	9	3	7	4	218	50	308
8	1	7	5	33	4	29	35	193	307
				Observer J.S.					
1	267	0	1	0	0	0	17	2	287
2	0	325	0	0	0	0	0	0	325
3	0	0	226	11	1	36	9	3	286
4	0	0	2	208	22	41	0	11	284
5	0	0	3	27	250	7	2	3	292
6	0	0	32	73	0	222	1	4	332
7	23	2	8	9	6	1	181	78	308
8	0	0	5	56	4	2	13	227	307

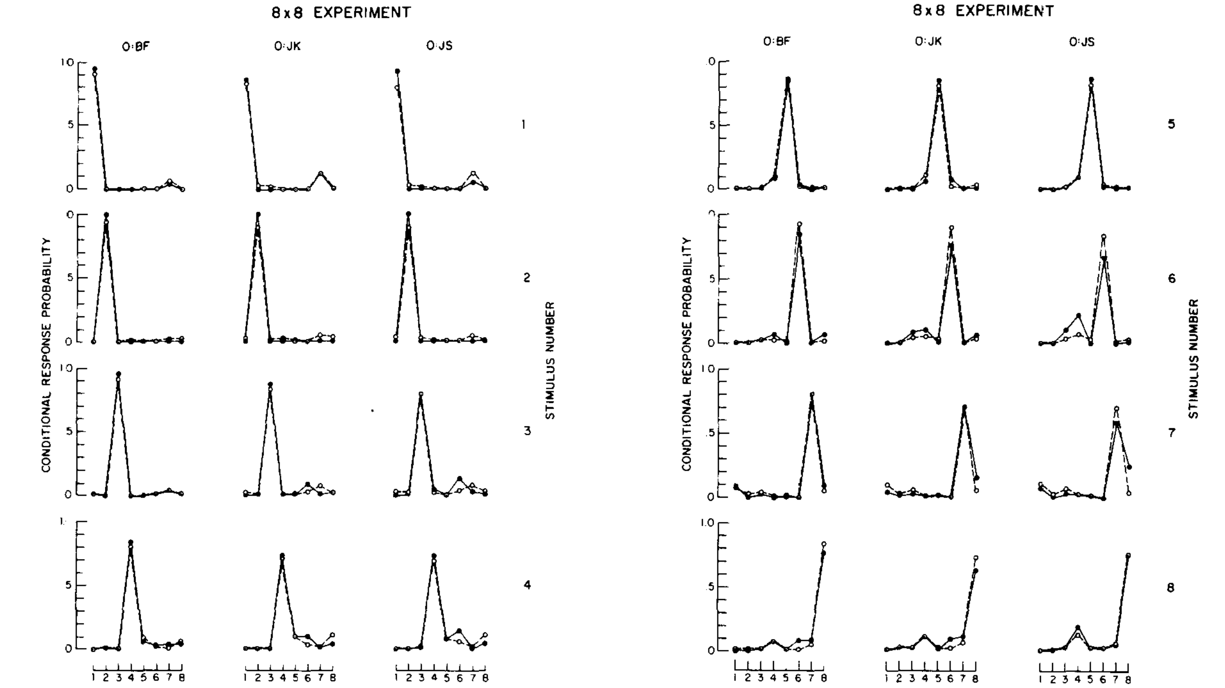


Figure 7. 8 by 8 experiment: Distribution of response probability for each stimulus (Stimuli 1 to 4 on left, Stimuli 5 to 8 on right) for each observer. Obtained distributions are given by solid lines and filled circles; distributions predicted by the model are given by dashed lines and open circles.

Model Analysis

Parameter estimation. Model parameters were estimated separately for each observer using a negative-gradient hill-climbing technique that sought to minimize the sum of squared deviations between the predicted and obtained confusion matrices for that observer. Under several different analysis conditions regarding the dimension salience weights, discussed below, the number of parameters estimated ranged from one (*a*, the sensitivity parameter) to five (*a* and four salience weights, w_1 to w_4).

Prediction of confusion matrices. We first fitted the model to the individual confusion matrices of our three observers using the three dimension salience weights for each observer provided by the INDSCAL analysis. The parameter values of individual observers, and the proportion of variance accounted for in the data of individual observers, and the average values are shown in Table 7A.

Estimating only a single parameter, *a*, for each observer, we found that we were able to account for 97% of the variance in the full confusion matrix, averaged across observers. On the other hand, if we included only identification errors (the off-diagonal elements) in the variance analysis, we accounted for only 21% of the error-matrix variance. There are at least four reasons why the variance accounted for, considering only errors, may be small; we take them up in turn in the next four paragraphs.

First, the obtained relative frequencies calculated for the off-diagonal cells contain the effects of both quantization error (the observed frequency of each response is integral) and sampling error. Since the predicted probabilities of most off-diagonal cells are very small, the range of variation in the measures to be correlated is small. We may expect, on these grounds, that the unpredicted variation due to quantization and sampling error is substantial relative to the small total amount of variation present.

Second, the brief stimulus duration (2 sec) may have resulted in some proportion of trials on which the stimulus was either not seen or not seen long enough for adequate encoding. Responses on these trials should represent pure guesses, uniformly distributed across the set of responses (assuming no response biases). While guessing undoubtedly occurred occasionally, it is unlikely that it represents a major contribution to the confusion matrix. If it did, we would expect to see a nonzero baseline response frequency across all responses for a given stimulus, an expectation which is not confirmed by the observed confusion matrices.

Third, the set of dimension salience weights derived by INDSCAL from the similarity-judgment task may not accurately represent dimension salience in the identification task. This possibility was examined by fitting the model with the salience weights, w_1 to w_3, and the sensitivity parameter, *a*, free to vary.

Table 7
8 by 8 Experiment: Estimated Parameter Values and the Proportion of Variance Accounted for by the Model for Each Observer

	Model Parameters					Proportion of Variance Accounted for	
Observer	w_1	w_2	w_3	w_4	a	Full Matrix	Error Matrix
(A) Three-Dimensional Model: w_1-w_3 INDSCAL Constrained							
B.F.	.34*	.26*	.41*		10.9	.987	.281
J.K.	.57*	.29*	.14*		8.1	.971	.211
J.S.	.33*	.39*	.28*		7.7	.942	.126
Mean	.41*	.31*	.28*		.89	.967	.206
(B) Three-Dimensional Model: w_1-w_3 Free Parameters							
B.F.	.73	.14	.13		10.8	.988	.290
J.K.	.55	.27	.18		8.1	.972	.215
J.S.	.49	.28	.23		7.8	.943	.124
Mean	.59	.23	.18		8.9	.968	.210
(C) Four-Dimensional Model: w_1-w_3 INDSCAL Constrained, w_4 Free Parameter							
B.F.	.16*	.12*	.19*	.54	9.4	.991	.480
J.K.	.34*	.18*	.08*	.40	7.5	.978	.376
J.S.	.19*	.23*	.16*	.43	7.3	.955	.305
Mean	.23*	.18*	.14*	.46	8.1	.975	.387
(D) Four-Dimensional Model: w_1-w_4 Free Parameters							
B.F.	.27	.08	.06	.59	9.5	.992	.543
J.K.	.30	.12	.11	.47	7.6	.978	.406
J.S.	.14	.14	.25	.47	7.4	.957	.325
Mean	.24	.11	.14	.51	8.2	.976	.425

**Parameter not estimated; value derived from INDSCAL analysis.*

As seen in Table 7B, the estimated parameter values change somewhat from their INDSCAL-derived values; however, there is no significant increase in proportions of variance accounted for either in the full matrix or in the off-diagonal cells. Thus, for a model of three dimensions, the INDSCAL-derived salience weights are nearly optimal relative to the best possible performance of the model with unconstrained choice of parameter values.

This leads us to consider the fourth possible reason for the error variance accounted for being small, namely, that one or more dimensions used by the observers in the identification task did not emerge in the INDSCAL analysis. As discussed earlier, we have good reason to suspect that the temporal periodicity present in the stimulus patterns is one such missing dimension. So we fitted the model to the confusion matrices a third time, using the INDSCAL-derived salience weights for each observer for the first three dimensions and allowing the salience weight for the fourth dimension (periodicity) to vary. By including periodicity as a fourth dimension, we found that the average proportion of variance accounted for in the full matrix increased from 97% to 98%, as shown in Table 7C, and that accounted for in the off-diagonal cells increased from 21% to 39%, a substantial improvement. Moreover, with an average salience weight of .46 assigned to periodicity, it was by far the most important dimension among the four in determining interstimulus distance.

Finally, we wished to determine if any further improvement in prediction might be obtained by allowing the salience weights (constrained to their INDSCAL values in the last fit) to vary. The result, shown in Table 7D, was essentially no further increase in the proportion of variance accounted for. The values of w_1, w_2, and w_3 changed very little on average from the INDSCAL values—indicating, as observed before, that the INDSCAL values were nearly optimal in terms of the model's ability to predict the full confusion matrix.

The pattern of results discussed above suggests that the effect of including periodicity as a fourth dimension is independent of, and additive to, the effect of freely estimating the first three salience weights. Including periodicity as a fourth dimension increases the proportion of variance accounted for by 1% and 18%, respectively, in the full and error matrices when INDSCAL constraints are used (Tables 7A and 7C) and by 1% and 22% in the full and error matrices when parameters are freely estimated (Tables 7B and 7D). Freely estimating weights w_1 to w_3 increases the proportion of variance accounted for by 0% in both the full and error matrices when only three dimensions are used (Tables 7A and 7B), and by 0% and 4% when periodicity is included as a fourth dimension (Tables 7C and 7D).

We turn now from summary measures of goodness of fit to the prediction of individual cells in the confusion matrix, using the four-dimensional model with freely estimated salience parameters. We choose this version of the model, in spite of the almost equally good fit noted above when INDSCAL-constrained parameters are used, for consistency with data reported in the next section on the partial identification tasks, where fits of the different model versions were not equally good. The distributions of predicted conditional response probabilities are plotted for each observer in Figure 7 as open circles connected by dashed lines, superimposed on the obtained distributions. The difficulty one has in separating obtained and predicted curves attests to the considerable accuracy of the predictions.

There are occasional deviations between predicted and obtained probabilities that appear to be systematic in that two of the three observers show the same pattern of deviations. For example, when Stimulus 3 was presented, both observers J.K. and J.S. made Response 6 more frequently and Response 7 less frequently than predicted. This, and other such examples, may result from the observers' use of dimensions that were available in the set of patterns but not included in the model analyses. Overall, though, the model that incorporates three INDSCAL-derived dimensions, and a fourth added, predicts quite well the individual cells of the confusion matrix.

THE 8 BY 4 PARTIAL IDENTIFICATION TASK

We may test the model in another way by considering how well it is able to predict the pattern of identification confusions when an observer is limited to responses associated with only a subset of the eight stimuli. In this section, we apply the model to three conditions of an 8 by 4 partial identification task, in which only four of eight stimuli—referred to as the "signals"—correspond to allowable identification responses, a different set of four stimuli in each of the three conditions.

The "signals" in Condition 1 were the four stimuli, 1, 2, 5, and 6, of Figure 3. These signals were not clearly distinguished from the remaining four "noise" stimuli on any of the four physical dimensions discussed previously. In Condition 2, the signals were Stimuli 3, 4, 5, and 6, the patterns with relatively high-frequency periodicities. In Condition 3, the signals were Stimuli 1, 3, 5, and 7, patterns which tended to have low values on our physical measure of midfrequency energy.

The analyses that follow are based on 230 trials for each observer in each condition. On the average, each of the eight stimuli was presented about 28 times in each condition. An initial block of 30 practice trials has been omitted from analysis for each condition.

Results

Error probability. Error rates can be defined only for the four "signals" in each condition. All responses on "noise" trials were necessarily errors, since the noise stimuli did not correspond to allowable identification responses. Using only the signal subsets of the confusion matrix for each condition, we found considerable variation in the error rates of the three observers averaged across conditions: 5%, 24%, and 8% errors for observers B.F., J.K., and J.S., respectively. There was also variation in the error rates of the three conditions averaged across observers: 8%, 21%, and 9% for Conditions 1, 2, and 3, respectively. Finally, the overall error rate in this experiment, 12%, was somewhat lower than that in the full 8 by 8 experiment (18%).

Confusion matrices. The raw confusion matrices are given in Table 8 for each observer and for each condition. The corresponding response distributions are plotted in Figures 8-10 for each stimulus, for each observer, and for each condition by the filled circles connected by solid lines. As in the first experiment, the response distributions for a given stimulus and condition are generally very similar across the three observers. The most notable exceptions are the distributions for Stimuli 6, 7, and 8 in Condition 2 (Figure 9B) and for Stimuli 2, 6, and 8 in Condition 3 (Figure 10). With the exception of Stimulus 6 in Condition 2, these are all instances in which the identification response corresponding to the presented stimulus was not among the set of allowed responses, a fact we will return to shortly in considering the model's predictions.

Model Analysis

Parameter estimation. The application of the model to the confusion matrices proceeded in much the same way as in the previous experiment. As before, the model was fitted to the data under four different conditions corresponding to combinations of two factors: (1) the inclusion or not of periodicity as a fourth dimension, and (2) the use of INDSCAL-derived or freely estimated values for the three salience weights, w_1 to w_3. In all cases, the psychological coordinates of the eight stimuli on the first three dimensions were taken from our INDSCAL analysis of the similarity-judgment data, and the coordinates of periodicity from our physical measure of that dimension. Model parameters were estimated separately for each observer in each of the three experimental conditions and for each of the four versions of the model.

Predictions of confusion matrices. The parameter values and proportion-of-variance values for each condition, averaged across observers, and the average values across conditions are shown for each of the four versions of the model in Table 9. Comparing the average proportion of variance accounted for by each model in Table 9 with the comparable average value for the 8 by 8 experiment in Table 7, it is clear that the model is accounting for somewhat less of the variance in the full matrix than before but for sub-

Table 8
8 by 4 Experiment: Matrix of Confusion Frequencies for Each Observer in Each of Three Conditions

Stimulus	Condition 1 Response					Condition 2 Response					Condition 3 Response				
	1	2	5	6	Total	3	4	5	6	Total	1	3	5	7	Total
							Observer B.F.								
1	28	0	0	0	28	32	0	0	0	32	32	0	0	1	33
2	0	33	0	0	33	1	26	0	0	27	25	0	0	0	25
3	9	0	1	18	28	25	0	0	1	26	0	30	0	2	32
4	0	0	11	19	30	0	24	0	4	28	0	0	23	8	31
5	0	0	27	1	28	0	2	21	0	23	0	0	23	0	23
6	0	0	2	28	30	0	3	0	42	45	0	6	7	18	31
7	23	1	2	1	27	16	1	1	2	20	2	1	0	27	30
8	2	7	9	8	26	4	15	0	10	29	0	1	1	23	25
							Observer J.K.								
1	28	0	0	0	28	28	0	3	1	32	29	0	0	4	23
2	0	32	1	0	33	0	26	1	0	27	0	0	0	25	25
3	4	0	3	21	28	20	0	0	6	26	1	26	1	4	32
4	1	2	9	18	30	0	16	1	11	28	0	1	25	5	31
5	2	0	20	6	28	0	2	14	7	23	0	1	22	0	23
6	0	0	8	22	30	3	9	12	21	45	0	5	22	4	31
7	18	1	3	5	27	5	1	8	6	20	4	1	1	24	30
8	7	1	7	11	26	0	10	11	8	29	0	0	10	15	25
							Observer J.S.								
1	28	0	0	0	28	32	0	0	0	32	29	0	0	4	33
2	0	33	0	0	33	0	27	0	0	27	0	0	25	0	25
3	1	0	0	27	28	24	0	0	2	26	0	32	0	0	32
4	0	3	5	22	30	0	23	0	5	28	0	2	29	0	31
5	0	0	22	6	28	0	1	22	0	23	0	0	23	0	23
6	0	0	1	29	30	1	5	0	39	45	0	23	8	0	31
7	14	5	4	4	27	14	5	1	0	20	1	3	0	26	30
8	3	1	4	18	26	2	25	0	2	29	0	7	7	11	25

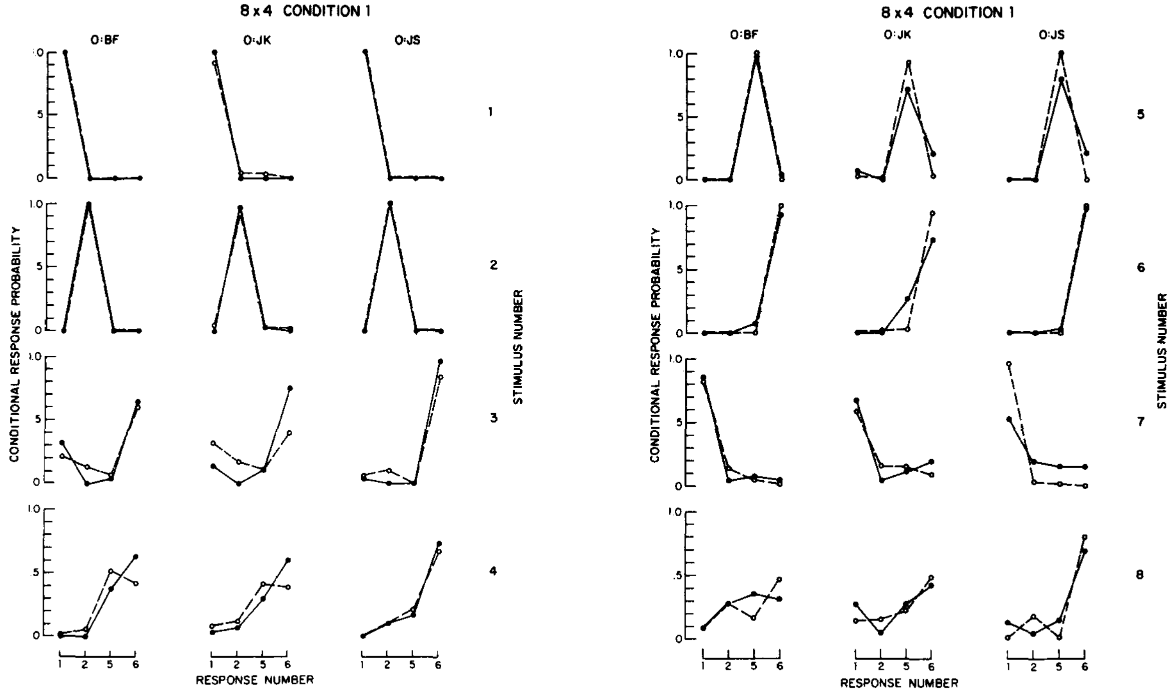


Figure 8. 8 by 4 experiment, Condition 1: Distribution of response probability for each stimulus (Stimuli 1 to 4 on left, Stimuli 5 to 8 on right) for each observer. Obtained distributions are given by solid lines and filled circles; predicted distributions are given by dashed lines and open circles.

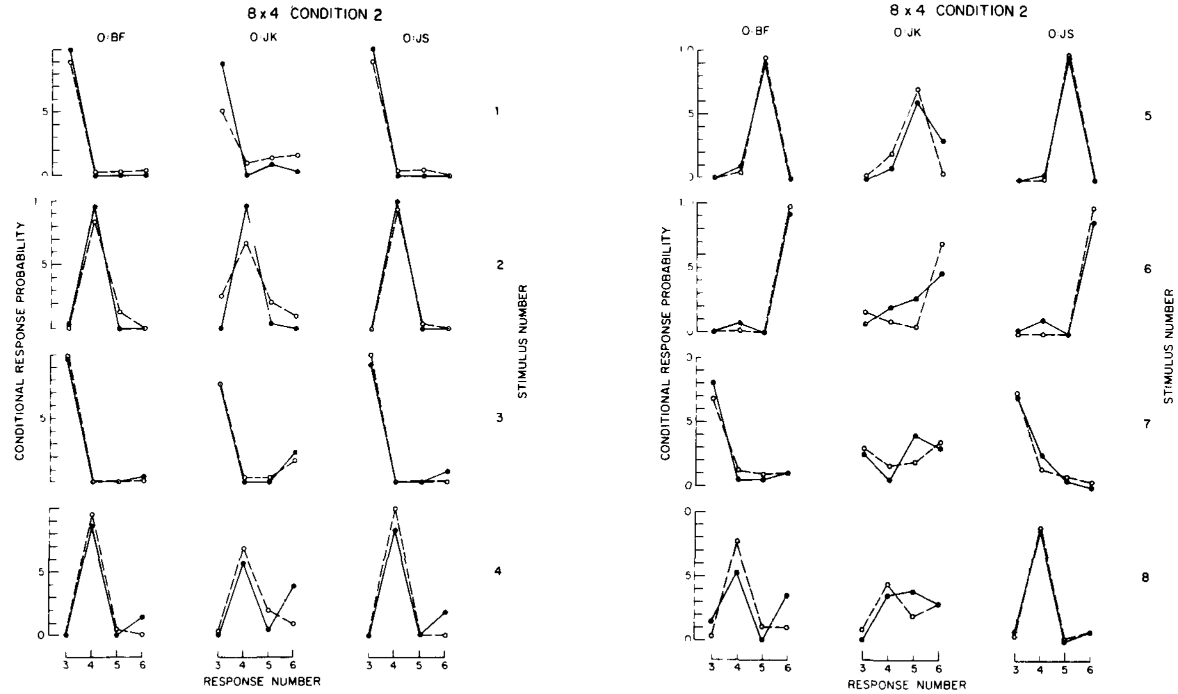


Figure 9. 8 by 4 experiment, Condition 2: Distribution of response probability for each stimulus (Stimuli 1 to 4 on left, Stimuli 5 to 8 on right) for each observer. Obtained distributions are given by solid lines and filled circles; predicted distributions are given by dashed lines and open circles.

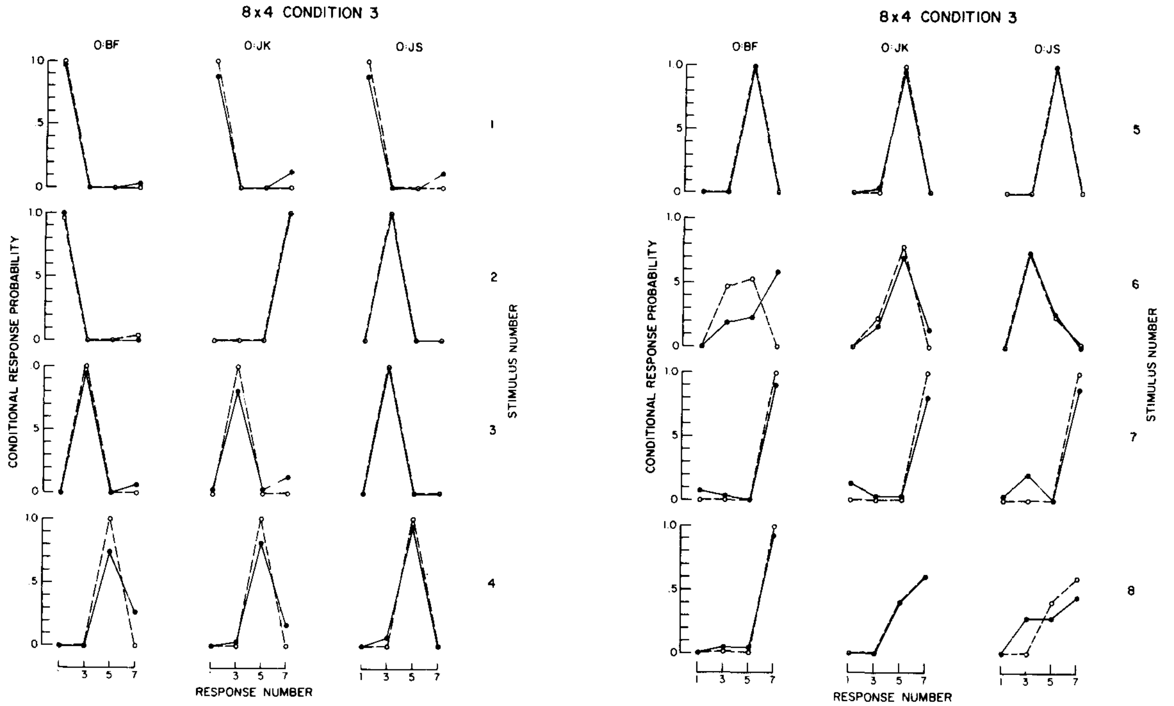


Figure 10. 8 by 4 experiment, Condition 3: Distribution of response probability for each stimulus (Stimuli 1 to 4 on left, Stimuli 5 to 8 on right) for each observer. Obtained distributions are given by solid lines and filled circles; predicted distributions are given by dashed lines and open circles.

Table 9
8 by 4 Experiment: Estimated Parameter Values and the Proportion of Variance Accounted for by the Model Averaged Across Observers for Each of the Three Conditions

Condition	Model Parameters					Proportion of Variance Accounted for	
	w_1	w_2	w_3	w_4	a	Full Matrix	Error Matrix
	(A) Three-Dimensional Model: w_1-w_3 INDSCAL Constrained						
1	.41*	.31*	.28*		5.6	.814	.580
2	.41*	.31*	.28*		17.5	.772	.682
3	.41*	.31*	.28*		8.0	.643	.322
Mean	.41*	.31*	.28*		10.7	.743	.528
	(B) Three-Dimensional Model: w_1-w_3 Free Parameters						
1	.31	.43	.26		10.6	.859	.693
2	.63	.33	.04		14.9	.859	.822
3	.44	.28	.28		16.7	.786	.672
Mean	.46	.35	.19		14.1	.835	.729
	(C) Four-Dimensional Model: w_1-w_3 INDSCAL Constrained, w_4 Free Parameter						
1	.29*	.23*	.19*	.29	7.3	.855	.682
2	.39*	.30*	.27*	.04	17.5	.773	.684
3	.17*	.15*	.14*	.55	12.2	.746	.511
Mean	.28*	.23*	.20*	.29	12.3	.791	.626
	(D) Four-Dimensional Model: w_1-w_4 Free Parameters						
1	.20	.32	.17	.31	12.0	.897	.791
2	.58	.20	.00	.22	9.5	.858	.822
3	.32	.09	.21	.39	90.1	.937	.888
Mean	.37	.20	.13	.31	37.2	.897	.834

**Parameter not estimated; value derived from INDSCAL analysis.*

stantially more of the variance in the error matrix. The first result is probably due in large part to the considerably smaller number of trials contributing to each of the confusion matrices—approximately 30—resulting in less stable estimates of the observed conditional response probabilities. The second result follows from the inclusion of noise trials on which observers are forced to make confusion errors. As a consequence, the amount of variability available to be explained is large in the error matrices of the 8 by 4 experiment relative to the amount in the error matrices of the 8 by 8 experiment.

Paralleling the results of the analysis of the 8 by 8 experiment, the pattern of changes in the variance accounted for across the four versions of the model suggests that the effect of including periodicity as a fourth dimension is independent of, and additive to, the effect of freely estimating the first three salience weights. Including periodicity as a fourth dimension increases the proportion of variance accounted for by 5% and 10% in the full and error matrices, respectively (Tables 9A and 9C). Freely estimating the salience weights w_1 to w_3 increases the proportion of variance accounted for by 10% and 20% in the full and error matrices, respectively (Tables 9A and 9B). Doing both—including periodicity as a fourth dimension and freely estimating the salience weights w_1 to w_3—increases the proportion of variance accounted for by 16% and 30% in the full and error matrices, respectively (Tables 9A and 9D).

The substantial improvement in the model's predictions, when the salience weights w_1 to w_3 are freed of their INDSCAL-derived values and independently estimated in each condition, is in contrast to the lack of improvement observed in the 8 by 8 experiment. The reason is apparent in Figure 11, which shows the estimated salience weights (including the periodicity dimension) for each observer, separately for the 8 by 8 experiment and each of the three conditions of the 8 by 4 experiment. The pattern of estimated weights clearly changes from one condition to another, suggesting that the observers modified their set of salience weights from condition to condition according to the composition of the set of four stimuli defined as signals. In Conditions 1 and 2, the patterns of weights appear quite similar across observers; in Condition 3, they appear relatively different.

The predicted response distributions for each stimulus and for each observer in each of the conditions are shown by the open circles connected by dashed lines in Figures 8-10. These predictions are based on the four-dimensional model with freely estimated salience weights. Scanning over the large number of distributions, the overall impression is that there is a remarkably good agreement between predicted and obtained distributions. It is worth noting that a 10% deviation between obtained and predicted probabilities corresponds to a difference of only about three responses, given that each obtained distribution is based on about 30 trials.

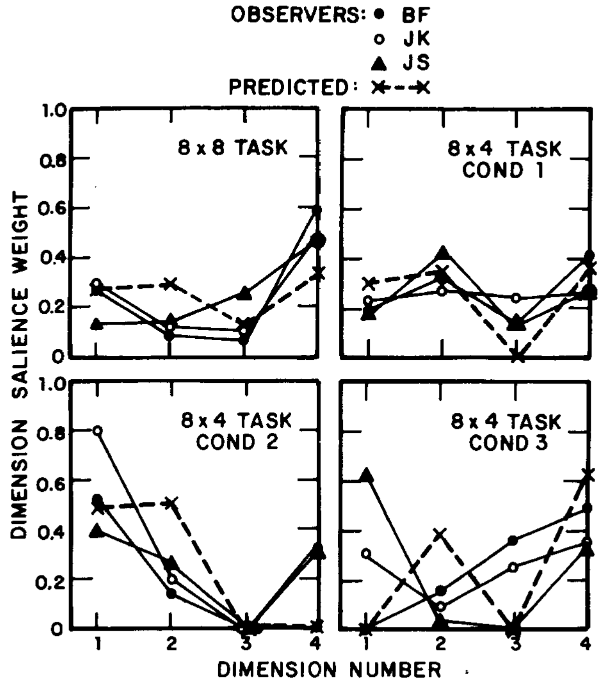


Figure 11. Estimated salience weights on the four dimensions for each observer for each condition in the 8 by 8 and 8 by 4 experiments (solid lines). The pattern of weights that maximizes probability of correct identification is given by the dashed line for each condition.

We mentioned earlier that most of the cases in which the observed response distributions for a given stimulus differed among the three observers were instances of noise stimuli. We now observe that most of the cases in which there is a substantial deviation between predicted and obtained response distributions are also instances of noise stimuli (e.g., Stimulus 6 for Observer B.F. in Condition 3). In fact, the structure of the model provides an insight into the source of these deviations. It will sometimes be true that the interstimulus distances between a particular noise stimulus and each of the four signal stimuli allowed as responses are all very large, and thus that the corresponding confusability weights are all very small. The observer would have no difficulty in rejecting all four responses as corresponding to the true "identity" of the presented stimulus. In this predicament, the observer most likely reinterprets the task to respond with the stimulus among the allowable four which is "most similar" to the presented stimulus, even though it is clearly incorrect. However, if the confusability weights for all four responses are extremely small, we may expect the observed response distribution to be very sensitive to small differences in salience weights and also to other decision processes.

A particularly good illustration of the problem is found in the response distributions to Noise Stimulus 2 in Condition 3 (Figure 10). Each observer made a single response exclusively—but a different response for each observer. In all likelihood, each observer was aware that the pattern was Stimulus 2 at each presentation. Having decided once which signal stimulus was "most similar" to Stimulus 2, the observer then remembered and gave that response consistently thereafter. It can be seen in Figure 10 that the model accurately predicted the different distributions for the three observers—but at a cost. The parameter-estimation algorithm has attempted to accommodate the extreme distribution by increasing the sensitivity parameter value in Condition 3 relative to Conditions 1 or 2 (see Table 9D), thus decreasing the error in predicting that particular response distribution. As a consequence, however, the prediction of other response distributions are now more extreme than they otherwise would be. This effect can be seen in the predicted distributions for most other stimuli in Condition 3.

Overall, however, the model, using the perceptual space derived from similarity judgments within the context of the full stimulus set, predicts quite well the confusion matrices for partial identification tasks in which different subsets of the stimuli are identified. The predictions are improved, as in the complete identification task, when a fourth dimension is added to the MDS-derived perceptual space.

ADAPTIVE TUNING

Adaptive Tuning in Identification

Our application of the model to the complete identification task and the three conditions of the partial identification task suggests that the observers were flexible in their use of perceptual dimensions to identify the stimuli, in that our estimates of the relative salience weights on the several dimensions varied from condition to condition (see Figure 11). Given that the same set of eight stimuli was presented in both tasks, we may ask what motivated the observers to adjust the pattern of salience weights as they did across the different conditions. While the set of stimuli remained constant across all conditions, the subset of stimuli that we required the observer to identify—the "signals"—changed from condition to condition. Furthermore, it was only for this subset of stimuli that the observer received discriminative feedback that indicated which stimulus had been presented. We believe that the observer was engaged in an adaptive tuning process in which the relative weighting of dimensions was adjusted in order to maximize the discriminability of the subset of stimulus patterns to be identified in that condition. This tuning process probably takes place gradually, over many trials, based on the feedback given the observer regarding the correctness of identification.

Given that observers were instructed to maximize their percentage of correct identifications, it seems likely that this criterion formed their basis for tuning.

To test this hypothesis, we have determined from the model, for each condition of the experiment, what pattern of dimension salience weights would maximize the probability of a correct signal identification. The optimal pattern of weights, assuming an average value for the sensitivity parameter, is plotted by the dashed line for each experimental condition of both tasks in Figure 11. While the detailed agreement of the observed and optimal weight patterns is not particularly good, there is a general correspondence of observed and predicted pattern shape across conditions, with the exception of Condition 3 of the 8 by 4 task. The comparison in this particular condition is probably not meaningful because of the large interobserver differences, as discussed earlier. In general, the observed patterns of dimension salience weights seem consistent with the hypothesis that observers are tuning their weighting of dimensions in order to maximize the probability of a correct identification.

Adaptive Tuning in Judgment of Similarity

The concept of adaptive tuning may also provide an explanation for the failure of periodicity to emerge as a dimension in the INDSCAL analysis of the similarity judgments obtained from our three observers. If we assume that the observers' perceptual spaces were the same in both the similarity-judgment and identification tasks, then periodicity could fail to emerge as a dimension in the INDSCAL analysis if the observers were according it zero or nearly zero weight. This would be analogous to our analysis of Condition 2 of the partial identification task in which Dimension 3 (contrast) was given zero weight by all observers, although this dimension presumably was available since it was used in other conditions. Thus, we suggest that periodicity was present in the observers' perceptual spaces in the similarity-judgment task but was given zero weight in the adapted pattern of dimension weights. Moreover, we believe that the particular pattern of weights resulted from an adaptive tuning process that sought to optimize some aspect of performance, as in the identification task. Since observers were instructed to map their perceived range of similarities into numbers so as to use the entire range from 1 to 10, we speculate that observers may have tuned dimension weights to obtain the maximal possible range of interstimulus distances over all pairs of stimuli. Further work is required to decide this issue.

DISCUSSION

A Validation of MDS Procedures and the Decision Model

Our approach in this paper assumes (1) that a set of complex stimuli can be represented perceptually as a set of points in a multidimensional psychological space, (2) that an MDS procedure can be used to derive the dimensions of that space and the relative loci of the stimuli within the space, and (3) that identification confusions can be predicted by a simple decision model based on weighted interstimulus distances. Our success in predicting confusion matrices for individual observers across several experimental conditions provides support for this approach.

We believe that this outcome provides a significant validation of the use of MDS procedures to reveal perceptual structure. Other evidence for the validity of MDS-derived spaces has relied largely on the intuitive reasonableness of the abstracted dimensions and stimulus configurations. The present outcome provides much stronger support in that the MDS-derived space is used to predict data in a different, independent ask. To be sure, the best prediction of identification behavior occurred when the MDS-derived space for our three observers was supplemented by an additional dimension. Nonetheless, the first three dimensions revealed by the MDS procedure for those observers were useful in the prediction.

Our results also support the decision model as a description of the identification process. The model accounts well for the changes in performance observed across different conditions of the complete and partial identification tasks. Of particular interest is the fact that the model accounts for these changes in performance in terms of changes in the relative salience of perceptual dimensions rather than in terms of changes in the structure of the perceptual space. In the model, a given stimulus is assumed to have a fixed location in the perceptual space, regardless of changing stimulus context.

The MDS/Decision-Model Approach as an Aid in Predicting Identification Performance

A new, particular identification task arises whenever new signals come under study. New signals are studied when new sources, sensors, settings, or displays are discovered, defined, or devised. Examples are replete in medical diagnosis, nondestructive testing of materials and equipment, military surveillance, analysis of biological microstructures, or study of the perceptually handicapped. Our experiments enhance the possibility that the present approach can be used to gain understanding about, and predict behavior in, any particular identification task.

Gaining understanding about a new set of signals, in this context, means isolating the perceptual dimensions that are useful and assigning the proper saliences to those dimensions. The practical import of the approach described here is that perceptual dimensions can be isolated quite simply and quickly. We suspect that observers can be economically trained to use the useful dimensions and to use them approximately with weights that maximize the probability of correct identifications.

We have pointed out that our application of an

MDS procedure to our three test observers did not define all of the dimensions that were found useful in the identification task but that the additional dimension in question (periodicity) did appear in the four-dimensional MDS solution yielded by 14 other judges, some of whom had more general experience with signals of the sort used than the three observers. In examining a new identification task, one might do well to employ a rather large number of judges having expertise with the general class of signals in question and to examine carefully the MDS solutions of larger dimensionality.

An alternative might be to ask the judges in the MDS task to rate the "confusability" of the stimulus pairs, with the identification task in their minds, rather than stimulus similarity. Our three test observers might have yielded the additional useful dimension (periodicity) under such instructions. Though our present experiments were constrained by our interest in validating MDS procedures as they are usually applied, MDS tasks undertaken specifically to analyze new sets of signals might possibly be adjusted in this manner to make them maximally useful.

REFERENCE NOTE

1. Hardzinski, M., & Pachella, R. G. *A psychophysical analysis of complex integrated displays*. [Technical Report 59 (014523-2-T)]. Ann Arbor, Michigan: University of Michigan, February 1977.

REFERENCES

Carroll, J. D. Individual differences and multidimensional scaling. In R. N. Shepard, A. K. Romney, & S. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences*. New York: Seminar Press, 1972.

Carroll, J. D., & Chang, J. J. Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika*, 1970, **35**, 288-319.

Carroll, J. D., & Wish, M. Models and methods for three-way multidimensional scaling. In R. C. Atkinson, D. H. Krantz, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology*. San Francisco: Freeman, 1973.

Howard, J. H., Jr. The psychophysical structure of eight complex underwater sounds. *Journal of the Acoustical Society of America*, 1977, **62**, 149-156.

Howard, J. H., Jr., & Silverman, E. G. A multidimensional scaling analysis of 16 complex sounds. *Perception & Psychophysics*, 1976, **19**, 193-200.

Klein, W., Plomp, R., & Pols, L. C. W. Vowel spectra, vowel spaces, and vowel identification. *Journal of the Acoustical Society of America*, 1970, **48**, 999-1009.

Luce, R. D. Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of Mathematical Psychology*. New York: Wiley, 1963.

Miller, J. R., & Carterette, E. C. Perceptual space for musical structures. *Journal of the Acoustical Society of America*, 1975, **58**, 711-720.

Morgan, B. J. T., Woodhead, M. M., & Webster, J. C. On the recovery of physical dimensions of stimuli, using multidimensional scaling. *Journal of the Acoustical Society of America*, 1976, **60**, 186-189.

Pachella, R. G., & Somers, P. The development of integrated multidimensional displays. *Naval Research Reviews*, in press.

Plomp, R., & Steeneken, H. J. M. Effect of phase on the timbre of complex tones. *Journal of the Acoustical Society of America*, 1969, **46**, 409-421.

Pols, L. C. W., Van Der Kamp, L. J. Th., & Plomp, R. Perceptual and physical space of vowel sounds. *Journal of the Acoustical Society of America*, 1969, **46**, 458-467.

Romney, A. K., Shepard, R. N., & Nerlove, S. B. (Eds.). *Multidimensional scaling: Theory and applications in the behavioral sciences* (Vol. II). New York: Seminar Press, 1972.

Shaw, D. J. A phonological interpretation of two acoustic confusion matrices. *Perception & Psychophysics*, 1975, **17**, 537-542.

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. Stimulus and response generalization: Deduction of the generalization gradient from a trace model. *Psychological Review*, 1958, **65**, 242-256. (a)

Shepard, R. N. Stimulus and response generalization: Tests of a model relating generalization to distance in psychological space. *Journal of Experimental Psychology*, 1958, **55**, 509-523. (b)

Shepard, R. N. Psychological representation of speech sounds. In E. David & P. Denes (Eds.), *Human communication: A unified view*. New York: McGraw-Hill, 1972.

Shepard, R. N., & Chipman, S. Second-order isomorphism of internal representations: Shapes of states. *Cognitive Psychology*, 1970, **1**, 1-17.

Shepard, R. N., Romney, A. K., & Nerlove, S. B. (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences* (Vol. I). New York: Seminar Press, 1972.

Stenson, H. H. The psychophysical dimensions of similarity among random shapes. *Perception & Psychophysics*, 1968, **3**, 201-214.

Swets, J. A., Green, D. A., Getty, D. J., & Swets, J. B. Signal detection and identification at successive stages of observation. *Perception & Psychophysics*, 1978, **23**, 275-289.

Tversky, A., & Krantz, D. H. The dimensional representation and the metric structure of similarity data. *Journal of Mathematical Psychology*, 1970, **7**, 572-596.

NOTE

1. Power metrics are the class of metrics such that the distance between the points $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ is given by

$$d(x,y) = \left[\sum_{i=1}^{n} \left| x_i - y_i \right|^r \right]^{1/r}$$

for $r \geq 1$.

(Received for publication February 28, 1979
revision accepted April 16, 1979.)