Perception & Psychophysics
1973, Vol. 14, No. 3, 471-482

Feature lists and confusion matrices

L. H. GEYER*†
Human Factors Program

and

C. G. DeWALD*
Operations Research Program, Department of Industrial Engineering
State University of New York at Buffalo, Buffalo, New York 14214

This report compares three feature list sets for capital letters, previously proposed by different investigators, on the ability of each to predict empirical confusion matrices. Toward this end, several variants of assumed information processes in recognition were also compared. The best model incorporated: (1) variable feature retrieval probabilities, (2) a goodness-of-match lower threshold below which guessing determines response, and (3) response bias on guessing trials. This model, when combined with one particular proposed feature list set, produced stress values of less than 9% in comparisons to empirical matrices for each of three different Ss. The feature retrieval probability vectors associated with these minimum-stress predictions were highly correlated ($\bar{r}$ = .83), suggesting considerable generality of process and feature sets between Ss.

In the view of many cognitive psychologists, perception is understood as an active process of synthesis (Neisser, 1967; Gibson, 1969). Further, the role of distinctive attributes, commonly termed features, is considered fundamental both to perception and to ongoing information processing in recognition. Neisser (1967, Chap. 3) presents an excellent review of the arguments favoring the view that "... recognition is mediated, in part, by a hierarchy of 'feature analyzers' [p. 46]." Indeed, the concepts of feature lists and list processing are ubiquitous sources of theory for cognitive psychologists who are information-processing oriented, a status made clear in Frijda's (1972) constructive review of processes required to simulate long-term memory.

However, until some specific set of feature lists has been shown to aid in understanding recognition data for some clearly defined class of stimuli, the feature list remains an unsupported hypothetical construct in recognition theory. This potential contribution as an exemplar increases the intrinsic interest in the discovery of the feature set people use to recognize the letters of the alphabet. The attempt to define the features of capital letters has received attention from several investigators, most notably Gibson and her associates (Gibson, 1969; Gibson, Osser, Schiff, & Smith, 1963; Gibson, Gibson, Pick, & Osser, 1962; Yonas & Gibson, 1967), but also Geyer (1970) and Laughery (1971). The bases for the feature sets proposed by these investigators is reviewed later; for the moment, it suffices to say that each of the three contains substantial degrees of judgment, clearly needing independent empirical support.

What constitutes empirical support for a proposed feature set? Gibson (1969) reviews three approaches she and her associates have taken: (1) correlation of confusion errors and features in common between letter pairs; (2) a multidimensional scaling analysis of the same confusion error data, and (3) comparisons of latencies for same/different judgments for letter pairs with features in common between the pair. The multidimensional scaling analysis was reported only out to three dimensions (features), and even there appeared to be bogging down in the familiar difficulty that the dimensions isolated failed to be namable (i.e., were not identifiable as features). The use of latency data, while ingenious, introduces a host of explicit and implicit assumptions that becloud the question of how much support a given analysis provides. The first approach, correlation of confusion errors with features in common, directly engages the assumed underlying processes while still facilitating unambiguous empirical comparisons. The strongest support for a feature list set would be correspondence between the full alphabet confusion matrix, generated on the basis of feature list similarities/differences, and empirical confusion matrices. Fortunately, the empirical data is already at hand for such comparison. Townsend (1971a, b) reports a study that provides full 26-letter confusion matrices eminently suitable to this purpose. During a series of tachistoscopic presentations of capital letters, he limited stimulus duration such that, on the average, correct recognition occurred half the time. He presented each of the 26 capital letters 25 times to each of the six Ss to obtain a 150 trials/row "average S" confusion matrix, then 150 times to each of two (different) Ss to obtain two 150 trials/row individual S confusion matrices. This report analyzes how effective each of the three aforementioned sets is in predicting Townsend's full alphabet confusion matrix data, i.e., how well the confusion matrix generated by feature list similarities/differences corresponds to his data.

*We are grateful to J. T. Townsend for his commentary upon an earlier draft and to I. Pollack for helpful editorial guidance.

†Now at Northeastern University, Boston, Massachusetts 02115.

Table 1
Gibson Feature Set and Lists

Feature Description	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
STRAIGHT																										
1 HORIZONTAL	1				1	1	1	1				1								1						1
2 VERTICAL		1		1	1	1		1	1		1	1	1	1		1		1		1					1	
3 DIAGONAL (/)	1										1		1									1	1	1	1	1
4 DIAGONAL (\)	1										1		1	1			1	1				1	1	1	1	
CURVE																										
5 CLOSED		1		1											1	1	1	1								
6 OPEN, VERTICAL										1											1					
7 OPEN, HORIZONTAL			1				1			1									1							
8 INTERSECTION	1	1			1	1		1			1					1	1	1		1				1		
REDUNDANCY																										
9 CYCLIC CHANGE		1			1								1						1				1			
10 SYMMETRY	1	1	1	1	1			1	1		1		1		1						1	1	1	1	1	
DISCONTINUITY																										
11 VERTICAL	1					1		1	1		1		1	1		1		1		1					1	
12 HORIZONTAL					1	1						1								1						1

An inherent facet of this study, therefore, is the explicit definition of the information processes employed in recognition by feature analysis. The approach taken was to investigate several alternatives, each differing moderately in detailed ways, but similar in general structure. The intent of this approach was to permit the exploration of a full range of processing assumptions, so that a failure of any proposed feature set to predict empirical data could be attributed to the inadequacy of the features and not to feature processing assumptions. This approach mandated that, instead of one predicted confusion matrix for each set of feature lists, it would be necessary to generate a (small) number of related matrices.

This report is organized as follows: The next section discusses the three feature sets studied, and the following section defines the information processing assumptions and the specific alternatives employed. Successive sections describe methodology, present the results, and discuss the findings.

FEATURE SETS REPRESENTING CAPITAL LETTERS

Gibson's Set

Gibson (1969) discussed the basis for selecting a feature set: "In selecting a list of features, the experimenter's intuition is the principal generator, but it can be assisted by evidence from experimental literature and by meeting ... the following criteria: (1) the features had to be critical ones, present in some members of the [grapheme] set but not in others so as to present a contrast; (2) they should be relational so as to be invariant under brightness, size, and perspective transformations; (3) they should yield a unique pattern for each grapheme; and (4) the list should be reasonably economical [pp. 86-87]." She proposes a set of 12 features to represent capital letters (Table 5-5, p. 88), reproduced here as Table 1. She notes that the rationale for selection includes: (1) the neurological research of Maturana, Lettvin, McCulloch, and Pitts (1960) with frogs, and Hubel and Wiesel (1962) with cats; (2) developmental studies with the human (Gibson, Gibson, Pick, & Osser, 1962; Piaget & Inhelder, 1956); and (3) stabilized retinal image research (Pritchard, Heron, & Webb, 1960). She reports supportive evidence from studies designed to test the adequacy of the proposed feature set, but none of these studies rigorously specified the information processes presumed to underlie recognition by feature analysis. Hence, comparisons to empirical data were limited to 26 letter-pair correlations rather than to the full 676 (26 x 26) comparisons of a confusion matrix. Also, since the confusion data were obtained for children, there were no comparisons based on Ss with fully developed reading capabilities. In contrast, such comparisons are a major objective of this study.

Laughery's Set

Laughery (1971) reported a theory of short-term memory (STM) in the form of a computer simulation model. Stimuli were presented to the model as the list of features, auditory or visual, he assumed represented individual capital letters. Inherent to the methodology of simulation, he did specify rigorously the information processes assumed to underlie recognition by feature analyses. His assumptions and those of this investigation are compatible. However, he ascribed a predominately

Table 2
Laughery Feature Set and Lists

Feature Description	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
1 VERTICAL, LEFT		1		1	1	1		1			1	1	1	1		1		1								
2 VERTICAL, CENTER									1											1					1	
3 VERTICAL, RIGHT								1					1	1												
4 HORIZONTAL, TOP					1	1			1	1										1						1
5 HORIZONTAL, MIDDLE	1				1	1	1	1																		
6 HORIZONTAL, BOTTOM					1				1			1														1
7 FULL SLANT, POSITIVE	1																					1	1	1		1
8 FULL SLANT, NEGATIVE	1													1								1	1	1		
9 FULL CURVE, CLOSED															1		1									
10 NOT USED																										
11 FULL CURVE, OPEN, RIGHT			1				1																			
12 FULL CURVE, OPEN, LEFT				1																						
13 FULL CURVE, OPEN, TOP										1											1					
14 HALF CURVE, OPEN LEFT, TOP	1															1		1								
15 HALF CURVE, OPEN RIGHT, TOP	1																		1							
16 HALF CURVE, OPEN LEFT, BOTTOM																			1							
17 PART SLANT, NEGATIVE											1		1				1	1					1		1	
18 INTERSECTION																	1							1		
19 TWO SLANTS, PARALLEL																							1			
20 PART SLANT, POSITIVE											1		1										1		1	

auditory or phonemic basis to STM; thereby, his work did not directly test the capability of his visual feature list to predict confusion errors in a recognition paradigm. (Phonemic feature lists are outside the scope of this report.) Laughery's visual feature lists covered both capital letters and arabic numerals, a condensed version employing only the features he ascribed to capital letters appears as Table 2. (Table 2 involves some renumbering of this subset of features for computational convenience, but his definition of features and their mapping to the alphabet is preserved.)

Geyer's Set

Geyer (1970) reported a theory of iconic storage (terminology from Neisser, 1967) in the form of a computer simulation model. As with Laughery's work, the simulation methodology ensured an explicit statement of the information processes assumed to underlie recognition by feature analysis.

Geyer acknowledged that his feature set was based on Gibson's set with certain deliberate modifications. The rationale for the changes, implied by his theory of metacontrast masking, emphasized the importance of attention to a defined segment (area of the visual stimulation), segmentation being accepted as a preattentive process in perception. The geometric flavor of this construct led to the following modifications to Gibson's feature set: (1) a feature could be present one or more times in the list representing a letter (e.g., the letter H has two vertical lines, L only one); (2) a curved line was a set of one or more curved *segments* (e.g., four for the letter O, only one for the letter J); (3) the concepts of cyclical redundancy and discontinuity, if relevant to recognition at all, were matters of analysis, not basic features; (4) internal protrusions (e.g., the letters W and M) were area-specific features analogous to "open." He also discriminated between symmetry about vertical and horizontal axes. These modifications result in the feature set shown in Table 3.

The prime purpose of his simulation work was to explicate recognition performance in metacontrast masking paradigms, hence that work did not produce full alphabet confusion matrices generated from unmasked displays. However, the processes specified in that effort are entirely compatible with those explored in this investigation. This study employs processes that are a modification of his model appropriate to recognition without masking.

INFORMATION PROCESSES IN RECOGNITION

This report attempts to test proposed sets of feature lists for capital letters by generating the confusion matrix consequent to each set. The crucial question for this approach is, "How reasonable is it that the research has employed valid assumptions concerning the processes the human uses to recognize letters?" Fortunately, our scope is considerably narrowed by the initial decision to explore *feature list* processing, but even so there are alternative processing assumptions. This study attempted to be complete; each of the feature sets to be tested was used to generate confusion matrices in each of the alternative processes which appear, collectively, to exhaust the set of variants. We are dealing, therefore, with two categories of information processing assumptions: (1) those which, it is argued, are basic to the feature list concept and (2) those which constitute plausible variants of the basic model.

Table 3
Geyer Feature Set and Lists

	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
EXTERNAL																										
1 HORIZONTAL					2	1						1								1						2
2 VERTICAL		1		1	1	1		2	1	1	1	1	2	2		1		1		1	2					
3 SLANT (/)	1																					1	1		1	
4 SLANT (\)	1																					1	1		1	
5 CONVEX SEGMENT		2	3	2			3			1					4	1	4	1	2		1					
OPEN																										
6 HORIZONTAL			1				1												2							
7 VERTICAL										1											1					
8 WEDGED, HORIZONTAL		1									1							1						2	1	2
9 WEDGED, VERTICAL											2		1	2				1				1	1	2	1	
10 INTERNAL PROTRUSION													1										1			
11 INTERSECTION, INTERNAL	2	1			1	1		2			1							1								
12 BAR-HORIZONTAL	1				1	1	1	1																		
13 BAR-SLANT, CROSSING																	1									
14 NOT USED																										
15 SYMMETRY, VERTICAL	1								1				1		1					1	1	1	1	1		
16 SYMMETRY, HORIZONTAL		1	1	1	1			1	1		1				1									1		

The basic feature list information processing assumptions, used in all model variants are as follows: (1) display of arrays of symbolic stimuli result in the existence of stimulus-evoked (SE) lists; (2) preliminary (preattentive) processes organize coherent segments (areas) of the visual field into separate SE lists, which are the content of the icon and which are available to be attended to in a central processing channel; (3) in the central processing channel, SE lists are compared to a set of lists stored in long-term memory (LTM) by parallel processes of the general type set forth in the pandemonium model (Selfridge, 1959); (4) correspondence of an SE list to an LTM list results in response selection of the name corresponding to the matched LTM list.

If an initially clear, unambiguous presentation of a letter is attended to immediately, then there will be an identical match of the SE list to one LTM list. Correct recognition follows, without possibility of confusion. This is the normal situation for most real-world presentations of letters; only in medical or psychological laboratories is it usual to find controlled stimulation at degraded levels of visual clarity or reduced energetic intensity. However, precisely these conditions generate the confusion data by which we can test inferences concerning feature lists and feature processing. Hence, the most interesting assumptions of the model concern processes whereby degraded representations produce naming responses. Plausible alternatives for these processes define the variants of the basic model employed in this research.

All such alternatives share the four assumptions of information processing noted above, and they share the additional assumption that the mechanism of stimulus degradation is that the SE list may not contain all the information (features) that would have attended a full-strength stimulation, i.e., features can be "lost" but, on the other hand, cannot appear spuriously in a degraded SE list. This assumption originated in earlier simulation (Geyer, 1970) in keeping with the spirit of a pandemonium-type model. More elaborate justification can be offered from inspection of Townsend's (1971a, b) data, but these arguments are best presented in terms of model performance. These assumptions define a model whose essence is shown in Fig. 1. The assumed underlying psychological processes are discussed next.

Formulation of the SE List

Following stimulation by a letter whose correct list

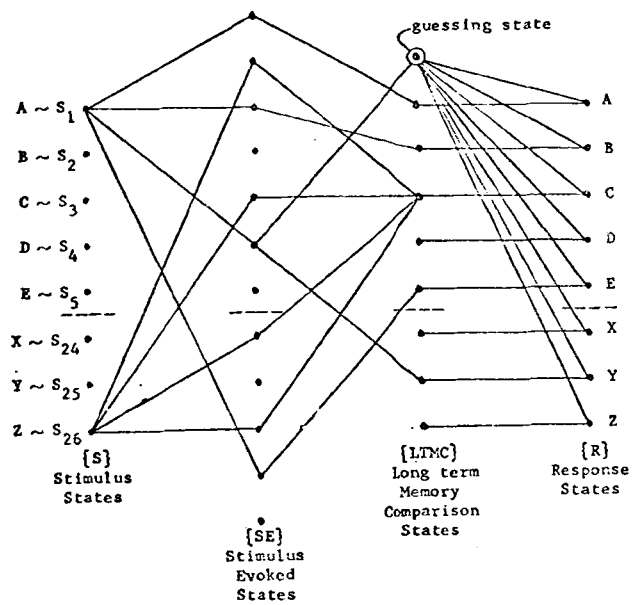

Fig. 1. General representation of recognition as a three-step process.

contains n features, there are 2^n possible SE lists. One of these is the correct list, one is the null list, all others are the possible combinations with one or more features being absent, but at least one present. For the letter O in the Gibson list, Table 1, there are four possible SE lists: (1) both features, 5 and 10, present; (2) Feature 5 present, 10 missing; (3) Feature 5 missing, 10 present; and (4) both missing, list empty. For a letter assumed to have six features (e.g., A in the Gibson list, Table 1), there are $2^6 = 64$ possible SE lists. While combinatorially complex, the assumed process of SE list formation is markovian. Each stimulus can result in 2^n next states. It remains to clarify the question of transitional probabilities, which introduces the first model variant. In one case, it was assumed that the probability of a feature carrying over from presence in a stimulus to presence in an SE list was the same for all features of the assumed set. In the other variant, it was assumed that some features might possess higher probability of being carried over than others, that individual values of p_i varied across the set $i = 1, N$, where N is the total number of features specified as the feature set (N = 15 for Geyer lists, 12 for Gibson lists, 19 for Laughery lists).

The first alternative is a very simplistic limiting case. It appeared deserving of analysis, but there is no evidence to support its psychological reality. Implicitly, features have been assumed to have some underlying neurological basis, there is no reason to assume that each neuronal network would build up to the same probability of availability ("strength") in a stimulation of marginal intensity, nor that all would persist at the same probability of availability in a decaying icon.

In either variant, there is a probability (P) vector of N elements, where N is the number of features employed in the assumed feature set. In the equal probability variant, designated EQ, each element, p_i, of this vector has the same value. In the variable probability variant, designated VP, the values of p_i may differ. The probability of not carrying over, q_i, is $1 - p_i$. Accordingly, the transitional probability for each SE list (state) is the product of all p_i for features carrying over times the product of all q_i for features missing.

Processing a Degraded SE List

The examples of processing in this section are based, arbitrarily, on the Gibson feature set (Table 1). Suppose that a controlled presentation of the letter E resulted not in the full list of features exactly denoting an E, i.e., 1, 2, 8, 9, 10, and 12, but a degraded SE list containing only 1, 2, 8, and 12 (9 and 10 absent). What response will be evoked, or what is the probability distribution of some set of responses? The degraded SE list shares four features with the six stored as the LTM list for E; it shares the same four with the five stored as the LTM list for F; and it contains all three of those stored as an L, plus an "extra" feature not present in the LTM list representing L. There are other examples of partial matches, but these three will suffice to illustrate information processes assumed in this study. All of the variants of models studied compute a goodness-of-match ratio by comparison of the SE list to the LTM list for each letter. For all variants, these comparisons are presumed to occur in a process basically similar to the "pandemonium" model (Selfridge, 1969; see also Neisser, 1967, pp. 74-76, for a summary description). In such a model, the best match will evoke response selection to the appropriate LTM list. In this study, the goodness-of-match comparison of an SE list to each LTM list is based upon a "hit ratio" (h) computed as follows: (1) for each feature held in common by the two lists, the numerator of the ratio is incremented by one; (2) for each feature in the SE list, but not in the LTM list, the numerator is decremented by one; (3) the denominator of the hit ratio is the number of features in the LTM list.

Assumptions 1 and 2 mean that the numerator of the hit ratio is a sort of communal content score, albeit one which reflects the theoretical assumption that features can be lost, but that intrusion errors are unlikely (i.e., are penalized in computing h). Assumption 3 is consistent with this theoretical assumption in making the number of features in the LTM list the standard against which the communal content score is compared, rather than the number of features in the possibly degraded SE list.

For an SE list that is an exact replica of an LTM list, h = 1.0. For the preceding example of a degraded SE list derived from the stimulus letter E, the following hs result: (1) for the comparison to E, $h = (4 - 0)/6 = .67$; (2) for the comparison to F, $h = (4 - 0)/5 = .80$; (3) for the comparison to L, $h = (3 - 1)/3 = .67$. Response selection processes differ slightly in different variants of the model, because some variants include the concept of a threshold on h, below which response involves guessing, but in all variants, if the response is not by guessing, it is via the letter(s) eliciting maximum h. In this example, the degraded representation of an E would be named an F.

The preceding discussion has been based upon processes which do not involve any sequential dependencies between comparisons, i.e., parallel processing. It may be objected that this is an unwarranted exclusion of a whole class of models—those which employ "sorting networks" or "decision trees" to sequentially test for the presence/absence of features in some predetermined order. Neisser (1967), argued that a sequential approach is not appropriate to pattern recognition. He cited several reasons, including evidence from his own research, concerning time for searching through lists for a single or multiple targets. Recent evidence from Sperling, Budiansky, Spivak, and Johnson (1972) would appear to reinforce this argument. In direct recognition of these arguments, this research was based on the pandemonium concept. However, it should

empirical confusion matrix (one of the three matrices reported by Townsend, 1971a, b). This iterative process begins with an initial solution for P having all p_i set equal at an arbitrary value. The arbitrary value was set, based on experience, to produce approximately 50% correct recognition on the average across the 26 stimulus letters, because this was the empirical condition employed by Townsend. However, the approximate correctness of this setting is of no scientific consequence, i.e., it does not influence the final outcome values but merely reduces the number of iterations, thereby economizing computer time.

For the EQ variants of the model, the iterative process proceeds, in response to a particular empirical matrix, as follows: (1) at each iteration, the markov-chain-generated confusion matrix is determined for a P vector; (2) the resultant stress is then computed; (3) the change in stress which would result from an increase of all elements of the P vector by a factor, G, or by a decrease by the factor 1/G is computed; (4) whichever change most reduces stress is adopted and a new iteration begins; (5) if neither alteration of the P vector reduces stress, then G is reduced further and the preceding computational steps, 3 through 5, repeated.

G is initially set at a substantial value, 1.16, or a 16% change. It is successively reduced one-half way toward 1.0, e.g., from 1.16 to 1.08, whenever neither change to the P vector reduces stress by as much as 1.0×10^{-6}. When G becomes less than 1.001, i.e., p would change by less than 0.1%, the process terminates and the resultant confusion matrix, value for p_i, and minimum stress is reported.

For the VP variants of the model, the process is similar, but at each iteration, the effect of changing each p_i individually by the factors G and 1/G is computed; the change that most reduces stress is incorporated, and the solution proceeds as for the EQ variant.

The transitional probabilities of the second stage of the model, i.e., response selection based upon some particular SE list, do not change as a result of changes in the transitional probability vector for the first stage. The second step probabilities are completely determined by the assumed feature set and response selection rules, i.e., the values of the guessing threshold parameters, T_2 and T_1, and whether or not the formulation includes response bias.

The solution program was written in FORTRAN IV for the SUNY at Buffalo CDC 6400.

Procedure

As noted, six model variants were studied: (1) equiprobable list formation with no guessing response (EQ/NG); (2) equiprobable list formation with guessing response and no response bias (EQ/WG-NRB); (3) variable probability of features in list formation with no guessing response (VP/NG); (4) variable probabilities of features in list formation with guessing response and no response bias (VP/WG-NRB); (5) equiprobable list formation with guessing and provision for response bias (EQ/WG-RB); and (6) variable probability of features in list formation with guessing response and response bias effects (VP/WG-RB). All of these variants were compared to each of the three confusion matrices reported by Townsend (1971a, b) for average performance of six Ss, S M.J. and S V.F. Thereby, effects due to individual S differences could be compared. In all model variants, each of the three feature sets proposed by Geyer (1970), Gibson (1969), and Laughery (1971) were studied. Thereby, each of the six model variants produced a simulated confusion matrix and resultant unexplained residual variance for nine conditions, three feature lists by comparison to three data matrices.

Table 4
Free Parameters of Model Variants/Feature List Combinations

Model	Geyer	Gibson	Laughery
EQ/NG	1	1	1
EQ/WG	3	3	3
VP/NG	15	12	19
VP/WG	17	14	21

When the model variant that produced least stress values across list and S comparisons was determined, that model was then studied with respect to variations in the initially assumed values for the guessing range parameters, T_2 and T_1. This study employed a sampling of feature list by S comparisons to search for a lower stress by varying T_2 and T_1. The redefined guessing parameter range was then used to rerun that model variant to obtain an improved solution.

Results

The criterion of model performance was stress, defined as the fraction of total variance of the empirical confusion matrix left as the unexplained residual variance from the prediction of the optimal synthesized model matrix. For the three empirical matrices, the total variance data are 6.22 for the average of six Ss data, 7.44 for the M.J. data, and 8.70 for the V.F. data. All stress values reported hereafter are the fractions of these values representing the unexplained residual variance.

Table 5 presents the resultant stress for each model variant, feature list set, and empirical data matrix comparison. Lines 1-18 of Table 5 present results for each of the model variants with the guessing threshold held constant (where employed) at $T_1 = 0.3$, $T_2 = 0.6$. From Lines 1-18, the following may be noted: (1) The variable probability (VP) configurations are always superior to the equal probability (EQ). (2) The guessing threshold variant (WG) is always superior to the purely deterministic (NG). (3) The variant that incorporates response bias (RB) is uniformly superior to the variant without response bias (NRB). (4) Thereby, the variant with variable probability, guessing threshold, and response bias (VP/WG-RB) is superior (Lines 15-18). (5) For the VP/WG-RB variant, the Geyer feature lists show least stress.

Determination of (Approximately) Optimal Guessing Threshold Range

The guessing threshold range parameters, T_2 and T_1, were originally set on the basis of indications from

Table 5
Residual Stress for Model Variants

Model Variant	Empirical Comparison Data			
	Six Ss	S M.J.	S V.F.	Average
1 EQ/NG ~ Geyer Lists	.374	.318	.279	.319
2 EQ/NG ~ Gibson Lists	.316	.309	.300	.308
3 EQ/NG ~ Laughery Lists	.322	.308	.233	.283
4 EQ/WG-NRB ~ Geyer Lists	.184	.157	.169	.169
5 EQ/WG-NRB ~ Gibson Lists	.235	.233	.225	.231
6 EQ/WG-NRB ~ Laughery Lists	.220	.232	.190	.213
7 VP/NG ~ Geyer Lists	.267	.197	.172	.207
8 VP/NG ~ Gibson Lists	.292	.255	.215	.250
9 VP/NG ~ Laughery Lists	.236	.239	.148	.203
10 VP/WG-NRB ~ Geyer Lists	.132	.118	.107	.117
11 VP/WG-NRB ~ Gibson Lists	.218	.193	.167	.190
12 VP/WG-NRB ~ Laughery Lists	.156	.156	.107	.137
13 EQ/WG-RB ~ Geyer Lists	.145	.144	.153	.147
14 EQ/WG-RB ~ Gibson Lists	.223	.217	.197	.211
15 EQ/WG-RB ~ Laughery Lists	.191	.208	.160	.184
16 VP/WG-RB ~ Geyer Lists	.112	.105	.102	.106
17 VP/WG-RB ~ Gibson Lists	.204	.187	.164	.183
18 VP/WG-RB ~ Laughery Lists	.141	.147	.104	.129
19 VP/WG-RB (Optimal Range) Geyer Lists	.087	.089	.083	.086
20 VP/WG-RB (Optimal Range) Gibson Lists	.145	.144	.128	.138

simulation of a different experimental paradigm (Geyer, 1970). However, that study did not include a sensitivity test of these parametric values. Having determined an otherwise optimal model configuration, the threshold range was varied systematically as part of this study. A threshold range of 0.5 to 0.7 was found to be the best choice for the Geyer and Gibson lists. The original range of 0.3 to 0.6 seemed best for the Laughery lists.

Performance of the VP/WG-RB Model with Approximately Optimal Guessing Threshold Parameters

Performance of this (final) model variant is shown on Lines 18, 19, and 20 of Table 5. The performance is best with the Geyer feature lists, comparisons to all three empirical data matrices yielding stress values of less than 9%. Neither the Laughery nor the Gibson lists yield stress values less than 10%.

It is worth noting that individual differences between Ss in the empirical data are substantial. If one of the three empirical matrices is used as a predictor of the other two, stress values range from .093 (M.J. as predictor of V.F.) to .147 (V.F. as predictor of average of six Ss). The VP/WG-RB model variant and Geyer feature list set achieves lower stress values for comparisons to each S matrix. This would appear to be an indication that the model and feature lists were synthesizing some human cognitive process(es) indigenous to the perception of capital letters. From this viewpoint, the degree of similarity of the three optimal P vectors corresponding to the three empirical confusion matrices would be indicative of the generality of the information process assumed. Very different P vectors would suggest that the different Ss employed different predominant subsets of features in recognition, whereas similar P vectors would indicate a generality of feature dominance across Ss. The Pearson product moment correlation for the three possible pairings of P vectors gave correlation coefficients of .872 ($p < .001$) for average of six Ss to M.J., .734 ($p < .002$) for average of six Ss to V.F., and .885 ($p < .001$) for M.J. to V.F. These results indicate that a predominant portion of process variance is common in all three simulations, but that variations due to individual differences are also substantial.

DISCUSSION

The results reported indicate that a very explicitly formulated set of information processes and a particular feature list set are capable of synthesizing confusion matrices which show an interesting degree of correspondence to three different empirical matrices. It is the explicit formulation of the information processes which gives these results their interest, however. Prior work with less structured models has produced lower levels of unexplained variance than those reported here. Townsend (1971a, b) reported on several models, the most successful of which produced stress values of .033 (average S), .024 (M.J.), and .025 (V.F.). However, this model employed 350 free parameters in optimizing stress, using minimal assumptions concerning the cognitive processes underlying recognition. It is difficult to compare these results directly with the VP/WG-NRB model with 17 free parameters.

Townsend's all-or-none (AON) model is to a

considerable degree analogous in performance to what one would obtain with our WG-RB model variants if it were possible to set the guessing state parameters such that for all SE states that produced h*s having exactly one arc (to the correct response) $h^* > T_2 = T_1$ and that for all other SE states $h^* < T_2 = T_1$ (i.e., only the unambiguously correct LTM comparisons evoked nonguessing responses). His AON model required 51 free parameters, and its performance is not very much better than the VP/WG-RM model performance (see Table 6). These results suggest that the structure of the latter model, having only one-third the parametric flexibility, cannot be completely remote from the optimal assignment of transition probabilities from stimulus through SE to LTMC states.

It is possible to compare the performance of these models to a prediction matrix based upon absolutely no psychological theory, only knowledge of the experimentally imposed constraint that stimulus energy was set such that S would average 50% correct responses. Using that fact, one can generate a uniform matrix composed of 0.5 in all main diagonal elements, and of an equal spreading of the remaining 0.5 response probability across the other 25 cells of each response vector. This uniform matrix contains no information processing assumptions. In any meaningful psychological context, it is not a model. Nonetheless, it can be used as a predictor matrix, and as such, it will generate residual sum squares and stress values in comparison to an empirical matrix. In comparison to Townsend's (1971a, b) data, the stress values generated are .123 (six Ss), .144 (S M.J.), and .137 (S V.F.).

At first glance, these values may seem surprisingly small, at least in comparison to the AON and feature processing models. However, that would underestimate the stress reducing power, in the uniform matrix, of knowing the true value of the average correct probability. Thereby, the residual contribution to stress along the main diagonal reflects solely failure to predict the letter-to-letter variation in correct recognition; the residual contribution to stress of off-diagonal cells reflects failure to predict cell-to-cell variation in confusability of different stimulus-response (S/R) pairs. Thereby, a model which predicts with less stress along the main diagonal is making some contribution to predicting relative recognizability, one which predicts better for the off-diagonal cells is making some contribution to relative confusability of different S/R pairs. Table 6 presents these comparisons to the uniform matrix for the VP/WG-RB model (with Geyer lists) and for the AON model.

Viwed in this light, the comparison of the uniform matrix to the AON model seems eminently sensible. The AON model, with its powerful parametric emphasis on correct recognition probabilities, has far superior performance to the uniform matrix along the main diagonal. On the other hand, the AON model is scarcely better than the uniform matrix at predicting confusion cell data. The small improvement is undoubtedly due to the AON model's capability for considering response bias, but the smallness of the difference suggests that the "all-or-none" assumption about recognition is highly suspect—i.e., that some S/R pairs are indeed more confusable than others.

Table 6
Comparative Stress Values

Predictor	Stress: Main Diagonal	Stress: Off Diagonal	Stress: Total
Uniform Matrix	.055	.082	.137
VP/WG-RB (Geyer Lists)	.018	.068	.086
AON, Townsend (1971a, b)	.001	.081	.082

On theother hand, the feature processing model is less superior than the AON model to the uniform matrix along the main diagonal, but shows a fairly substantial advantage in predicting confusion errors. These results also seem reasonable, since the feature retrieval probability parameters of this model would have become optimized for best overall prediction of both recognition and confusion performance, thereby attending to the relative confusability of S/R pairs.

Finally, many of the less adequate feature processing model variants and/or assumed feature lists produced higher stress than the uniform matrix. Analysis of the main-diagonal and off-diagonal components of stress would indicate the following: (1) the EQ variants of the model were notably inferior with respect to both recognizability and confusability predictions, but more so for confusability; (2) in general, all of the VP variants were superior to the uniform matrix with respect to recognizability predictions; but (3) the assumptions of a guessing threshold and response bias were needed to make confusability predictions superior to the uniform matrix; and (4) even in the VP/WG-RB variant, only the Geyer lists showed confusability predictions superior to the uniform matrix.

In summary, the results and these comparisons offer moderate encouragement; however, it would seem that Townsend's results with his more heavily parametered models provide an indication that there is substantial room for improvement in our present model and/or feature lists. The key question is whether the problem lies with the information processes (model) or with the assumed feature lists. The discussion now takes up the question of which assumptions it would appear that the study supported, and which seem to be priority candidates for ongoing refinement. This discussion considers first feature list assumptions, then information processing assumptions.

Feature List Sets

For model variants using (1) individually variable feature transition probability and (2) a guessing threshold, the Geyer feature lists produced lowest stress

values in comparison to each of the three empirical matrices. It should be remembered that the Geyer lists differ from the Gibson lists, from which they were derived, in only two facets; (1) the repetition of a feature is explicitly allowed, and (2) a small number of features were changed to better accommodate a predominantly geometrical organization of the SE list. The former "difference" is one which Gibson herself never addressed directly, therefore it may represent no difference at all. Information tests with the Gibson lists modified to incorporate multiple feature entries, but without the other modifications, did show decreased stress, although insufficient to close the gap in stress values reported for the two lists. Further, the Laughery lists, which were generally second in terms of minimal stress for the more successful model variants, also approximated multiple feature entries, albeit by explicit differentiation and enumeration. Pending further work, it seems reasonable to understand these findings as indicating: (1) Features that are phenomenologically repeated in the stimulus are repeated in the stimulus-evoked feature list used for recognition processing. (2) The high correlation of feature probability vectors across simulation of different Ss means that when we obtain a more refined definition of the list of relevant features, it will enjoy considerable generality between individuals. (3) The modified Gibson feature list presented in Table 3 as the Geyer list may be a promising beginning toward the explicit understanding of the features underlying capital letters.

The Information Processing Model

This study strongly supports assumptions of (1) variable feature probability, (2) a guessing threshold, and (3) the importance of response bias. The model also used assumptions which were not tested at all in this study. In particular, certain assumptions concerning the goodness-of-fit criterion for the pandemonium-type transition process from SE state to LTMC state and the assumption of independence of feature retrieval probabilities in the transition from stimulus to SE state require further study. These are discussed in more detail in the following sections.

Stimulus Evoked List Degradation

This study assumed that degradation of the SE list could only result from loss of features, that no spurious intrusions occur. The assumption does seem consistent with some of the detailed confusion error pairings found by Townsend, e.g., E is seen as F more often than F is seen as E. However, dividing the hit criterion by the size of the LTM list tends toward that result also, so it may be that noise intrusions are not inimical to simulation performance. Perhaps noise intrusions should be tried in future work.

Computation of Goodness-of-Fit

This study employed a criterion of maximum hit ratio, $h^*(e)$ (defined previously) as the sole basis for response selection. The question arises, should other hit ratios that are nearly as large as $h^*(e)$ receive some weighted portion of the transition probability from SE list to response selection? Other questions which could be systematically explored for improved performance of the model are: (1) Should the hit ratio be decremented for features in the SE list and not in the LTM comparison list (as was the case herein)? (2) Should the denominator of the hit ratio reflect the size of the LTM comparison list (as was the case herein)?

Independence of Feature Retrieval Probabilities

All of the reported work assumed that the value for p_i, the probability of the ith feature carrying over from the stimulation to the SE list, was independent of the context of other features also present in the stimulation. This assumption is more justifiable on the basis of an initial simplification for efficiency of study than on psychological grounds. On the latter basis, interactive facilitation and/or inhibition seems entirely reasonable. The implications of such a generalization do not appear to be severe for the formal model, but the solution process is expected to be seriously complicated. Nonetheless, this is a promising avenue toward increased explanatory power. It would seem quite apparent that this exploration must include not only solution process development, but that the empirical evidence base must be generalized to more than one level of energeticness of stimulation (fraction correct trials averaged across all letters). Such extension of Townsend's (1971a, b) work will also permit potentially interesting comparisons of the optimal P vectors across different stimulation levels.

APPENDIX A

Consider the basic form of the sorting net; each branch point is based on a yes-no decision, i.e., in the case of recognition by features, a present-absent decision. Inevitably, there are just as many negative (absent) branches as there are positive (present). At the bottom of a basic sorting net, there is always one branch that is the completely empty route, i.e., the complete absence of features. The apparent efficiency of a decision tree is dependent upon using these negative branches as constructive information for decision making. However, if we assume that an SE list may fail to contain some feature from a particular stimulation, due to the low energy of that stimulation and/or due to iconic decay prior to formation of the SE list, then a decision based solely upon absence of that feature will suppress the possibility of the SE list's being responded to as a degraded representation of the original stimulus. It follows that such use of negative decision branches suppresses symmetrical confusion errors in favor of a determined selection of one of the presumably

confusable responses. Elaboration of the network to eliminate this limitation inherently involves suppressing the decision potential of the negative branches by adding logically redundant tests.

A hypothetical example may make this process clearer. Suppose that a very simple symbol set is composed from the three features f_1, f_2, and f_3 and that the members of this set are three symbols: S_1 is composed of List 1 = (f_1, f_2), S_2 is composed of List 2 = (f_1, f_3), S_3 is composed of List 3 = (f_2, f_3). The most efficient possible sequential sorting net, shown in Fig. 2A, would always sort $L(f_3)$ to S_3, whereas presumably it would be more realistic to sort $L(f_3)$ to equal probability of S_3 and S_2. Similarly, $L(f_2)$ would always be sorted to S_3, whereas S_3 and S_1 should be equally probable, and $L(f_1)$ would always sort to S_2, whereas S_1 and S_2 should be equally probable. Finally, the empty list, L(blank), would always sort to S_3, whereas S_1, S_2, and S_3 should be equiprobable. To correct these deficiencies, the "redundant" tests shown in Fig. 2B are needed.

The point of this discussion is that the elaborated network analogous to Fig. 2B for recognition of capital letters would produce transitional predictions for the SE to LTMC states which cannot be distinguished from the parallel process model presented in this paper. Thereby, it is true that for the data used in this study, and for the purpose of investigating the merits of particular feature sets, no loss in generality is implied by assuming a parallel process.

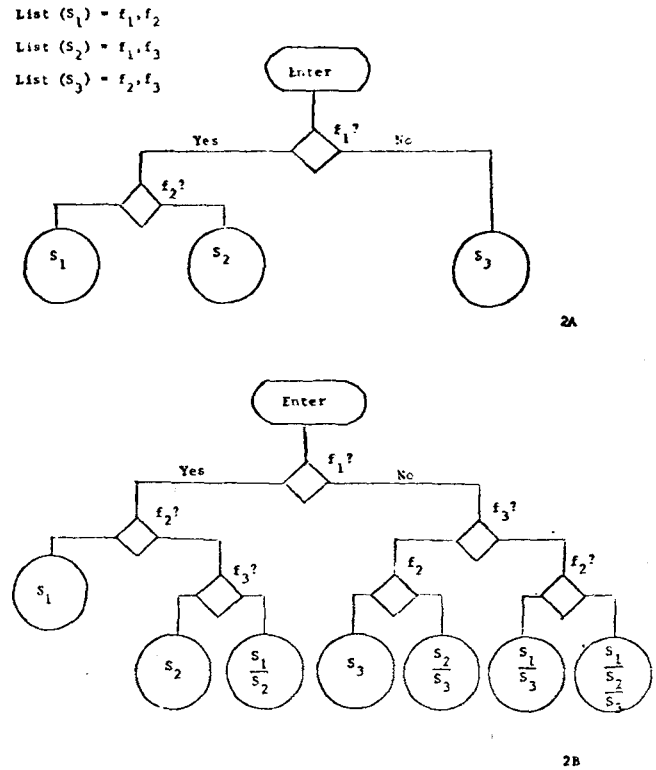

Fig. 2. Comparative representation of sequential recognition processes as decision trees without (2A) and with (2B) capability to accommodate symmetrical confusion errors.

APPENDIX B

Details of the model and problem formulation are as follows: Let (1) Φ denote the ordered set of features: f_1, f_2, $\cdots$, f_N, where N denotes the number in the set; (2) p_i denote the feature retention probability that if feature f_i exists in a stimulus list, then it will be present in the resultant stimulus-evoked (SE) list; (3) S_l denote the list of features characterizing the lth letter in the alphabet; (4) $J(S_l)$ denote a list (or sublist of features present in the stimulus list characterization of the lth letter and present in a resulting SE list (state); (5) $K(S_l)$ denote the list (or sublist) of features present in the stimulus list characterizing the lth letter and not present in the resulting SE corresponding to $J(S_l)$; (6) $e[J(S_l)]$ denote the SE state corresponding to the set of features $J(S_l)$; (7) $t_{l,J(S_l)}$ denote the one-step transition probability from the lth letter to the SE state $e[J(S_l)]$; (8) $t'_{J(S_l),L}$ denote the one-step transition probability from SE state $e[J(S_l)]$ to the Lth LTMC state; (9) $T_{l,L}$ denote the two-step transition probability from the lth letter (state) through the SE lists (states) to the Lth LTMC (state); (10) $r_{l,L}$ denote the empirical response probability in the confusion matrix from Stimulus Letter l to Response Letter L; (11) h(e,L) denote the hit-ratio of the SE state and the (LTMC) letter L; and (12) $\bar{r}$ denote the mean of the empirical response coefficients. Then the transition probabilities are defined as follows:

$$t_{l,J(S_l)} = \prod_{J(S_l)} p_j \prod_{K(S_l)} (1 - p_k) \tag{1}$$

for each stimulus letter and SE state; and

$$T_{l,L} = \sum_{\substack{\text{all} \\ J(S_l)}} [t_{l,J(S_l)} \cdot t'_{J(S_l),L}] \tag{2}$$

$$= \sum_{\substack{\text{all} \\ J(S_l)}} \left[t'_{J(S_l),L} \cdot \prod_{J(S_l)} p_j \prod_{K(S_l)} (1 - p_k) \right] \tag{3}$$

for each (l,L).

For the no-guessing variants of the model, the L = 0 LTMC state is empty, and $T_{e,L}$ also represents the three-step transitional probability from Stimulus Letter l to Response Letter L, $T^b_{e,L}$.

With guessing allowed, in addition to the 26 states, L = 1, 2, $\cdots$, 26, corresponding to the response letters A through Z, respectively, the state L = 0 is included in the model to permit guessing responses. "Guessing" is effected by the insertion of two thresholds, $1 \geqslant T_2 \geqslant T_1 \geqslant 0$ into the transition probabilities.

For each SE state e, compute the hit ratio h(e,L) and let $h^*(e) = \max[h(e,L)]$ over all response states L; then for each SE state, the transition probabilities are defined as follows:

For each SE state $e[J(S_e)]$ with $h^*(e) > T_2$,

$$t'_{J(S_e),L} = \begin{cases} \dfrac{1}{M} & \text{for each of the M response letter states L with } h(e,L) = h^*(e) \quad (4a) \\ 0 & \text{all other L} \quad (4b) \end{cases}$$

For each SE state $e[(J(S_1)]$ with $T_1 \leqslant h^*(e) \leqslant T_2$,

$$t'_{J(S_e),L} = \begin{cases} \dfrac{h^*(e) - T_1}{M(T_2 - T_1)} & \text{for each of the M response letter states L with } h(e,L) = h^*(e) \quad (5a) \\ \dfrac{T_2 - h^*(e)}{T_2 - T_1} & \text{for } L = 0 \quad (5b) \\ 0 & \text{all other L} \quad (5c) \end{cases}$$

For each SE state $e[J(S_1)]$ with $h^*(e) < T_1$,

$$t'_{J(S_1),L} = \begin{cases} 1 & \text{for } L = 0 \\ 0 & \text{all other L} \end{cases} \quad (6a)$$

The range of guessing is controlled by varying the interval (T_1, T_2). "No guessing" in the model is achieved by setting $T_1 = T_2 = 0$ in the above formulas.

If no response bias is permitted, then transitions from the $L = 0$ LTMC state to each of the $L = 1$ to 26 response states all have probabilities of 1/26. On the other hand, if response bias is present, the set of three-step transition probabilities with response bias $T^b_{l,L}$ is computed from the following formulae

$$T^b_{l,L} = T_{l,L} + b_L \left(1.0 - \sum_{L=1}^{26} T_{l,L}\right), \quad \text{all } (l,L) = (1,1), \cdots, (26,26) \quad (7)$$

where the response bias associated with letter L (b_L) is defined as

$$b_L = \epsilon_L \Big/ \sum_{L=1}^{26} \epsilon_L, \quad L = 1, 2, 3, \cdots, 26 \quad (8)$$

where

$$\epsilon_L = \begin{cases} \sum_{l=1}^{26} (r_{l,L} - T_{l,L}) \text{ if } \sum_{l=1}^{26} (r_{l,L} - T_{l,L}) > 0 \\ 0 \quad \text{otherwise} \end{cases} \quad (9)$$

The ensuing optimization problem is to determine the vector P of feature probabilities $(p_1, p_2, \cdots, p_N)$ which minimizes the objective *stress* function,

$$S(P_1, P_2, \cdots, P_N) = \frac{\sum_{\substack{\text{all} \\ l,L}} (T^b_{l,L} - r_{l,L})^2}{\sum_{\substack{\text{all} \\ l,L}} (r_{l,L} - \bar{r})^2} \quad (10)$$

REFERENCES

Frijda, N. H. Simulation of human long-term memory. Psychological Bulletin, 1972, 77, 1-31.

Geyer, L. H. A two channel theory of short term visual storage. (Doctoral dissertation, SUNY at Buffalo) Buffalo, N.Y: University Microfilms, 1970. No. 71-7165.

Gibson, E. J. *Principles of perceptual learning and development.* New York: Meredith, 1969.

Gibson, E. J., Gibson, J. J., Pick, A. D., & Osser, H. A developmental study of the discrimination of letter-like forms. Journal of Comparative & Physiological Psychology, 1962, 55, 897-908.

Gibson, E. J., Osser, H., Schiff, W., & Smith, J. An analysis of critical features of letters, tested by a confusion matrix. In: A basic research program on reading. Cooperative Research Project No. 639, U.S. Office of Education, 1963.

Hubel, D. H., & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. Journal of Physiology, 1962, 160, 106-154.

Laughery, K. R. Computer simulation of short-term memory: A component decay model. In G. T. Bower and J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory.* Vol. VI. New York: Academic Press, 1971.

Maturana, H. R., Lettvin, J. Y., McCulloch, W. S., & Pitts, W. H. Anatomy and physiology of vision in the frog (*Rana pipiens*). Journal of General Physiology, 1960, 43, 129-175.

Neisser, U. *Cognitive psychology*. New York: Meredith, 1967.

Piaget, J., & Inhelder, B. *The child's conception of space.* New York: Humanities Press, 1956.

Pritchard, R. M., Heron, W., & Hebb, D. O. Visual perception approached by the method of stabilized images. Canadian Journal of Psychology, 1960, 14, 67-77.

Selfridge, O. G. Pandemonium: A paradigm for learning. In *The mechanization of thought processes.* London: H.M. Stationary Office, 1959.

Sperling, G. The information available in brief visual presentations. Psychological Monograph, 1960, 74, No. 11.

Sperling, G., Budiansky, J., Spivak, J. F., & Johnson, M. C. Extremely rapid visual search: The maximum rate of scanning letters for the presence of a numeral. Science, 1971, 174, 307-311.

Townsend, J. T. Alphabetic confusion: A test of models for individuals. Perception & Psychophysics, 1971a, 9, 449-454.

Townsend, J. T. Theoretical analysis of an alphabetic confusion matrix. Perception & Psychophysics, 1971b, 9, 40-50.

Yonas, A., & Gibson, E. J. A developmental study of feature-processing strategies in letter discrimination. Paper presented at the Eastern Psychological Association, Boston, Massachusetts, April 1967.

(Received for publication December 11, 1972;
final revision received June 11, 1973.)