

On the nature of perceptual information during letter perception

STEPHEN J. LUPKER

University of Western Ontario, London, Ontario N6A 5C2, Canada

Letter perception has been traditionally viewed as a process in which individual features are accumulated over time. In order to test this notion, a special stimulus set was created having little or no featural redundancy. Using a masking paradigm, confusion matrices were generated at each of eight interstimulus intervals. Few, if any, of the predictions made by the feature accumulation models were upheld. Instead, it is suggested that letter perception is better thought of as a global-to-local process. When a letter is presented, an observer initially perceives a large array of perceptual data. Over time, a clearer view of the stimulus emerges as the perceptual system brings the letter into focus. Thus, global information about the letter is available quite early in processing, while the letter's more local aspects become available only after relatively extensive perceptual processing.

Everyone knows that sentences are made up of words, words are made up of letters, and letters are made up of lines or features. However, the functional utility of each of the subunits in perceiving the larger units continues to be a hotly debated topic. The aim of this investigation is to shed some light on the role and the identities of the "features" involved in the letter-perception process. Two general models of this process will be described and then evaluated in their ability to predict the nature and frequency of errors in a masking experiment. A good starting point would be to define clearly what is meant by the letter-perception process.

The basic framework borrows heavily from the general information processing model proposed by Massaro (1975). When a letter is presented to an observer, its visual representation is established in an iconic or preperceptual visual storage (PVS). The perceptual process can then begin to resolve figure-ground information from this representation in order to ascertain the important "featural" characteristics of the letter. This process takes time and, depending on the display parameters, all the perceptual information may not be resolved before PVS has either fully decayed or has been interfered with by a masking stimulus. The acquired perceptual information is

then processed by a decision/naming stage where the appropriate abstract character name is generated. These names are then stored in short-term memory (STM) where they can be rehearsed, recoded and/or output as responses. If no name exists, the perceptual information itself may be held in STM until a response based on this information must be made.

Feature-Accumulation Models

The question being addressed concerns the nature of the buildup of perceptual information during the perceptual process. Specifically, when only partial perceptual information about a letter is available, what is the nature of that information? The conventional way of viewing this process, at least since the pattern recognition work of the late 1950s and, specifically, the pandemonium model of Selfridge (1959), has been as a feature accumulation process. Initially, the observer has no information about the nature of the visual image. Over time, the individual features of the letter become available from the representation in PVS. The features will continue to be accumulated until a feature list sufficient to identify the letter has been acquired. The decision system can then produce the appropriate letter name.

The nature of the feature list describing a given letter depends, of course, on the identities of the features presumed to be functional in the process. The simplest version of the general feature accumulation model was proposed by Rumelhart and Siple (1974). In their theory, the features correspond to the line segments of the letters. Thus, the feature set necessary to describe a given letter will be a list of horizontal, vertical, diagonal, and presumably, curved lines. More sophisticated versions of the general model (e.g., Gibson, 1969; Lindsay & Norman, 1972) assume that relational information is also ex-

This paper is based on parts of a doctoral dissertation submitted to the Department of Psychology, University of Wisconsin-Madison. I am indebted to Lola Lopes, Dom Massaro, and Gregg Oden for many useful comments and suggestions both prior to and during the writing of the original manuscript. Special thanks are owed John Theios for his guidance, suggestions, and review of earlier drafts. I would also like to thank Bill Krane and Albert Katz for their contributions to the final manuscript. Requests for reprints should be addressed to Stephen J. Lupker, Department of Psychology, University of Western Ontario, London, Ontario N6A 5C2, Canada.

tracted from the iconic representation along with the more basic featural information. Thus, properties like angles and symmetry may also become part of the feature list. Presumably, these relational features must be, in some sense, secondary because a feature like an angle could not logically exist unless two basic line features were also present. Thus, relational features could not be accumulated without the simultaneous accumulation of two line features. On the other hand, line features can exist and be identified without the observer acquiring any information about the nature or existence of the angle between them. Thus, on the average, relational features should be the last features to be added to the feature list.

Initial support for the accumulation models arose from the analysis of empirically obtained confusion matrices of the 26 uppercase letters. In such a matrix, certain confusions will arise frequently, generally involving letters having physical similarities. The assumption made in all situations was that letters confused frequently must share features. Therefore, by noting the featural overlap existing in frequently confused letter pairs, the features which are functional in the letter perception process could be ascertained.

Unfortunately this type of analysis is lacking in a number of ways. For example, any two letters which are confused frequently will share a number of physical attributes. Which of these are presumed to be functional in the perception process probably reflects more on the investigators' preconceived notions about what constitutes a feature than on the perceptual information the observer actually acquires.

However, even beyond this, it would seem that a confusion matrix based on all 26 letters is a very poor tool for evaluating the nature of partial perceptual information. If perception is an accumulation process, incorrect responses should be a result of the acquisition of an incomplete feature list. However, in a task where the potential stimuli are the 26 letters, it is not the case that an incomplete feature list will necessarily result in an incorrect response. This is due to the fact that each of the letters of the alphabet has a certain amount of intracharacter redundancy in a task of this sort. That is, given partial featural information from the presented letter, the observer can sometimes use the fact that the stimulus must be one of the 26 letters in order to fill in missing features and correctly identify the letter. Some letters have very little redundancy and all features must be accumulated before accurate identification is possible. The letter T might be an example. Others, for example Y, can be uniquely distinguished from the other 25 letters on the basis of minimal featural information. Thus, the response to the presentation of a particular letter may tell us little about the actual perceptual information used in making that response.

In order to make a statement about the nature of partial perceptual information, the redundancy problem must first be dealt with. The present investigation attempts to do that through the creation of a stimulus set having little or no intracharacter redundancy. The stimulus set was composed of the four uppercase letters in the Roman alphabet having only two line features (L, T, V and X) as well as the four line features comprising them (|, -, /, \). Thus, if perception is an accumulation process, partial featural information about any of the letters will not allow an accurate identification. Instead, the observer will have to choose a response from among those stimuli consistent with the perceptual information acquired.

Additionally, four two-feature nonletter characters (|, |, ^, ^) were included in the stimulus set. These were created by a different juxtaposition of the two features in each of the letters used. Thus, the controls on redundancy are still maintained. Partial information from any of these characters would only allow the observer to narrow the response candidates to a set consistent with the acquired information.

The purpose of including these stimuli was to evaluate the role of relational featural information in the feature-accumulation process. If perception is a feature-accumulation process, these stimuli can be distinguished from the letters only on the basis of the relational information which should be acquired later in perceptual processing. Thus, the particular confusions the two-feature stimuli generate should give some indication of how relational information is used in completing the perceptual process. Additionally, a comparison between the letters and the characters with similar features should shed some light on the role of familiarity in extracting relational information.

There was one other important difference between the present study and more traditional studies. Since the question being addressed concerns the development of the percept over time, a masking paradigm was used in order to generate confusion matrices at eight different points in perceptual processing. The particular types of confusions frequent at each of these points should provide a clearer picture of exactly how perceptual information does become available over time.

At this point, it is important to consider whether the stimulus set does exercise the appropriate controls on redundancy. Potentially, problems could arise if the line features either singly or when contained in letters or characters carried unwanted length information, orientation information, or spatial location information. This extra information, if present, could allow the existence of nonperceived features to be inferred, thus destroying the controls on redundancy. To guard against the first of these

possibilities, the lengths of all the lines of the same orientation were made exactly equal. However, the diagonal lines contained in the different characters did differ slightly in orientation. Additionally, it was, of course, necessary to vary the relative spatial locations of the line features in order to create the different letters and characters. Nonetheless, because of the small size of the stimulus field (less than $.30^\circ$ visual angle in both height and width) these variations in orientation and spatial location were essentially imperceptible. Additionally, the mask was made over four times larger than the stimulus field in order to prevent the subjects from using its relation to any perceived feature as a spatial location cue. Thus, it is unlikely that undesired length, orientation, or spatial location information was available to the subject. However, because the influence of these factors can not be ruled out, their implications for the feature accumulation model will be discussed later.

If it is the case that this extra information is not available to the subject, the predictions of the feature-accumulation model are quite straightforward. The basic notion inherent in this model is that featural information builds up over time. When processing is terminated early, no features will have been acquired and the errors should generally be random guesses regardless of the stimulus presented. A little later in processing, the basic line features should begin to become available. This information could be sufficient to allow the identification of single features but not of two-feature stimuli. Thus, the masking functions for the single features should rise most rapidly. It would not be essential that these four masking functions be equivalent. The perceptual system may be more finely tuned to perceiving, say, horizontal lines than vertical lines.

Complete information about the two-feature stimuli (i.e., both line features and relational information) should not be available until later in processing. Thus, masking functions for these stimuli would be expected to rise more slowly than those for the single-feature stimuli. In particular, the masking function for a letter or nonletter character should rise no more rapidly than the masking functions of its component features because both features must be perceived before the two-feature stimulus can be identified. In fact, it would be expected that the identifiability of any two-feature stimulus would, in some sense, be predictable from the identifiability of its component features. The type of errors observed when two-feature stimuli are presented would be expected to be a function of the interval between the stimulus and the mask. As mentioned above, when processing is terminated early, errors to these stimuli should be random guesses. Slightly later, when line features begin to become available, errors may start to involve

one of the line features contained in that stimulus. Finally, a little later in processing, both line features may be available but relational information may yet be missing. Thus, at this point, errors should involve other two-feature stimuli having the same two line features.

The Global-to-Local Model

In recent years, an alternative way of conceptualizing the buildup of perceptual information has been suggested (Bouma, 1971; Eriksen & Schultz, 1978). In this scheme, perception is not viewed as a feature-accumulation process but as a focusing process. An observer's initial perception of the letter is much like what one would see if the presentation had been completely out of focus. Early in processing, the letter's more global aspects, such as its general shape, or what Bouma termed its "envelope," would be available. (A letter's envelope is defined as the smallest polygon without indentation which fully encloses the letter.) From this information, the observer can determine such rudimentary things as the height-to-width ratio, whether the letter has an ascender or a descender, etc. Depending on the nature of the set of potential stimuli, this may be sufficient information to allow the observer to identify the stimulus. Later in processing, the more local information, such as where the gaps in the envelope are and how the inner parts are arranged, will become available. Only at this point will the observer have the information which is more traditionally thought of as featural information.

Evidence supporting this view of perception is, if anything, less substantial than the evidence supporting the feature-accumulation model. Bouma's (1971) argument for this type of model is based on an empirically obtained confusion matrix of the 26 lowercase letters. As before, Bouma made the assumption that letters confused frequently must share features. Thus, by noting the physical similarities existing in frequently confused letter pairs, he could ascertain the perceptual information functional in the process. His general finding was that letters confused frequently were those having similar envelopes and not those having similar features in the more traditional sense. However, as noted earlier, this type of analysis is somewhat lacking and probably reflects more on the investigator's biases than on the perceptual information the observer actually obtains. Thus, it seems that a more complete exploration of this type of model is in order.

Unfortunately, deriving predictions from a global-to-local type of model are a bit complex. The model states that the perceptual data available at any point during perceptual processing, in some sense, resembles a blurred image of the stimulus. Over time, the image becomes more well defined, until, with

sufficient processing time, the local features become clear. At a particular ISI, correctly identifying a stimulus will be difficult to the extent that other stimuli have similar outlines. Thus, in order for this model to predict the masking functions and confusions generated by particular stimuli at particular ISIs, it is necessary to know the extent to which all 12 stimuli would be perceptually defined at each ISI. Fortunately, a reasonable approximation to these predictions can be produced by considering the presumed nature of the perceptual data at a single, random point in processing. Stimuli having similar outlines at this point should have relatively similar outlines at other points also. Therefore, any stimulus difficult to perceive at this point, because many other stimuli have similar outlines, should be difficult to perceive at other points, also. Additionally, the particular confusions predicted at this point would generally also be predicted at other points, both earlier and later in processing. So, while the predictions which are obtained in this manner may not be totally precise, they should be fairly close.

In order to determine what the perceptual data might look like at some random point in processing, a blurred image of each stimulus was created by defocusing a slide containing all 12 stimuli. The defocused images were then photographed and a tracing made of their fuzzy outlines. The tracings obtained are shown in Figure 1, surrounding the corresponding stimuli. The predictions were then made on the basis of the relationships between these 12 outlines.

Because perception is viewed as a focusing process, at each point in processing, the observer should have a wider array of perceptual data than is actually contained in the stimulus. From this, two general predictions can be derived. First, errors should seldom result from the perception of a stimulus that is smaller than the presented stimulus. Thus, the responses appropriate for the single-feature stimuli should be given erroneously on very few trials. This is in oppo-

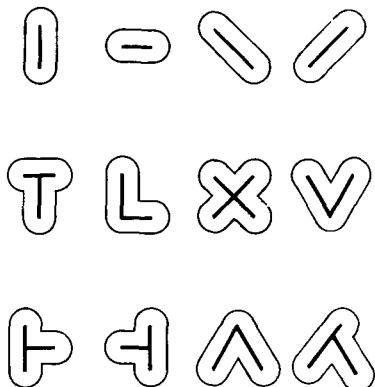


Figure 1. Complete stimulus set. Each stimulus is surrounded by the tracing of that stimulus's blurry outline.

sition to the feature-accumulation models which predict that single features can often be perceived when two-feature stimuli are presented, especially at relatively brief ISIs. Second, when perception is terminated early, a broad array of perceptual data will be perceived. When this occurs, subjects should often give a response appropriate for a very large stimulus. The largest outlines in this stimulus set belong to the X, the \wedge , and, to a lesser extent, the V and the Λ . Thus, responses appropriate for these stimuli should be given erroneously on a fairly large number of trials. This prediction also stands in opposition to that made by the feature-accumulation models. These models predict that the features sufficient to perceive large characters could not be perceived unless those characters were actually presented.

On a more specific level, the stimuli having the most distinct outlines should be the easier stimuli to perceive at any ISI. This implies that the thin tubular outlines of the four single features should make them the most perceptible stimuli. The vertical line is a possible exception because the T, the \vdash , the \dashv , and to some extent the L also have long tubular shapes extending in a vertical direction. Thus, the \vdash may often be confused with these stimuli which will cause its masking function to rise more slowly.

In fact, all stimuli containing the vertical line have long tubular shapes extending in a vertical direction. Thus, confusions among these stimuli should be frequent. However, the positioning of the horizontal line should cause some pairs to be confused less often than others. Specifically, the T which is "top heavy" and the L which is "bottom-right heavy" should not be confused very often. Nor should the two half-H characters be confused very often since they "lean" in opposite directions. However, both half-H characters are somewhat similar to the T and the \vdash is quite similar to the L, so these confusions should appear more frequently.

With respect to the masking functions for these stimuli, since the T is similar to at least three other stimuli (i.e., \vdash , \dashv , and \dashv), its masking function should rise more slowly than that of the L, which is similar only to the \dashv and, to some extent, the \vdash . Further, because the only difference between the two half-H characters is that the \dashv is similar to the L, its masking function should rise more slowly than that of the \dashv .

As was noted earlier, the two-feature stimuli created from the diagonal lines have envelopes which are similar in the sense that they are fairly wide. Thus, in general, these stimuli will tend to be confused with one another. However, as with the horizontal-vertical stimuli, some of these pairs should be confused less often than others. In particular, confusions should be less frequent between the more "top-heavy" V and the more "bottom-heavy" Λ and \wedge than between

MASKING FUNCTIONS	FREQUENT CONFUSIONS
I < -, /, \	[I, T] [I, L]
T < L	[I, H] [I, F]
X < V	[T, H] [T, F]
F < H	[L, F] [X, V]
	[X, ^] [X, v]
	[^, ^]

Figure 2. Summary of the specific predictions of the global-to-local model.

the latter two stimuli. However, the X, which has a rather square outline, would be somewhat confusable with all three of these stimuli.

With respect to the masking functions, since the X should be confused with the V, the ^, and the v, it should be more difficult to identify than the V at any ISI. On the other hand, there is no reason to expect that the masking functions for the ^ and the v will be any different. A summary of these predictions is provided in Figure 2.

METHOD

Subjects

Eleven University of Wisconsin undergraduate volunteers (three males and eight females) participated in this experiment. All received course credit for their participation.

Apparatus

The subjects were seated at a table in a semidarkened room in front of a Tektronix (Type RM 503) oscilloscope (P15 phosphor). The oscilloscope was positioned 1.54 m away from the subject. To maintain this distance, a chinrest was attached to the edge of the table. On the table in front of the subjects was a box with four response buttons, all of which were operable. A Digital Equipment Corporation PDP-8 computer was programmed to present the stimuli and to record the stimulus, response, and interstimulus interval (ISI).

Procedure

A simple masking paradigm was used. On each trial, the subject was shown 1 of 12 stimuli (subtending a visual angle of approximately .20° in width and .30° in height, depending on the stimulus presented), followed by a 250-msec masking stimulus at one of eight ISIs. The duration and brightness of the stimulus display, as well as the amount of illumination in the experimental room, were varied from session to session and subject to subject in order to maintain each subject's performance at around 50% overall. The average duration of the stimulus display was between 4 and 5 msec. The stimuli were the four single features, the four letters and the four two-feature characters shown in Figure 1. Each stimulus was presented equally often.

The mask was a grid of hexagons subtending a visual angle of .60° in both height and width. Thus, the mask was much larger than the stimuli. This was done to prevent the subject from using some sort of relative position information in identifying the presented stimulus. The ISI values used were 10, 20, 30, 40, 50,

75, 100, and 200 msec. Due to the nature of the stimulus display, it was felt that these ISIs would yield the most complete description of the masking functions for all 12 stimuli.

The subject was first read the instructions and allowed some time to study a sheet of paper containing the 12 stimuli. The subject was told that on each trial one of these stimuli would be displayed and would be subsequently followed by a masking stimulus. Each stimulus would be displayed equally often and the subject's job would simply be to identify the stimulus presented on each trial.

In order to begin the experiment, the subject first rested his chin on the chinrest and fixated on the dot in the middle of the screen. After fixating, the subject initiated a trial by pressing any of the four buttons on the box in front of him. After the stimulus and the mask had been presented, the subject was to decide which stimulus had appeared and then give the response in two parts. First, the subject indicated if the stimulus was a single feature, a letter, or a two-feature nonletter character. He was to press the first button if it was a single feature, the second button if it was a letter, and the third button if it was a two-feature character. This would cause a small plus sign to appear in the middle of the screen. The plus sign indicated to the subjects that they were now to indicate specifically which of the four possible stimuli of that type had been presented. They could do this by pressing one of the four buttons because the four stimuli within each category were arbitrarily assigned to the four buttons. (A sheet with the 12 stimuli and the stimulus-response mapping was next to the subject at all times.) After the parameters of the trial were recorded, the subject was informed whether his response was correct or incorrect. The fixation dot then reappeared and the next trial was ready to be initiated.

Each subject was run for 5 consecutive days with two ½-h sessions per day. Each session contained two replications of each cell of the 12 (stimuli) by 8 (ISI) design, and, therefore, consisted of 192 trials. A short rest was allowed between sessions.

RESULTS

The exact nature of confusion matrices is always strongly determined by the biases subjects have for guessing certain stimuli. Thus, in order to analyze the perceptual factors involved, the effects of these biases must somehow be removed. What appears to be the best way of doing this is to appeal to the choice model of Luce (1963; Townsend & Ashby, Note 1).

In Luce's model, it is assumed that there exist bias parameters, b_i , for each stimulus i and similarity parameters, S_{ij} , representing the figural similarity between stimuli i and j . It is generally assumed that $S_{ii} = 1$, $S_{ij} = S_{ji}$, and the sum of the b_i s equals one. The probability of responding with response j when stimulus i is presented, P_{ij} , is then as written in Equation 1,

$$P_{ij} = \frac{S_{ij} \cdot b_j}{\sum_{k=1}^n S_{ik} \cdot b_k}, \tag{1}$$

where n is the number of stimuli in the stimulus set. This model has $[n(n-1)]/2 - 1$ parameters (here 65 for each ISI condition) and generally provides a very good fit to the data.

In the present set of data, bias and similarity parameters were estimated for each of the eight confusion matrices according to the formulas given in the Appendix of Townsend (1971). The root mean squared deviation between the observed and predicted P_{ij} s increased essentially monotonically from .0127 at an ISI of 10 msec to .0259 at an ISI of 200 msec. Chi-square analyses were not performed, because in every condition a number of expected cell frequencies were below five. However, these deviation values are quite in line with those obtained for the choice model by Townsend (1971), and, thus, the model appears to do a fairly good job of mimicking the data.

The similarity parameters were then used to estimate what the masking functions for the 12 stimuli would have been like had the subjects been unbiased. That is, Equation 1 was reduced to

$$P_{ij} = \frac{S_{ij}}{\sum_{k=1}^n S_{ik}}$$

The masking functions for the three stimulus types with adjusted probability correct as a function of ISI are shown in Figure 3. As is apparent, single features were most perceptible, followed by letters and characters. This result is, generally, in accord with the predictions of both types of models.

The individual masking functions for the various stimuli are shown in Figure 4 for the horizontal and vertical features, letters and characters and in Figure 5 for the diagonal features, letters and characters. The reader should recall that feature accumulation models predict that a letter or character should be no more perceptible than its individual features. Yet, as Figure 4 shows, the L was one of the most perceptible stimuli in the stimulus set, while one of its components, the vertical line, was one of the least perceptible stimuli. In fact, at some ISIs, the vertical line was less perceptible than every two-feature stimulus having it as a component. Clearly, there is no evidence that

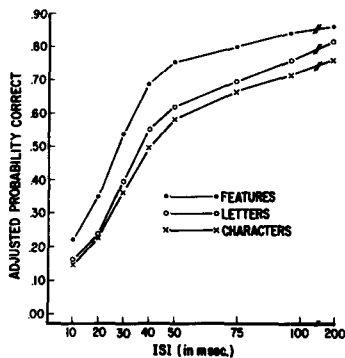


Figure 3. Idealized masking functions for the three different stimulus types.

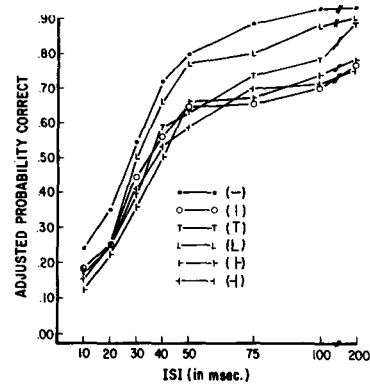


Figure 4. Idealized masking functions for the horizontal and vertical features, letters and characters.

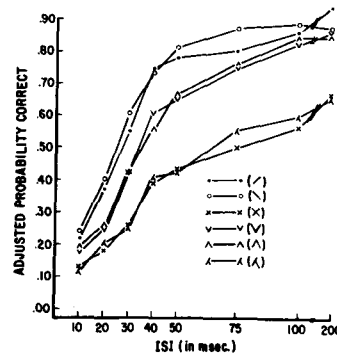


Figure 5. Idealized masking functions for the diagonal features, letters and characters.

the perceptibility of the two-feature stimuli in Figure 4 is a function of the perceptibility of their component parts. Much the same conclusion can be drawn from Figure 5. The two diagonal features are quite easily perceived, yet the two least perceptible stimuli in the entire stimulus set (X and ^) were created from these highly perceptible features. It appears that this aspect of the data speaks against the idea of perception as a feature-accumulation process.

On the other hand, the individual masking functions are predicted fairly well by the global-to-local model. It was expected that the | would be hard to perceive, because of its similarity to a number of other stimuli, while the other single features would be quite perceptible. Among the letters, it was expected that the L would be more perceptible than the T and the V more perceptible than the X. Among the characters, it was expected that the † would be more perceptible than the ⊥. Except for this final prediction, all of these results were obtained. What was not expected, however, was that the ^ would be so much more perceptible than the / or that the X would be so difficult to perceive. Nonetheless, the overall pattern of results seems to be fairly well in line with the predictions of the global-to-local model.

Table 1
Idealized Overall Stimulus Confusion Matrix

Stimuli	Responses											
		-	/	\	T	L	X	V	†	‡	Λ	∧
	919	18	24	20	159	41	78	43	111	184	49	119
-	23	1178	20	23	59	41	99	41	88	71	46	73
/	30	21	1157	36	33	28	145	68	51	45	43	105
\	27	24	37	1192	24	30	125	54	60	26	53	105
T	160	44	25	16	967	43	108	61	80	107	47	101
L	46	33	22	24	47	1088	93	58	179	34	34	103
X	54	52	79	63	75	59	686	133	102	94	115	248
V	42	30	53	39	65	53	192	986	85	60	50	109
†	95	54	30	34	68	138	122	72	870	77	60	135
‡	177	50	33	18	103	29	128	58	89	906	35	132
Λ	49	35	34	39	48	32	166	51	75	40	994	199
∧	86	39	58	56	72	66	254	76	112	100	138	701

The similarity parameters were next used to estimate what the eight confusion matrices would have been if the subjects had been unbiased. A visual examination of these eight matrices revealed that the pattern of errors changed very little over ISI conditions. Thus, only the overall confusion matrix is reported in Table 1. The first thing to notice is that, regardless of the stimulus presented, error responses are not random guesses. This was true even at the briefest ISI. Instead, errors almost always involve stimuli having more perceptual data than the presented stimulus. Thus, even after only 10 msec, the observer is not left with an absence of perceptual data. Additionally, the pattern of this information does not appear to change qualitatively over time. Thus, there is little support for the feature accumulation model in this portion of the data.

Along similar lines, single features occurred as errors to the presentation of two-feature stimuli only rarely. Again, this is true at all ISIs. The single exception to this is the set of two-feature stimuli containing the vertical line (in particular, T, †, and ‡). These stimuli were perceived as the vertical line with some regularity. Thus, if perception is a feature accumulation process, it appears that only when a vertical line is contained in the display do the observers ever acquire one line feature without the other. On the other hand, a large number of confusions between the vertical line and the stimuli containing it were predicted by the global-to-local model because all of these stimuli have long, tubular shapes extending in a vertical direction. At the same time, the global-to-local model did not predict that any of the other single features would appear as error responses with any frequency. Thus, this result seems to support the notion of perception as a global-to-local process.

Probably the predominant feature of this confusion matrix is the frequency of error responses appropriate for the stimuli X and ∧, even when the presented stimulus had only horizontal and vertical features. While the frequency of these error responses is a

bit larger than anticipated (31% of the errors made), this type of result is in line with the predictions of the global-to-local model. That is, since perception is viewed as a focusing process, whenever this process is terminated prematurely the perceived information should be more compatible with the larger stimuli. Thus, X and ∧ should be more likely to appear as error responses than other, smaller, stimuli. On the other hand, the feature-accumulation models would not predict that these two responses would be popular, because the features sufficient to perceive large characters should not be perceived unless those characters were actually presented.

In order to analyze the other confusions predicted by the global-to-local model (see Figure 2), two subsets of the overall matrix were extracted from Table 1 and are listed in Table 2. Among the stimuli in the left half of Table 2, 7 of the 10 possible confusions (14 cells) were anticipated to be frequent. As the reader can see, the only exceptions to this were the pairs (|, L) and, to a lesser extent, (T, †). Of the three pairs anticipated to be infrequent, only the pair (†, ‡) had any noticeable frequency. In the right half of Table 2, four of the six possible confusions (eight cells) were anticipated to be frequent. If an arbitrary criterion of 110 is chosen as the cutoff point between frequent and infrequent confusions, this yields a perfect separation of the eight cells predicted to be frequent from the four cells predicted to be infrequent. So, again, the global-to-local model seems to do a fairly good job of describing the data.

DISCUSSION

The feature-accumulation model being examined here was a general one based on a few seemingly reasonable assumptions. Features were assumed to be accumulated over time, with line features becoming available first, followed by the relational information necessary to put the line features together properly.

Table 2
Idealized Error Matrices for Selected Subsets of Stimuli

Stimuli	Responses					Stimuli	Responses			
		T	L	⊥	⊥		X	V	^	^
		159	41	111	184	X		133	115	248
T	160		43	80	107	V	192		50	109
L	46	47		179	34	^	166	51		199
⊥	95	68	138		77	^	254	76	138	
⊥	177	103	29	89						

Further, the assumption was made that the line features were processed independently, with factors like lateral inhibition, capacity limitations, or intracharacter redundancy having little, if any, effect. Thus, the existence of any one feature in the display was assumed to neither inhibit nor facilitate the processing of any other feature in the display. Finally, it was assumed that the subjects would base their responses solely on the features they perceived rather than inferring the existence of features not yet perceived.

There are three aspects of the present results which argue against the feature-accumulation model of perception based on these assumptions. First, the perceptibility of the component parts of a two-feature stimulus in no sense predicted the perceptibility of that stimulus. Specifically, all stimuli containing a | were perceived at least as well as, if not better than, the ⊥, and the two stimuli most difficult to perceive (X and ^) were composed of two features which were quite easily perceived. Second, the errors made when single features were presented were not random but almost always involved stimuli having more figural information than was available in the display. Finally, there was no evidence that the pattern of errors made when two-feature stimuli were presented changed as anticipated over ISIs. That is, error responses to these stimuli were never random, even at the briefest ISI, nor was there much evidence that they involved component features at any ISI. Instead, at all points in processing, errors to two-feature stimuli involved other two-feature stimuli which did not necessarily have the same two component features. Taken together, these results present a strong argument against the general feature-accumulation model presented above.

Owing to the current popularity of feature-accumulation models, a closer look at some of the present assumptions seems warranted. The first assumptions of no capacity limitations and no lateral inhibition of one feature by another feature both imply that the processing of one feature does not inhibit the processing of any other feature. If either of these assumptions were incorrect, it would not follow that the perceptibility of a two-feature stimulus should be a function of the perceptibility of its component parts. Instead, the processing of the two-feature stimuli

would be slowed to some unspecified extent, as was the case with the diagonal stimuli, specifically X and ^ . However, this was not the case with the horizontal-vertical stimuli, specifically, the L, which was perceived much better than one of its component features. Additionally, if there had been a slowing due to lateral inhibition or capacity limitations, the observer's perceptual information should have been in an incomplete state (either no features or one feature) for a relatively long period of time. If so, errors to the presentation of two-feature stimuli should have reflected these partial information states. Clearly they did not. Thus, while relaxing these two assumptions allows some masking function results to be accounted for within the feature-accumulation framework, much of the data is yet to be explained.

The effects of redundancy and the efforts to control it were discussed earlier. If redundancy had not been successfully controlled, this would have allowed the subjects to use perceived information to infer the existence or nonexistence of features not yet perceived. This would have had the effect of speeding up the processing of two-feature stimuli in relation to their component features, as was the case with the L and, perhaps, to a much lesser degree, the other horizontal-vertical stimuli. For example, it may have been that the position of the horizontal line in the L was detectable and allowed the subjects to infer the existence and position of the vertical line and, hence, to identify the stimulus. However, even if redundancy had been a factor in the perception of the L, it did not seem to contribute to any of the other results in the present study. Thus, as before, this assumption does not seem to be the cause of the failure of the feature-accumulation model.

The final assumption is that the subjects reported only what they actually perceived rather than guessing that another feature was actually contained in the display. It may be argued that, without definitive knowledge that a feature is absent, subjects may report it as present if it is compatible with a perceived feature. This type of strategy would also speed up the processing of two-feature stimuli relative to their component features and, thus, could account for the rapidly rising masking function of the L. It would also partially explain the error re-

sults. That is, if nonperceived features were being inferred to be in the display, error responses would involve stimuli having more features than the presented stimulus. However, the use of this strategy would still not explain why, when horizontal-vertical stimuli were presented, stimuli having totally incompatible features (X and \wedge) would appear as error responses so often. Nor would it explain the relationships among the masking functions for the diagonal stimuli. So, again, even though this assumption might be in error, much of the data still cannot be explained within the feature accumulation framework.

It is quite conceivable that while none of these assumptions alone was responsible for the failure of the feature accumulation model, all of them together may have led to the present set of results. For example, if it is assumed that only diagonal features laterally inhibit one another and that redundancies and guessing strategies are only relevant to horizontal-vertical stimuli, most of the present results can be explained. However, the totally post hoc nature of these new assumptions makes them uninviting. Thus, it appears that the applicability of the feature-accumulation model to the present set of data must be regarded as somewhat limited.

On the other hand, though the predictions derived from the global-to-local model were less precise, this model seems to handle the data much better. All but one of the predicted relationships between masking functions were observed and errors generally involved stimuli having more perceptual information than the presented stimulus. Specifically, the two largest stimuli, X and \wedge , were given as responses quite often while single features appeared very seldom as erroneous responses. Additionally, the individual confusions which occurred frequently were generally the ones that were expected. Thus, it appears that the general pattern of results does support the idea of perception as a focusing process in which perceptual information is acquired in a global-to-local manner.

There were, however, two aspects of the data which, from the standpoint of the global-to-local model seemed to be a bit surprising. Both of these involved the stimuli X and \wedge , and both were mainly matters of degree. That is, the model predicts that these stimuli should appear as error responses quite often, yet it was not anticipated that over 30% of the total errors would involve the responses appropriate for these two stimuli. Additionally, while these two stimuli do appear to be highly confusable with a number of stimuli, including each other, thus, making them generally difficult to perceive, it was not expected that they would be so much more difficult to perceive than all of the other stimuli.

However, the inability of the model to predict the strength of these effects may reflect more the treat-

ment given it here than any inadequacy in the model. The reader should recall that the predictions obtained were generalizations derived from considering the potential state of the perceptual information at some random point in processing. As such, any a priori predictions of the *degree* to which these effects would manifest themselves were not really possible. So, in retrospect, these findings may not be so surprising after all. In fact, if these two effects had not occurred (i.e., if the two stimuli had been highly perceptible and had appeared seldom as error responses), those results would have been quite surprising and damaging to the global-to-local model.

The Global-to-Local Model and Spatial Frequency Models

Another test of the global-to-local model has recently been provided by Navon (1977). Using large letters composed of small letters as stimuli, Navon demonstrated that information about the larger letters is available much earlier than information about the smaller letters. This result, demonstrated in a number of paradigms, also allowed Navon to make a strong argument for the global-to-local model.

Navon's studies, however, are also important for another reason. In recent years, theories of visual perception based on spatial frequency channels have emerged (Graham, 1976; Sachs, Nachmias, & Robson, 1971). These theories are based on the idea of size-tuned receptive channels which respond not to stimulus "features," but, rather, to particular spatial frequencies. If it is assumed that the time necessary for a channel to respond is a function of its characteristic spatial frequency, with lower frequency channels responding earlier, these models could be regarded as special cases of the general global-to-local model. As such, it would seem that a model of this sort might account for Navon's data.

However, in each of Navon's paradigms, he included a control condition where only a single small letter was presented. Navon found that information about the single small letter was available just as rapidly as information about the larger letter composed of small letters. Therefore, the absolute position of the particular spatial frequency channel whose output is necessary in order to respond correctly did not determine the subject's level of performance. From this result, Navon argued that a model incorporating the idea of receptors sensitive to particular spatial frequencies would not be the most parsimonious way of describing his data.

A further case against spatial frequency models has been made in a recent paper by Coffin (1978). Reanalyzing a series of alphabetic confusion matrices published in the last 20 years, Coffin found that a general spatial frequency analysis could account for no more than 16% of the variance in any of the matrices. He concluded that, while spatial frequency

coding may exist, a model based solely on spatial frequency coding can not provide a full account of perceptual performance. Therefore, because of the results of Coffin (1978), Navon (1977), and others (e.g., Kinchla & Wolf, Note 2), which argue against the general spatial frequency model, it did not seem appropriate to apply this type of model to the present data.

The model of letter perception offered here is, then, an extension of Bouma's (1971) work. Perception is viewed as a process in which an initial array of perceptual data is focused, over time, in order to reveal actual figure and ground. This initial perception presumably involves all of the true figure as well as much of the ground around it. At subsequent points in processing more and more of the ground is lost as the percept comes to take on the form of the letter. Eventually, the perceptual process will have removed all of the extraneous data, leaving only the local features, the entities on which the feature models are based.

If perception is stopped, by means of a masking stimulus, at any point in processing, the observer will generate a "best guess" as to the letter's identity on the basis of the pattern of perceptual information available as well as any contextual constraints acting in the situation. This process may be somewhat analogous to the word recognition process as described by Morton (1969). The pattern of perceptual information would be compared against the expected perceptual information from a variety of potential stimuli. The results of these comparisons could then be combined with any a priori expectations to produce strength values for each of the stimuli. The stimulus with the largest strength value would, then, become the "best guess" response. Stimuli having similar general shapes will be confused often, making these stimuli appear to be difficult to perceive, as was the X in the present study. However, the perceptibility of stimuli will appear to change from situation to situation as the identities of the other potential stimuli change. For example, in an experiment involving an X and eight straight lines of differing orientations, the X may not be mistaken for any of the other stimuli very often, making it appear to be quite perceptible.

Frequent error responses to a particular stimulus will generally only involve other stimuli having at least as much figural information. Additionally, there will generally be some similarities between the general outlines of any two stimuli which are confused often. How the similarity between stimuli should be measured is problematic. In the present study, a very general set of symmetric similarities combined with biases provided an adequate account of the data. Future research should lead to a better specification of the nature of similarities in this perceptual realm, thus, providing a more precise set of predictions.

In summary, the present experiment was designed to provide a test of two general types of models of letter perception, the feature-accumulation model and the global-to-local model. Only a few of the predictions of the feature-accumulation model were upheld, while the global-to-local model seemed to provide a very adequate description of the data. It is suggested that letter perception might be better viewed as a focusing process wherein the observer's initial "view" of the letter is a blur of perceptual data. Over time, the excess figural information is gradually lost until, after relatively extensive processing, the letter's local aspects finally become available.

REFERENCE NOTES

1. Townsend, J. T., & Ashby, F. G. *Toward a theory of letter recognition: Testing contemporary feature processing models*. Unpublished manuscript, November 1977.
2. Kinchla, R. A., & Wolf, J. *The order of visual processing: "Top-down," "bottom-up," or "middle-out."* (Psychology Research Report No. 21), Department of Psychology, Princeton University, December 1977.

REFERENCES

- BOUMA, H. Visual recognition of isolated lower-case letters. *Vision Research*, 1971, 11, 459-474.
- COFFIN, S. Spatial frequency analysis of block letters does not predict experimental confusions. *Perception & Psychophysics*, 1978, 23, 69-74.
- ERIKSEN, C. W., & SCHULTZ, D. W. Temporal factors in visual information processing. In J. Requin (Ed.), *Attention and performance VII*. New York: Academic Press, 1978.
- GIBSON, E. J. *Principles of perceptual learning and development*. New York: Appleton, 1969.
- GRAHAM, N. Spatial-frequency channels in human vision: Detecting edges without edge detectors. In C. S. Harris (Ed.), *Visual coding and adaptability*. Hillsdale, N.J.: Erlbaum, 1976.
- LINDSAY, P. H., & NORMAN, D. A. *Human information processing: An introduction to psychology*. New York: Academic Press, 1972.
- LUCE, R. D. Detection and recognition. In R. D. Luce, R. B. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. I). New York: Wiley, 1963.
- MASSARO, D. W. *Experimental psychology and information processing*. Chicago: Rand-McNally, 1975.
- MORTON, J. Interaction of information in word recognition. *Psychological Review*, 1969, 76, 165-178.
- NAVON, D. Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, 1977, 9, 353-383.
- RUMELHART, D. E., & SIPLE, P. Process of recognizing tachistoscopically presented words. *Psychological Review*, 1974, 81, 99-118.
- SACHS, M. B., NACHMIAS, J., & ROBSON, J. G. Spatial frequency channels in human vision. *Journal of the Optical Society of America*, 1971, 61, 1176-1186.
- SELFRIDGE, O. *Pandemonium: A paradigm for learning*. In *Symposium on the mechanization of thought processes*. London: H. M. Stationery Office, 1959.
- TOWNSEND, J. T. Theoretical analysis of an alphabetical confusion matrix. *Perception & Psychophysics*, 1971, 9, 40-50.

(Received for publication August 2, 1978;
revision accepted January 22, 1979.)