Theoretical analysis of an alphabetic confusion matrix*

J. T. TOWNSEND
Purdue University, West Lafayette, Indiana 47907

A study was undertaken to acquire a confusion matrix of the entire upper-case English alphabet with a simple nonserifed font under tachistoscopic conditions. This was accomplished with two experimental conditions, one with blank poststimulus field and one with noisy poststimulus field, for six Ss run 650 trials each. Three mathematical models of recognition, two based on the concept of a finite number of sensory states and one being the choice model, were compared in their ability to predict the confusion matrix after their parameters were estimated from functions of the data. In order to ascertain the facility with which estimates of similarity among the letters could lead to a psychological space containing the letters, η_{ij}, the similarity parameter of the choice model was input to an ordinally based multidimensional scaling program. Finally, correlation coefficients were computed among parameters of the models, the scaled space, and a crude measure of physical similarity. Briefly, the results were: (1) the finite-state model that assumed stimulus similarity (the overlap activation model) and the choice model predicted the confusion-matrix entries about equally well in terms of a sum-of-squared deviations criterion and better than the all-or-none activation model, which assumed only a perfect perception or random-guessing state following a stimulus presentation; (2) the parts of the confusion matrix that fit best varied with the particular model, and this finding was related to the models; (3) the best scaling result in terms of a goodness-of-fit measure was obtained with the blank poststimulus field condition, with a technique allowing different distances for tied similarity values, and with the Euclidean as opposed to the city-block metric; and (4) there was agreement among the models in terms of the way in which the models reflected sensory and response bias structure in the data, and in the way in which a single model measured these attributes across experimental conditions, as well as agreement among similarity and distance measures with physical similarity.

The study reported in these pages sought to pursue several related goals. The broad goal was establishment and investigation of a tachistoscopic confusion matrix obtained from human observers attempting to identify single members of the complete upper-case English alphabet. Despite the abundance of data concerning legibility of alphanumeric characters (e.g., see Cernog & Rose, 1967), some of the results are of marginal significance to the psychologist because (1) of the highly specific applied nature of the study, (2) too few trials or subjects were run, (3) only a few of the letters of the alphabet were used, or (4) no attempt was made to use mathematical models to separate the response-bias characteristics from the stimulus or sensory characteristics. This study was intended to complement the literature by rectification of these limitations.

Within the above-stated broad goal was the subsidiary aim of examining the ability of three simple mathematical models to predict average behavior in a complete identification confusion study. Because of the large array of data represented by the 26 by 26 confusion matrix, it was required of the models that they be of sufficiently simple structure that parameter estimation be feasible. One of the few extant models that possesses this property without simplifying assumptions is the choice model (Luce, 1959, 1963a). However, work with finite-state detection models (e.g., see Atkinson & Kinchla, 1965; Luce, 1963b) also suggested the possibility of extension of this type of model to more complex situations. Because of the very explicit nature of assumptions of these models about the sensory and bias processes, they can be helpful in investigation of perceptual characteristics by manipulation of these assumptions. To this end, two generalizations of the activation model suitable for application in the present experiment were developed. As will be seen, one of these is like the choice model in that it possesses structure for description of psychological similarity, but the other assumes only two psychological sensory states relating to perfect information or no information at all. It was of interest to learn if better predictions are obtained with models that assume sensory confusion states as opposed to the two-state type of model just mentioned as well as to compare the choice model with the activation representative. In addition, a

simple correlational analysis was carried out in order to acquire a rough idea as to what extent the models seemed to be consistent in different conditions and, among themselves, to be measuring or reflecting the same attributes within the data. One limitation, of course, in the present type of study is that the models purport to explain individual behavior. Hence, the work reported here must be viewed as testing the ability of the models to handle a large amount of averaged human confusion data but not as providing a critical test of the models' assumptions that are meant to apply at the level of the individual. An experiment is in progress collecting long-term confusion data at the individual level.

A second major type of analysis included in the present study was the use of an ordinal-based multidimensional scaling procedure applied to estimated similarity parameters in the hope of learning more about the characteristics of psychological space for letter identification, when all letters are in the stimulus population. The results of this and the analysis with the substantive models are discussed, interrelated, and compared to a simple physical measurement of interletter similarity.

Apart from the hypothesis-testing function of an experiment, there exist aspects that perhaps should be viewed as of an information-gathering nature. In this context, latencies and confidence ratings were collected and the summary results commented on, although the models employed here were not developed to the extent necessary to make predictions concerning these.

THEORETICAL DEVELOPMENTS

It was of interest in the present study to compare two models selected from a class of finite-state models with the choice model (Luce, 1963a), and a "threshold" finite-state model with models postulating sensory confusion. The class of finite-state models from which the two representatives were selected was motivated by the simple finite-state models used in some detection experiments (Atkinson & Kinchla, 1965; Kinchla, Townsend, Yellott, & Atkinson, 1966) and multisymbol recognition situations (Townsend, 1966). In order to provide continuity with the intuitions and structure developed earlier, we will refer to the finite class of models as the general activation model.

General Activation Model

The nature of the activation model in our development supposes a mapping of the set of N possible stimulus events to a (finite) set of n hypothetical internal sensory states. On the basis of the

particular internal state activated, S is assumed to make a decision as to the stimulus event, thus yielding a mapping of the n internal states to the set of N possible responses. For the complete identification experiment with N stimuli, the sensory activation matrix would appear as $S = (a_{ij})$ where $i = 1, 2, \cdots, N$ and $j = 0, 1, 2, \cdots, n$. The second subscript refers to the hypothetical sensory states $s_0, s_1, \cdots, s_n$. Similarly, we may write the decision- or bias-process matrix as $D = (b_{i,j})$, where $i = 0, 1, 2, \cdots, n$ and $j = 1, 2, \cdots, N$. Finally, we obtain the general theoretical confusion matrix by premultiplying D by S:

$$C = (c_{jk}) = \left(\sum_{j=0}^{n} a_{i,j} b_{jk} \right)$$

where $i = 1, 2, \cdots, N$ and $k = 1, 2, \cdots, N$, where, in general, the only constraints are that

$$\sum_{j=0}^{n} a_{ij} = 1 \text{ and that } \sum_{j=1}^{N} b_{ij} = 1.$$

The interpretation of the internal sensory states we wish to make in this paper is in terms of sensory confusions among the various stimuli. For example, depending on the actual set of stimuli, it might be possible that S is certain that the presented stimulus is one of a particular subset of stimuli (the cardinality possibly being equal to zero or to the number of possible stimuli) and hence, S must choose from this subset on the basis of learning or motivational factors.

It should be noted that one can devise continuous-state models that possess properties that reasonably mimic finite-state models. For example, if the sensory results of each stimulus presentation could be described by a uniform distribution, then there would exist, in general, areas of nonoverlap of probability density; if an observation occurred in such an area, a reasonable strategy would be for S to respond with the associated response, with probability one. On the other hand, if an observation were obtained at a point where, say, k stimuli gave rise to equal a priori densities, S might well guess on the basis of his knowledge of the presentation probabilities. For example, he might choose the signal with the highest a priori or a posteriori probability (which maximizes average number correct), or he might employ a Bayesian probability matching strategy. The second strategy would occur when S chooses, say, S_t from the k alternatives with probability

$$\frac{P(S_i)^{(k)}}{\sum_{j=1}^{k} P(S_j)^{(k)}}$$

In this expression, $P(S_j)^{(k)}$ represents the a priori probability of stimulus S_j and the superscript merely indicates that the stimulus is a member of the k-stimuli subset of equal-density stimuli at the observed point. Under the expressed circumstance this above quantity is both the Bayesian probability (hence the term "Bayesian probability matching") of stimulus S_i given the particular observed point and the conditional probability of S_i given the stimulus must be in the k-stimuli subset. We will not attempt to distinguish such theories from the finite-state theories here.

In order to obtain tractable special cases and to investigate the relative importance of pairwise sensory confusions as opposed to "pure" guessing (where S is in a state of complete uncertainty regarding the presented stimulus), we develop the all-or-none and overlap activation models for application to letter-recognition experiments.

All-or-None Activation Model

The first specialization posits that S either obtains such information from the display as allows him to respond perfectly or he is thrown into an uncertain state where he has no partial information and must respond by guessing. Note that S could be in this uncertain state either because he fails to detect anything at all or because such characteristics of the letter as he observes do not aid in identifying it. If we index the letters of the alphabet by the numbers 1 through 26 and refer to S_i as stimulus i (i.e., the ith letter of the alphabet) and R_j as response j, we may write the appropriate sensory activation matrix as: $S = a_{i,j}$ where $i = 1, 2, \cdots, 26$, $j = 0, 1, 2, \cdots, 26$, and where

$$a_{ij} = \begin{cases} 1 - \sigma_i \text{ when } j = 0, \\ \sigma_i \text{ when } j = i, \\ 0 \text{ otherwise,} \end{cases}$$

where, as before, j refers to sensory state s_j. These sensory states are related to the set of overt responses by the decision matrix $D = (b_{ij})$, where $i = 0, 1, \cdots, 26$, $j = 1, 2, \cdots, 26$, and

$$b_{ij} = \begin{cases} p_j \text{ when } i = 0 \\ 1 \text{ when } i = j, \\ 0 \text{ otherwise.} \end{cases}$$

Hence, s_1 through s_{26} represent "certain" sensory states since $P(R_i/s_i) = 1$; but in the "uncertain" state, s_0, S is assumed to guess S_i with probability p_i, which may depend on learning and motivation variables. Multiplying these two matrices yields the theoretical confusion matrix $C = (c_{ij})$ where $i = 1, 2, \cdots, 26$, $j = 1, 2, \cdots, 26$, and $c_{ij} = \sigma_i \delta_{ij} + (1 - \sigma_i) p_j$ where

$$\delta_{ij} = \begin{cases} 1 \text{ when } i = j, \\ 0 \text{ otherwise.} \end{cases}$$

A few comments are in order concerning related work on this model. For the two-signal case, when $\sigma_1 = \sigma_2$, this model reduces to that developed for two-interval forced-choice detection experiments by Atkinson and Kinchla (1965) and by Kinchla and associates (1966). Smith (1968) has studied properties of confusion matrices generated by this mode, which he terms a "pure perceptibility" model, and he has shown that it can be viewed as a special case of the choice model; this result will be mentioned again after the brief presentation of the choice model.

The all-or-none activation model requires estimation of $2N - 1$ parameters, 51 in the present experiment.

Overlap Activation Model

The basic intuition for the overlap activation model is that in addition to a unique activation state relating a letter in a one-one fashion to the correct response, there exist partial-information states such that S is in a state of uncertainty with respect to *two* letters; thus, this model assumes that pairwise similarity can affect interletter confusability. When in a state of confusion, S is postulated to respond according to the relative magnitude of the two concerned response-bias probabilities. We thus express the sensory activation matrix as composed of entries for the pairs $(S_i, t_{j,k})$, where $i = 1, 2, \cdots, 26$, $j = 1, 2, \cdots, 26$, and $k = j, j+1, \cdots, 26$, and where the entry for $(S_i, t_{j,k})$ is given by

$$\xi_{ijk} = \begin{cases} \xi_{jk} \text{ when } j = i, k = i, \text{ or both,} \\ 0 \text{ otherwise.} \end{cases}$$

Note that $t_{i,i}$ is the certain state for S_i, and $t_{i,j}$ is the uncertain state for S_i and S_j, $i \neq j$, and can proceed from presentation of either S_i or S_j ($\xi_{ij} = \xi_{ji}$). The corresponding decision matrix is given by entries corresponding to the pairs $(t_{j,k}, R_m)$ where j,k are as before and $m = 1, 2, \cdots, 26$. The entry for $(t_{j,k}, R_m)$ is equal to one when $j = k = m$, to $g_j/(g_j + g_k)$ when $j = m \neq k$, to $g_k/(g_j + g_k)$ when $k = m \neq j$, and to 0

otherwise, where g_i is the response bias parameter for R_i.

The theoretical confusion matrix, given as the result of the two foregoing processes is written $C = (c_{i,j})$ where $i = 1, 2, \cdots, 26$, $j = 1, 2, \cdots, 26$,

$$c_{ii} = \xi_{ii} + \sum_{k \neq i}^{26} \xi_{ik} \left(\frac{g_i}{g_i + g_k} \right),$$

and

$$c_{ij} = \xi_{ij} \left(\frac{g_j}{g_i + g_j} \right) \text{ when } i \neq j.$$

Now, it is clear that this model represents a deviation from the all-or-none model since not only is stimulus confusion allowed, but a pure, no-information, guessing state is precluded. It is also a one-step extension in the sense that we could have postulated the existence of k-wise confusion states. In general, this last possibility generates an unmanageable number of parameters. However, if one is willing to posit further theoretical structure, then the number of parameters may be reduced. This approach was taken by Nakatani (1968) in the development of a hybrid model that results in a set of confusion states, where from two to all N objects may be confused. Although Nakatani's work was independent of the present results, both have employed the expedient of using the ratios of the relevant bias, for example, $g_i/(g_i + g_j)$, to describe the S's guessing bias for a particular confusion state.

The overlap model is favored by properties relevant to the present aims: (1) It allows comparison of a two-way confusion activation model with an activation model that allows only N-way confusion or no confusion, and (2) it possesses just one more parameter, $[N(N + 1)]/2$, than the choice model, which allows for a fair comparison between them for the large amount of data reported herein. Hence, for the present experiment, 351 parameters were estimated for the overlap activation model and 350 for the choice model.

Choice Model

The choice model requires less discussion, being a straightforward application of that model as presented by Luce (1963a). As does the overlap activation model, this model postulates an influence of pairwise similarity on stimulus confusions. Both the similarity parameters and bias parameters are assumed to lie on ratio scales. Given a particular stimulus, the probability of any possible response is assumed to be the strength of that

response, as given by multiplying the scale values of the bias and similarity parameters, relative to the total strength relating the given stimulus to other stimuli. For the present study, this results in the theoretical confusion matrix $C = (c_{ij})$, where $i = 1, 2, \cdots, 26$, $j = 1, 2, \cdots, 26$, and

$$c_{ij} = \frac{\eta_{ij} \beta_j}{\sum_{k=1}^{26} \eta_{ik} \beta_k}.$$

In addition, the similarity parameters are symmetric ($\eta_{ij} = \eta_{ji}$), and it is assumed that $\eta_{ii} = 1$ for all i, thus fixing the unit. As indicated above, this results in $[(N(N + 1))/2] - 1$ (350) free parameters to be estimated from the data.

For our purposes, the most important part of Smith's (1968) result relating the threshold to the choice model can be shown readily by viewing the relationship that is assumed to obtain between the confusion matrix probabilities and the similarity parameters:

$$\eta_{ij} = \sqrt{\frac{P(R_i \mid S_j) P(R_j \mid S_i)}{P(R_i \mid S_i) P(R_j \mid S_j)}}.$$

If we now substitute the theoretical expressions for $P(R_i \mid S_j)$ from the threshold model, we obtain

$$\eta_{ij} = \sqrt{\frac{(1 - \sigma_j)p_i \cdot (1 - \sigma_i)p_j}{[\sigma_i + (1 - \sigma_i)p_i] \cdot [\sigma_j + (1 - \sigma_j)p_j]}}$$

$$= \sqrt{\frac{(1 - \sigma_j)p_j}{\sigma_j + (1 - \sigma_j)p_j}} \cdot \sqrt{\frac{(1 - \sigma_i)p_i}{\sigma_i + (1 - \sigma_i)p_i}},$$

and this shows that for the threshold model and the choice model to hold for a given set of data, η_{ij} must be factorable into independent numbers associated with S_i and S_j. Hence, although it can be developed from entirely distinct conceptions, the threshold model can be viewed as a special case, when one neglects underlying assumptions giving rise to the parameters, of the choice model. To see that they are, nevertheless, not the same theory, it suffices to note that for the two-signal case, the threshold model generates linear and the choice model curvilinear isosensitivity curves. Or, to put it another way, if the threshold model provided good fits to several sets of data generated by varying motivational conditions, then only the p_j parameters would change; but by Smith's result, this would perturb not only the bias parameter β_j, but also the similarity parameter η_{ij} which in terms of the theory should remain

constant for constant stimulus and sensitivity conditions.

Similarly, a mapping from the parameters of the choice model to the overlap model or vice versa may be provided. Under some circumstances, this may imply nontestability of the two models against one another. However, the models do not predict identical relationships in the data, and in certain cases data may support one and falsify the other. For example, suppose that $\eta_{ij} = \eta_{kl} = \eta$, for $i \neq j$ and $k \neq l$, and that $\beta_i = \beta_j = \beta$, for all i, j. The mapping from the choice model parameters to the overlap model parameters is then given by

$$\xi_{ij} = P(R_i \mid S_j) + P(R_j \mid S_i)$$

$$= \frac{2\eta}{1 + (N - 1)\eta}, \quad i \neq j,$$

$$\xi_{ii} = \frac{1 - (N - 1)\eta}{1 + (N - 1)\eta},$$

where N is again the number of experimental stimuli. But, for $\eta > 1/(N - 1)$, a reasonable possibility for confusable stimuli, $\xi_{ii} < 0$, which is contradictory to the overlap model. Presently, we know of no corresponding theoretical results against the choice model when we map from the overlap to the choice model. Thus, it may be that the choice model is the more general of the two despite the almost equal number of parameters.

Multidimensional Scaling Model

Although the preceding substantive models possess sufficient structure to describe interletter (sensory) confusability and one (choice model) posits a function relating the similarity scale to distance in a metric space, they cannot be used facilely to test assumptions about the appropriate metric or the number of dimensions in a hypothesized underlying visual space. To this end, a program developed by Kruskal (1964a, b) was employed to obtain information about possible dimensions of confusion, as well as to test applicability of the Euclidean vs the city-block metric to an interletter confusion matrix. Kruskal's method, an extension of Shepard's ordinal scaling procedure (Shepard, 1962a, b), permitted this to be accomplished with a minimum of substantive assumptions.

METHOD

Apparatus

A Gerbrands two-field, two-mode tachistoscope (Model T-2B-1) was used to present the letters. A simple mechanical shutter was employed to increase the

number of independent fields to three.

Stimuli were presented on 8 x 11 in. white index cards, using an IBM Executive Directrix typewriter, one letter per card, with the stimulus population consisting of one deck made up of five English upper-case alphabets in a quasirandom (i.e., shuffled) order. A prestimulus fixation point was placed approximately 10 min below the locus of the letter presented in the stimulus field. A single letter subtended an angle of about 30 min and the fixation point an angle of about 6 min at S's eye. The possible visual fields were white with fixation point, white with letter, and noise. The noise field was produced by typing alternately with "normal" and "expanded" spacing modes on the Executive Directrix typewriter, and successive rows of letters were made to overlap. The characters typed were selected haphazardly from the upper-case English alphabet. Use of the resulting noise field made perception of presented letters much more difficult.

The luminance for CI (without noise), the first experiment, was at all times 5.6 fL, and for CII (with noise), the second experiment, the luminance was 5.6 fL for the prestimulus and stimulus fields and for the noise field was 4.0 fL.

Procedure

The stimulus exposure sequence for CI was: white prestimulus field with fixation point; stimulus field containing a randomly selected letter; poststimulus field containing fixation point. The sequence for CII was: white prestimulus field; stimulus field containing a randomly selected letter; noise field with jumbled letters; white poststimulus field containing fixation point.

The trials were S-paced, S initiating a stimulus exposure following an alerting buzzer sounded by E. The displayed letter followed the press of the start button by 1 sec. The only instruction to S with regard to his response speed was that a letter response was to be made within 5 sec after the stimulus exposure, after which he was to rate his sensory accuracy by giving a confidence rating (CR) of 1 to 4. Ss were instructed that a CR of 1 meant absolute certainty, 2 relative certainty, 3 relative uncertainty, and 4 absolute uncertainty, i.e., guessing at random. Following S's response, E told him the letter that had been presented. Each trial consumed about 15 sec, with approximately 5 sec intervening between feedback and the next buzzer sound. Latencies were recorded by E on each trial from a Hunter timer linked to a voice-operated relay and a microphone positioned close to S's mouth.

Two days of practice and calibration preceded 5 experimental days in CI. A

stimulus duration was selected for each S such that his overall probability correct was about 0.5, that is, at threshold. The selected duration was then used for that S the remainder of CI. Following CI, 1 day was taken to recalibrate Ss for CII. However, performance with the poststimulus noise proved much more variable; therefore, each S was recalibrated every day during the five experimental sessions. Both conditions included 20 warm-up trials at the beginning of each session.

Each experimental condition consisted of calibration followed by five experimental sessions, each session occupying approximately 1 h during which 130 trials of five pseudorandomized (shuffled) alphabets were presented. Each row of the resulting confusion matrices thus contained 150 points for both conditions. The amount of data per individual was insufficient to estimate stability but the group-average confusion matrices appeared quite constant from the first half to the second half of the conditions. All the models assume steady-state behavior is generating the response proportions.

The study required 13 consecutive days, including two weekends. A S was run at the same time each day.

Estimation

In all three models, estimation of the sensory and bias parameters was effected by setting each parameter equal to a function of confusion-matrix values associated with the parameter within a particular theory. The exact formulae used are given in the appendix.

The stimulus durations, chosen to yield probability-correct values close to 0.50, were large enough to allow zero probability of confusion entries in the confusion matrix. This caused a problem in estimation since the functions used to estimate the parameters often contained observed ($P(R_i/S_j)$) terms in the denominator. In order to allow estimation of all relevant parameters, a convention was uniformly adopted across the models. Since each row in the confusion matrix was based on 150 trials, it was decided to take the best estimate of the "zero" confusion probabilities (in terms of sum of squared deviations of predicted from observed values) lying between 0.000001 and 0.01. That is, the computer performed a series of iterations, and on each iteration exactly one value in the above range was tried for all the zero confusion entries. For each such value, the set of parameters was estimated for the model at hand and the sum of squared deviations computed; that value associated with the smallest sum of squared deviations was chosen as

the best estimated overall value of the zero entries. The rationale behind the convention was that if the "true" confusion probabilities were greater than 0.01, then one or more confusions would have occurred on the average in the 150 trials represented in each row of the confusion matrix. On the other hand, if the probability of confusion were less than 0.01 but greater than zero, one would expect to observe cases where the entry for 150 trials turned out to be zero. Thus, in a sense, one extra parameter was estimated for each of the models. The calculated sum of squared deviations varied between the obtained minimum and about 1.50 for the overlap and choice models but varied only a few 100ths for the threshold model. As it turned out, when estimates were also obtained for "zero" confusion probabilities between 0.01 and 1.00, in only one case was the sum of squared deviations reduced and that by only 0.02; this occurred for the choice model in CI.

It should be noted that in addition to the analytical difficulties associated with division by zero, the scaling properties of Luce's choice theory require stronger assumptions when the set of choice alternatives consists of some perfectly and some imperfectly discriminable pairs (Luce, 1959). The overlap activation model and the choice model required estimation of 351 and 350 parameters, respectively, and the all-or-none activation model required estimation of 51 parameters. The number of degrees of freedom associated with the empirical confusion matrix was 650.

Subjects

Six females recruited from introductory psychology classes at the University of Hawaii were employed as Ss. Their vision was required to be 20/20 after correction.

RESULTS AND DISCUSSION

Tables 1 and 2 show the empirical confusion matrices for the two experimental conditions. Table 3 shows predictions from the three models compared with the empirical values for the letter "P"—a symbol possessing curved- as well as straight-line features. The table illustrates the superior ability of the overlap and choice models to reflect similarity as well as the tendency to altered structure from CI to CII.[1] Although the predictions do seem to follow the general pattern of the confusions in the empirical matrix, a numerical index of precision may be helpful in interpretation. An idea of relative accuracy of the three models can be gained from Table 4, which indicates the sum of squared deviations of predicted from empirical points over all 676 cells for each confusion matrix.

Table 1
Empirical Confusion Matrix: Condition 1

Stimulus	Response																									
---	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	.58	.00	.00	.01	.00	.01	.00	.02	.02	.03	.10	.02	.00	.02	.01	.02	.00	.05	.00	.04	.00	.01	.00	.06	.00	.01
B	.02	.26	.02	.05	.02	.00	.04	.07	.01	.02	.01	.01	.03	.08	.03	.03	.03	.18	.01	.01	.06	.01	.00	.01	.00	.00
C	.01	.01	.50	.01	.05	.03	.03	.02	.01	.01	.03	.04	.00	.00	.07	.03	.01	.02	.01	.04	.01	.01	.01	.01	.01	.01
D	.01	.04	.01	.46	.01	.00	.05	.04	.01	.01	.00	.00	.01	.01	.12	.02	.05	.05	.00	.02	.06	.01	.00	.01	.01	.00
E	.01	.03	.03	.01	.36	.07	.00	.03	.03	.04	.06	.11	.00	.00	.03	.05	.00	.01	.00	.07	.01	.01	.00	.01	.03	.01
F	.00	.01	.00	.01	.02	.33	.01	.03	.09	.05	.03	.07	.01	.01	.00	.03	.00	.01	.01	.18	.01	.01	.00	.01	.07	.01
G	.01	.01	.08	.01	.01	.01	.34	.09	.01	.01	.02	.03	.00	.04	.11	.03	.03	.03	.01	.01	.07	.02	.00	.00	.01	.00
H	.01	.01	.01	.01	.00	.01	.01	.50	.01	.01	.01	.01	.03	.15	.03	.02	.00	.04	.00	.03	.04	.01	.03	.00	.01	.00
I	.00	.00	.00	.00	.01	.03	.01	.01	.57	.08	.01	.11	.00	.00	.01	.01	.00	.02	.00	.09	.00	.01	.00	.00	.03	.00
J	.01	.00	.00	.00	.00	.01	.01	.02	.15	.48	.01	.03	.00	.00	.03	.01	.00	.01	.01	.08	.04	.01	.00	.02	.05	.01
K	.03	.01	.01	.01	.01	.02	.00	.06	.05	.01	.50	.04	.01	.02	.04	.03	.00	.03	.00	.01	.01	.01	.01	.05	.03	.00
L	.01	.00	.01	.01	.01	.01	.00	.03	.14	.03	.02	.60	.00	.01	.03	.01	.00	.01	.00	.07	.00	.00	.00	.00	.02	.00
M	.00	.01	.00	.01	.00	.00	.01	.10	.00	.01	.01	.00	.62	.08	.05	.01	.01	.01	.01	.01	.01	.01	.05	.01	.00	.00
N	.03	.01	.00	.01	.00	.00	.01	.06	.00	.00	.03	.00	.07	.54	.03	.01	.03	.04	.01	.01	.02	.02	.04	.01	.01	.01
O	.01	.01	.06	.05	.00	.01	.11	.01	.00	.02	.01	.00	.01	.03	.51	.02	.10	.01	.00	.02	.02	.00	.01	.00	.00	.01
P	.01	.02	.02	.01	.01	.09	.01	.05	.03	.01	.01	.01	.01	.01	.06	.52	.00	.06	.00	.03	.01	.01	.00	.01	.01	.01
Q	.01	.01	.01	.01	.00	.00	.11	.06	.00	.01	.01	.01	.00	.03	.28	.01	.36	.01	.00	.00	.05	.00	.01	.01	.00	.00
R	.00	.04	.01	.00	.01	.01	.01	.09	.02	.01	.03	.03	.02	.05	.03	.05	.00	.49	.01	.02	.01	.01	.01	.03	.01	.00
S	.01	.01	.03	.01	.02	.00	.02	.06	.03	.03	.06	.03	.00	.03	.04	.05	.00	.03	.43	.03	.01	.02	.00	.03	.01	.02
T	.01	.00	.00	.01	.04	.05	.01	.01	.16	.04	.01	.05	.01	.01	.04	.04	.00	.00	.01	.42	.01	.01	.01	.01	.05	.01
U	.00	.00	.01	.01	.00	.00	.01	.11	.01	.00	.01	.02	.02	.05	.07	.01	.01	.01	.00	.02	.55	.04	.04	.01	.00	.01
V	.00	.00	.00	.01	.01	.00	.00	.03	.02	.02	.01	.01	.00	.01	.02	.02	.01	.01	.00	.02	.07	.60	.02	.01	.09	.01
W	.01	.00	.00	.01	.00	.00	.00	.08	.00	.01	.02	.01	.05	.07	.05	.01	.01	.02	.00	.02	.05	.08	.45	.03	.01	.01
X	.01	.00	.00	.00	.01	.00	.00	.03	.02	.01	.07	.01	.00	.04	.02	.01	.00	.00	.00	.05	.00	.03	.01	.55	.08	.04
Y	.01	.00	.00	.00	.00	.01	.00	.05	.02	.03	.03	.03	.00	.03	.06	.01	.00	.00	.01	.06	.01	.08	.01	.03	.51	.01
Z	.01	.01	.01	.01	.01	.00	.00	.03	.01	.06	.03	.00	.00	.00	.02	.00	.01	.00	.01	.03	.01	.01	.01	.05	.03	.66

In order to acquire some intuition on the size of these errors, we may compare the obtained values in Table 4 with a prediction following from random prediction (i.e., on the average, uniform or equiprobable confusions and probability correct) on the part of a model. That is, all confusion entries are simply 1/26. If we let P_{ij} refer to entries in the empirical confusion matrix, then the sum of squared deviations under random prediction would be

$$\sum_{i,j=1}^{26} \left(P_{ij} - \frac{1}{26}\right)^2 = \sum_{i,j=1}^{26} P_{ij}^2 - 1.$$

This index is largest when one value on each row is 1 and the rest are 0; it then attains the value of 25. Table 4 indicates the sum of deviations squared, as well as this (equiprobable) index, for the two conditions. Also, it shows the sum of squared deviations for the main diagonal only and the index figured only on the

main diagonal. This table can be used to discuss several points. First of all, the choice model and the overlap model are about equal in their ability to fit the data for CI, although the overlap model appears a little more accurate in CII. Secondly, it appears that the large increment in number of parameters (in fact, a factor of seven) for the two activation models permitting sensory confusion does not seem to purchase a corresponding increase in fit, relative to the calculated index. In fact, it

Table 2
Empirical Confusion Matrix: Condition 2

Stimulus	Response																									
	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	.83	.01	.00	.01	.00	.01	.00	.01	.01	.01	.01	.02	.00	.00	.02	.00	.00	.00	.02	.01	.00	.03	.00	.00	.00	.01
B	.02	.36	.01	.05	.01	.04	.03	.07	.02	.03	.01	.03	.01	.02	.10	.05	.00	.01	.02	.04	.01	.03	.01	.01	.01	.01
C	.00	.00	.79	.01	.00	.01	.03	.02	.01	.01	.00	.01	.01	.00	.05	.01	.00	.01	.02	.00	.00	.01	.00	.01	.01	.00
D	.01	.01	.01	.73	.01	.01	.00	.03	.00	.01	.00	.00	.00	.01	.11	.02	.03	.01	.01	.00	.01	.01	.00	.01	.00	.01
E	.03	.01	.03	.00	.43	.09	.01	.01	.05	.01	.03	.09	.01	.00	.06	.01	.00	.01	.03	.03	.00	.01	.00	.02	.03	.01
F	.05	.01	.02	.03	.05	.42	.01	.05	.04	.01	.01	.05	.01	.00	.05	.03	.01	.00	.01	.07	.01	.03	.00	.01	.04	.01
G	.00	.00	.11	.01	.01	.01	.57	.04	.00	.01	.01	.00	.00	.01	.12	.02	.01	.01	.00	.00	.03	.03	.00	.00	.00	.01
H	.05	.01	.00	.03	.03	.05	.03	.19	.09	.01	.01	.05	.03	.03	.10	.03	.01	.02	.02	.05	.05	.07	.00	.01	.01	.03
I	.05	.01	.01	.02	.01	.03	.01	.07	.38	.03	.00	.12	.00	.01	.07	.01	.00	.01	.01	.05	.01	.05	.00	.00	.02	.01
J	.03	.00	.00	.02	.01	.03	.01	.03	.08	.49	.01	.03	.00	.00	.06	.01	.01	.00	.02	.05	.01	.06	.01	.00	.01	.02
K	.03	.00	.00	.00	.00	.03	.00	.05	.01	.01	.61	.05	.01	.00	.04	.01	.00	.01	.00	.02	.01	.01	.00	.06	.03	.02
L	.03	.01	.03	.02	.02	.02	.01	.03	.04	.00	.02	.59	.02	.00	.07	.01	.00	.00	.01	.03	.01	.00	.01	.00	.02	.00
M	.09	.01	.01	.01	.01	.01	.04	.06	.04	.03	.01	.03	.41	.05	.08	.02	.00	.01	.00	.00	.00	.05	.01	.01	.01	.03
N	.04	.00	.01	.02	.01	.02	.01	.05	.05	.01	.01	.05	.01	.39	.09	.01	.01	.01	.01	.07	.01	.05	.02	.00	.03	.00
O	.01	.00	.01	.02	.00	.01	.07	.01	.01	.00	.00	.01	.00	.00	.72	.02	.06	.00	.02	.01	.00	.01	.00	.01	.00	.01
P	.00	.02	.01	.02	.01	.01	.00	.02	.03	.01	.01	.02	.00	.00	.05	.67	.01	.02	.03	.01	.01	.01	.00	.01	.02	.01
Q	.02	.00	.03	.00	.00	.02	.02	.01	.01	.00	.00	.00	.00	.00	.32	.01	.52	.01	.03	.01	.00	.00	.00	.00	.00	.01
R	.02	.01	.01	.03	.00	.01	.00	.08	.01	.00	.01	.03	.01	.00	.03	.13	.01	.52	.01	.03	.01	.02	.00	.00	.03	.01
S	.03	.01	.01	.01	.00	.00	.03	.07	.01	.01	.00	.01	.00	.00	.01	.00	.00	.01	.78	.01	.00	.01	.00	.00	.00	.01
T	.01	.00	.01	.01	.01	.03	.01	.01	.08	.03	.01	.05	.00	.01	.03	.01	.00	.01	.01	.58	.01	.03	.00	.00	.03	.02
U	.03	.02	.01	.01	.01	.00	.01	.03	.05	.02	.02	.01	.01	.00	.01	.03	.00	.00	.04	.01	.53	.05	.00	.02	.01	.01
V	.02	.01	.01	.01	.01	.01	.00	.01	.02	.00	.00	.01	.01	.01	.05	.01	.00	.00	.01	.01	.00	.69	.01	.02	.05	.01
W	.05	.01	.01	.01	.00	.03	.01	.05	.01	.02	.00	.02	.02	.03	.09	.04	.00	.01	.03	.02	.03	.37	.09	.03	.03	.01
X	.03	.01	.00	.00	.00	.01	.01	.06	.03	.00	.06	.04	.00	.00	.03	.01	.00	.00	.01	.03	.00	.07	.01	.47	.09	.03
Y	.01	.00	.01	.01	.01	.01	.00	.04	.03	.00	.02	.04	.01	.01	.07	.01	.00	.02	.03	.04	.00	.22	.00	.01	.36	.03
Z	.01	.00	.00	.00	.00	.01	.00	.00	.01	.01	.00	.01	.01	.00	.01	.01	.00	.00	.01	.01	.00	.01	.00	.00	.03	.88

is possible to estimate how much the choice and overlap model contribute with their extra 199 parameters via their similarity structure as compared with free assignment of parameters to those 199 cells in the threshold confusion matrix associated with the largest deviations from the empirical matrix. That is, we simply insert the observed confusion probabilities in those specified cells, recompute the sum of squared deviations, and contrast this number with the sum of squared deviations for the other two models. We find that the theoretically assigned parameters account for approximately half of the error reduction (.44 for CI, .49 for CII) produced by this discounting of the 199 largest deviations. As might also be expected from Table 4, the off-diagonal entries were disproportionately represented in the lists of 199 largest squared deviations for the threshold model. Thus, the all-or-none model may be more efficient in this sense. Thirdly, it is apparent that the overlap model found greatest difficulty on the main diagonal, the all-or-none with the off-diagonal cells, and the choice model distributed the error between the diagonal and off-diagonal cells more than the other two models. One reason the all-or-none activation model performs as well as it does may therefore be due to its potential for fitting the diagonal values associated as they are with greater possibility of variation.

This last result suggests that an activation model that included the all-or-none and the overlap might describe much of the structure of the data:

Table 3
Representative Theoretical Predictions

Condition 1
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
P	EMP	.01	.02	.02	.01	.01	.09	.01	.05	.03	.01	.01	.01	.01	.01	.06	.52	.00	.06	.00	.03	.01	.01	.00	.01	.01	.01
	AON	.02	.02	.02	.02	.02	.02	.02	.04	.02	.02	.02	.02	.02	.02	.02	.53	.02	.02	.02	.02	.02	.02	.02	.02	.02	.02
	OLP	.01	.02	.02	.01	.02	.04	.01	.04	.02	.01	.02	.01	.01	.01	.05	.56	.00	.05	.01	.04	.01	.02	.00	.01	.01	.00
	CHC	.01	.01	.02	.01	.02	.03	.01	.05	.02	.01	.02	.01	.01	.01	.05	.57	.00	.06	.00	.05	.01	.01	.00	.01	.01	.00

Condition 2
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
P	EMP	.00	.02	.01	.02	.01	.01	.00	.02	.03	.01	.01	.02	.00	.00	.05	.67	.01	.02	.03	.01	.01	.01	.00	.01	.02	.01
	AON	.01	.01	.01	.01	.01	.01	.01	.02	.01	.01	.01	.01	.01	.01	.04	.68	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
	OLP	.00	.02	.01	.02	.01	.02	.01	.02	.02	.00	.01	.02	.00	.00	.05	.64	.01	.05	.02	.01	.01	.01	.01	.00	.01	.01
	CHC	.00	.01	.01	.02	.01	.02	.00	.03	.02	.00	.01	.02	.00	.00	.06	.70	.00	.03	.00	.01	.01	.02	.00	.00	.01	.01

EMP = empirical. AON = all-or-none activation model. OLP = overlap activation model. CHC = choice model

Table 4
Sum of Squared Deviations of Theoretical from Empirical Points

Condition	Model			
	Choice	All-or-None	Overlap	Equi-probable
	Entire Confusion Matrix			
CI	.23	.49	.26	6.97
CII	.36	.49	.23	8.14
	Main Diagonal			
CI	.07	.02	.18	5.98
CII	.14	.06	.18	7.84

$$P(R_j/S_i) = \sigma_{ij} \frac{g_j}{g_i + g_j} + \left(1 - \sum_{k=1}^{N} \sigma_{ik}\right) g_j, \quad i \neq j,$$

and

$$P(R_i/S_i) = \sigma_{ii} + \sum_{k \neq i}^{N} \sigma_{ik} \frac{g_i}{g_i + g_k} + \left(1 - \sum_{k=1}^{N} \sigma_{ik}\right) g_i.$$

Two subcases of this model of interest could be extruded by (1) letting all σ_{ik} be equal but allowing the σ_{ii} to be different (this case would have just one more parameter than the all-or-none model), or (2) letting all the σ_{ii} be equal but allowing the σ_{ij} to be different. This case would have just one more parameter than the choice models and the same as the overlap model. These cases should give some idea of the relative importance of uniqueness of

the probability of the certainty states vs the uniqueness of the probabilities of the sensory confusion states. Too, they would perhaps aid in the interpretation of the significance of the number of parameters per se. For example, suppose the first subcase above fit no better than the all-or-none model and the second subcase fit as well as but not better than the overlap model. This would provide an indication that the number of parameters might have more influence than the particular structure of the models. Another possibility suggested by Smith (1968) might be to fit probability mixtures of any two activation models or even an activation model and the choice model, i.e., aMI + (1 – a)MII, where MI refers to Model I and MII to Model II and a is less than 1 and greater than 0. To the extent that the sum of errors squared is independent of the value of a are the two models explaining the same aspects of the data. Estimation techniques and computer programs are currently being developed to allow fits of these subcases to the data.

One interesting outcome was that CII was not better fit by the all-or-none model than was CI. The absence of fading afterimages undegraded by noise does not seem to have diminished the sensory confusability of the stimuli. It appears that the combination of the increase of stimulus duration necessary for adequate performance in CII and the effects of the mask did not radically affect the processing characteristics of the Ss, at least with reference to the fitting ability of the models tested here. It does appear that the choice model predicts less successfully in CII, although it is not clear why.

Figures 1a and 1b show the curves for a fairly typical S relating, in the first part, probability correct to latency and, in the second, probability correct to confidence rating. As the figures indicate, there is a continuous decrease in probability correct as latency increases for both conditions. Such a relationship between accuracy and latency has also been found to hold for multisymbol processing tasks (Estes & Wessel, 1967; Townsend, 1966), and suggests that the phenomenon may be connected with the same mechanisms in both cases, perhaps with the processing characteristics of individual symbols. Other characteristics, shown in Fig. 1a, are the longer mean latency and standard deviations for CI than for CII. This is probably connnected with the fact that after recalibration on CII, the Ss' average performance was better than on CI. To be sure, performance actually decreased in CII on letters H, I, M, N, W, X, and Y, indicating again a possible shift in confusion structure, perhaps due to the presence of the poststimulus noise mask. Only one S did not exhibit this pattern; he had a larger mean in CI but a smaller standard deviation.

Figure 1b shows the typical relationship of probability correct to confidence rating. Although this result has been taken as indicative of the continuity of perceptual states (as, in fact, it may be), Krantz (1969) and Wickelgren (1968) have pointed out, in essence, that one must consider the S's processes that evaluate his perceptual states and lead to the distribution of confidence ratings found in the data.

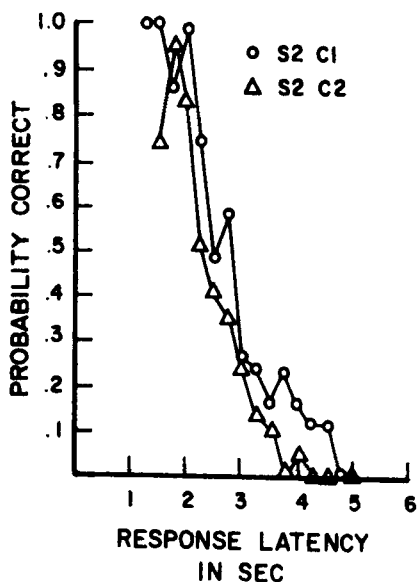In addition to acting as parameters of

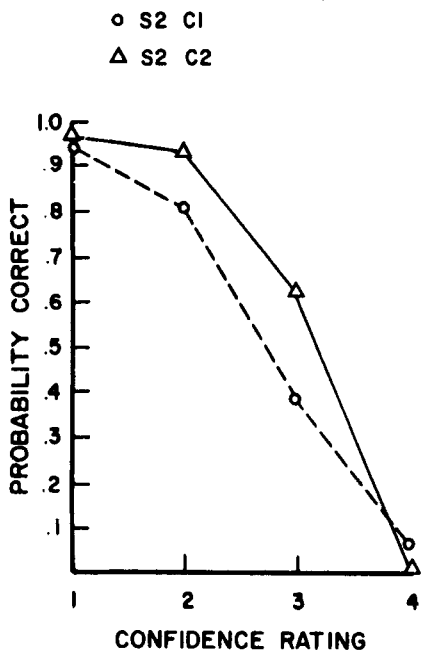Fig. 1a. Probability correct as a function of latency for a typical S.

Fig. 1b. Probability correct as a function of confidence rating for a typical S.

similarity in the choice model, η_{ij} as estimated by the formula has been suggested as possessing several properties appropriate for a measure of similarity (Shepard, 1958); for this reason, the resulting values were used as input data for the multidimensional scaling analysis. The present application of Kruskal's (1964b) program included two fit techniques as well as fitting the Euclidean and city-block metrics. One technique, referred to as the primary approach, allows two pairs of

letters having "tied" similarity scores to have different distances in psychological space without penalty in terms of stress. The other, the secondary approach, augments the stress value when tied pairs of letters fail to be associated with equal distance.

As can be seen in Figs. 2a and 2b, in general the Euclidean metric is superior to the city-block metric, the primary approach is superior to the secondary approach, and CI is more easily fit into a multidimensional space than is CII. The exception to this is CII, r = 1, which, except for spaces spanning three and two dimensions, is associated with lower stress values than is CII, r = 2. The reason for this exception is not clear at present.

Each of the eight stress curves in Fig. 2a reaches Kruskal's (1964a, b) criterion of .10 in less than 10 dimensions, and rather nice "elbows" are present. Also, from Klahr's (1969) recent results we can also infer the "significance" of at least the primary curves. For example, in 50 sets of randomly generated sets of 16 points, 5% of the cases were associated with stress values of .170 or less for three dimensions when fit with the Kruskal program. We would expect this result in even fewer cases with our greater number of points (26). Nevertheless, there was some inconsistency of spatial configurations for the two experimental conditions, for the two metrics, and even of the primary as opposed to the secondary approach. 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 either are very elementary and obvious or unless they turn out to be equivalent to some previously hypothesized dimensions. From examination of letters having relatively large projections on various dimensions one could infer the importance of direction of lines and angles, roundness, and vertical linearity, but, again, these did not appear in the same form for the different conditions and approaches, and the letters associated with one of these attributes were not always the same or even completely consistent with the attribute.

The difference in the ability of CI and CII to be fit into a multidimensional space may partially follow from the recalibration procedure carried out in CII. To take an extreme example, if a S's performance varied because of daily fluctuations in attention, equal accuracy, as given by the recalibration, may not have implied the same underlying visual space. The confusion matrix for CII would then be a composite of several underlying spaces.
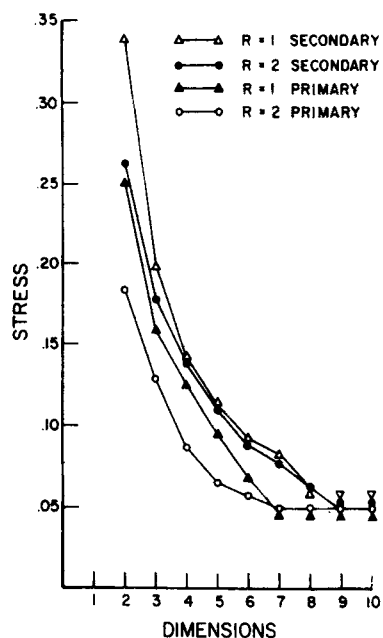

Fig. 2a. Kruskal measure of fit as a function of number of dimensions in the space, Condition I.
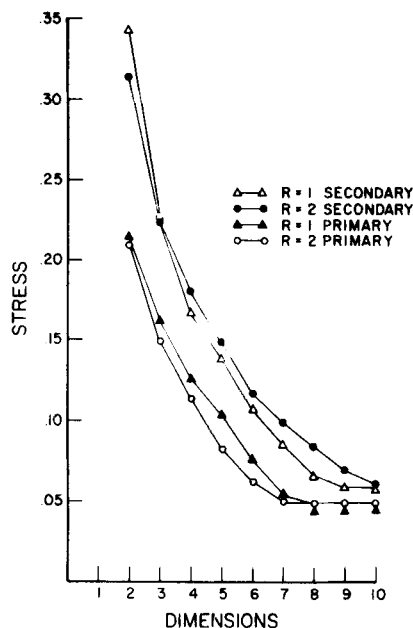

Fig. 2b. Kruskal measure of fit as a function of number of dimensions in the space, Condition II.

Two of the possible dimensions mentioned above, roundness and vertical linearity, correspond to two of the dimensions reported by Künnapas (1966). Nevertheless, despite this finding and despite the hopeful outcome of Brown and Andrews's (1968) comparison of the scaled space of a discrimination and judgment task, it should not be surprising if results should differ in the Künnapas and the

Model		OLP CII	CHC CI	CHC CII	AON CI	AON CII
AON	CI					.48
OLP	CI	.53	.88		.53	
	CII			.89		.83
CHC	CI			.23	.70	
	CII					.60

* See text for explanation of headings.

present type of studies; one demands conscious evaluation of perfectly perceptible stimuli and the other immediate identification of severely degraded images produced by tachistoscopic exposure.

The finding that a Euclidean metric is generally most appropriate for the present data may at first seem bizarre. Torgerson (1958) and Shepard (1964), for example, have suggested that certain types of visual dimensions may be conducive to the appearance of a city-block metric, such dimensions being "perceptually distinct" and "compelling." Letters, such as "O" and "I," for instance, may seem to be made up of perceptually distinct types of elements. There are several possibilities that might be related to this finding. Some of the structure that appears to be independent may actually be related. For instance, the extent of linearity of, say, an ellipse is inversely related to its degree of roundness, and perceptual blur due to tachistoscopic conditions might produce several gradations of linearity, thus producing perceptual dimensions better suited to a Euclidean metric. Also, of course, since letters are made up of several kinds of elements, this would tend to promote perceptual continuity among the letters themselves. A third possibility is that restricted (in time) observations such as those in the present study could lower the probability of sampling information from several dimensions simultaneously and thus improve the fit of a Euclidean metric relative to the city-block metric (e.g., see the discussion by Hyman & Well, 1967). It should be mentioned that the lack of clear-cut results relating to particular dimensions of similarity and perhaps even the support of a Euclidean metric may be due to an artifact caused by scaling group data, especially if Ss employ different dimensions in their identification process. An experiment is in progress that includes the gathering of long-term data for individual Ss and the application of the present analyses to them.

In an effort to obtain a rough idea of the extent to which the various models were reflecting the same underlying structure in the data, correlation coefficients were obtained for analogous parameters in the three models, and these are shown in Tables 5 and 6 for bias and similarity parameters, respectively. In Tables 5 and 6, AON = all-or-none activation model, OLP = overlap activation model, and CHC = choice model. The two conditions are given by CI and CII. Table 6 includes correlations with and among two possible metrics and a crude physical measure of interletter similarity. The metrics are ones suggested by Luce (1963a) to be associated with the choice model (d_c) and the other is that obtained from application of the scaling program (d_s). The measure of physical overlap or similarity (PHOVLP) was found by taking blown-up versions of the font used in the experiment and superimposing them in pairs on a 10 by 10 grid (the letters were eight units high and about five units wide, on the average). Then, the relative intersection was computed for each pair, the intersection divided by the union of cells covered in the matrix.

Bearing in mind that we usually interpret the correlation coefficient r or r^2 as referring to a linear relationship, the obtained values are surprisingly large. Within the bias table, the correlations within a condition are higher than across conditions; this may indicate an actual change of biases by Ss from CI and CII. Also, the highest correlations are obtained between the overlap activation model and the choice model, suggesting that despite their differing heritage, these two models reflected bias processes in a similar way, and different from the all-or-none model, by virtue of their ability to depict stimulus similarity.

Among the similarity parameter correlations, the most salient result was the remarkable r of .97 of the choice model in the two conditions, as contrasted with the lowest correlation of biases for the choice model across conditions, which, of course, we would expect if bias changed and stimulus similarity did not, and a model reflects these properties. To be sure, the differences in the scaling fits indicate that, at the least, the psychological spaces may not have been identical for the two conditions. When examining the correlations with the supposed metrics, we may note, as a reference, the correlation of about .75 obtained by correlating η_{ij} with $-\ln(\eta_{ij})$ (dc). The overlap parameters correlated slightly higher with $-\ln(\eta_{ij})$ than they do with η_{ij}. Also, a favorable relationship is found between the metric $-\ln(\eta_{ij})$ and the metric from the multidimensional scaling program. It is interesting to note that if the relationship between η_{ij} in CI and CII were by way of a similarity transformation (y = ax, "a" constant) as choice theory demands if the similarity structure is the same in CI and CII, then the correlation coefficient for $-\ln(\eta_{ij})$ for the two conditions would also be large. The fact that it is not (r = .63) tells us that the relationship must have been affine (y = ax + b, "a," "b" constants) rather than strictly simply a stretch or shrinkage relationship. This probably followed from fewer estimated similarities of zero in CI than in CII. Of course, conventional application of the theory would assume no zero similarities.

Concluding the correlation analysis is that of the physical overlap measure, a measurement that does not suffer from the built-in dependencies that characterize the other numbers in the correlation tables. The results are again substantially in the expected direction with the highest correlations found with the choice similarity parameters and with the scaling metric.

The approach assumed in this paper has been of a twofold nature. On the one hand, mathematical models were employed as investigatory structures to reveal and explicate various aspects of the data; the substantive recognition models and the scaling technique were complementary in this aim. On the other hand, there was a theory-testing facet of the study which

Model		OLP CII	CHC CI	CHC CII	d_c CI	d_c CII	d_s CI	d_s CII	PHOVLP
OLP	CI	.58	.43		−.67		−.61		.48
	CII			.64		−.67		−.54	.53
CHC	CI			.97	−.75		−.70		.70
	CII					−.76		−.65	.65
d_c	CI					.63	.86		−.69
	CII							.75	−.58
d_s	CI							.57	−.67
	CII								−.54

* See text for further explanation.

especially manifested itself in the evaluation and comparison of the recognition models. Although this approach proved fruitful in delineating some of the characteristics of alphabetic confusion and suggesting the type of model necessary to deal with visual confusions, the spatial configurations resulting from the scaling analysis were less than satisfactory, in terms of describing a psychologically intuitive set of dimensions. To some extent, this may be a function of the sheer complexity of the English alphabet; it may prove valuable to build up a psychophysics of simple visual elements, such as lines at various orientations, and then begin to construct more complex alphabets with these. In this way the laws relating to the emergent properties of the more complex stimuli may become known.

APPENDIX

Although there exist maximum likelihood estimation procedures for the choice model and the all-or-none model, for convenience and homogeneity of estimation technique, all parameters were estimated by setting each parameter equal to a function of the data [i.e., the modified method of moments (Bush, 1963)], the particular function, of course, being related to the particular model and the specific parameter. For simplicity, we will here denote $P(R_j/S_i)$ as P_{ij}.

All-or-None Activation Model

It will be recalled that there exist no similarity parameters as such in this model, only stimulus parameters reflecting the likelihood of obtaining sensory information, which is related in a one-one fashion to the (correct) response (o_i). In addition, if the S does not find himself in this certain state, he must guess from all 26 alternatives and he picks alternative "i" with probability p_i. This model possesses a total of 51 parameters for this experiment as compared with 650 degrees of freedom in the data. The estimates for the sensory parameters were:

$$\hat{o}_i = \left\{ P_{ii} \quad (1 - P_{ii}) \left[\frac{\frac{N}{2}}{\sum\limits_{\substack{j=1\\ \neq i}}^{N} \sum\limits_{\substack{k=1\\ \neq i,j}}^{N} \left(\frac{P_{jk}}{P_{jl}}\right)} \right] \right.$$
$$\left. + \sum\limits_{\substack{l=1\\ \neq i}}^{N} \left[1 - \frac{P_{il}}{\frac{\frac{N}{2}}{\sum\limits_{\substack{j=1\\ \neq i,j}}^{N} \sum\limits_{k=1}^{N} \frac{P_{jk}}{P_{jl}} + N \quad 2}} \right] \right\} \frac{1}{N}$$

$$(i = 1, 2, \cdots, N)$$

where, as before, N is the number of stimuli and responses in an experiment. The estimates for the guessing bias parameters were given by:

$$\hat{p}_i = \frac{\frac{N}{2}}{\sum\limits_{\substack{j=1\\ \neq i}}^{N} \sum\limits_{\substack{k=1\\ \neq i,j}}^{N} \frac{P_{jk}}{P_{ji}} + N \quad 2} \quad (i = 1, 2, \cdots, N).$$

Substitution of the theoretical expressions for the confusion matrix entries, P_{ij}, will suffice to show that the above are the appropriate formulae.

Overlap Activation Model

In this model, structure existed to measure or reflect pairwise stimulus similarity. The parameters assigned this function we called ξ_{ij}, designating the overlap or similarity of stimulus i and stimulus j. To denote the probability that "perfect" sensory information was obtained, we used the symbol ξ_{ii}. As this model included no pure guessing state, only perfect information and pairwise confusion states, we employed the assumption that when in a confusion state, say between S_i and S_j, the S gave response R_i with probability $g_i/(g_i + g_j)$, where g_i is the response strength or response bias (like p_i in the all-or-none activation model); this implies the constant ratio rule on the guessing biases.

The expressions which yielded the sensory parameters for this model were:

$$\hat{\xi}_{ij} = P_{ij} + P_{ji}. \quad (i,j = 1, 2, \cdots, N, i \neq j)$$

$$\hat{\xi}_i = P_{ii} \quad \sum\limits_{\substack{j=1\\ \neq i}}^{N} P_{ji}.$$

and the guessing biases were obtained from the formula:

$$\hat{g}_i = \frac{1}{\frac{1}{N} \sum\limits_{j=1}^{N} \frac{P_{ij}}{P_{ji}}} \quad (i = 1, 2, \cdots, N).$$

Choice Model

The choice model and the relevant formulae are well known; hence, we shall simply list them and remark that this model possesses the same number of similarity and guessing parameters less one as the overlap model for a total of 350. The similarity parameter functions are:

$$\hat{\eta}_{ij} = \left[\frac{P_{ij}P_{ji}}{P_{ii}P_{jj}} \right]^{1/2} \quad (i, j = 1, 2, \cdots, N)$$

and the bias parameters are estimated by:

$$\hat{\beta}_j = \frac{1}{N} \sum\limits_{k=1}^{N} \left[\frac{P_{jj}P_{kj}}{P_{jk}P_{kk}} \right]^{1/2}.$$

It may be observed that instead of selecting a single letter to use in estimating all the bias ratios, in order to possibly give more stability to the estimates, for each bias parameter, each letter was used as the denominator and the average taken. In terms of the model, this simply results in an overall multiplication of the β_j by the constant

$$\frac{1}{N} \left[\sum\limits_{k=1}^{N} \frac{1}{\beta_k} \right]$$

which, of course, cancels out in the expressions for P_{ij}.

REFERENCES

ATKINSON, R. C., & KINCHLA. R. A. A learning model for forced-choice detection experiments. British Journal of Mathematical & Statistical Psychology, 1965, 18, 183-206.

BROWN, D. R., & ANDREWS, M. H. Visual form discrimination: Multidimensional analyses. Perception & Psychophysics, 1968, 3, 401-406.

BUSH, R. R. Estimation and evaluation. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), Handbook of mathematical psychology. Vol. 1. New York: Wiley, 1963. Pp. 429-469.

CERNOG, D. Y., & ROSE, F. C. (Eds.), Legibility of alphanumeric characters and other symbols. II. A reference handbook. Washington, D.C: National Bureau of Standards Miscellaneous 262-2, 1967.

ESTES, W. K., & WESSEL, D. L. Reaction time in relation to display size and correctness of response in forced-choice visual signal detection. Perception & Psychophysics, 1966, 1, 369-373.

HYMAN, R., & WELL, A. Judgments of similarity and spatial models. Perception & Psychophysics, 1967, 2, 233-248.

KINCHLA, R. A., TOWNSEND, J. T., YELLOTT, J. I., & ATKINSON, R. C. Influence of correlated visual cues on auditory signal detection. Perception & Psychophysics, 1966, 1, 67-73.

KLAHR, D. A Monte Carlo investigation of the statistical significance of Kruskal's nonmetric scaling procedure. Psychometrika, 1969, 34, 319-330.

KRANTZ, D. H. Threshold theories of signal detection. Psychological Review, 1969, 76, 308-324.

KRUSKAL, J. B. Multidimensional scaling by optimizing goodness-of-fit to a nonmetric hypothesis. Psychometrika, 1964a, 29, 1-28.

KRUSKAL, J. B. Nonmetric multidimensional scaling: A numerical method. Psychometrika, 1964b, 29, 115-130.

KÜNNAPAS, T. Visual perception of capital letters. Scandinavian Journal of Psychology, 1966, 7, 189-196.

LUCE, R. D. Individual choice behavior. New York: Wiley, 1959.

LUCE, R. D. Detection and recognition. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), Handbook of mathematical psychology. Vol. 1. New York: Wiley, 1963a. Pp. 103-189.

LUCE, R. D. A threshold model for simple detection experiments. Psychological Review, 1963b, 70, 61-79.

NAKATANI, L. H. A confusion-choice stimulus recognition model applied to word recognition. Technical Report No. 31, 1968. Human Communication Laboratory, University of California, Los Angeles.

SHEPARD, R. N. Stimulus and response generalization: Deduction of the generalization quotient from a trace model. Psychological Review, 1958, 65, 242-256.

SHEPARD, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function: I. Psychometrika, 1962a, 27, 125-140.

SHEPARD, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function: II. Psychometrika, 1962b, 27, 219-246.

SHEPARD, R. N. Attention and the metric structure of the stimulus. Journal of Mathematical Psychology, 1964, 1, 54-87.

SMITH, K. Models of confusion. Paper delivered to Psychonomic Society, St. Louis, 1968.

TORGERSON, W. S. Methods of scaling. New York: Wiley, 1958.

TOWNSEND, J. T. Choice behavior in a cued-recognition task. Technical Report

No. 103, 1966, Institute for Mathematical Studies in the Social Sciences, Stanford University.

WICKELGREN, W. A. Testing two-state theories with operating characteristics and a posteriori probabilities. Psychological Bulletin, 1968, 69, 126-131.

NOTE

1. The complete set of theoretical predictions may be obtained from the American Documentation Institute.