JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR 19, 503–524 (1980)

Experimental Tests of a Hierarchical Model of Word Identification

JAMES C. JOHNSTON

Bell Laboratories

AND

JAMES L. MCCLELLAND

University of California at San Diego

According to hierarchical models of word perception, identification of a word is mediated by identification of its component letters. In this paper a hierarchical model is presented which explains why people are more accurate in perceiving a briefly presented letter when it appears in a word than when it appears alone (the Word–Letter Phenomenon). The model is consistent with what is presently known about the conditions under which the phenomenon is and is not obtained and makes two new predictions: (1) The sizeable Word–Letter Phenomenon that can be obtained using a mask made up of letter features should be greatly reduced if the mask consists of complete letters. (2) The size of the Word–Letter Phenomenon should be the same whether or not mask letters spell a word. Both predictions run counter to the widely accepted principle that interference increases with the similarity of target and mask. Nevertheless, the experiments reported in this paper confirm both predictions.

Ever since psychologists began to study reading, many researchers have argued that the identification of a word is mediated by an analysis of its component letters (see Huey, 1908; more recently, Estes, 1975; Geyer, 1970; Gough, 1972; Henderson, 1975b; Johnston, 1978; McClelland, 1976). Models of this kind (here called *hierarchical* models) are attractive partly because they would simplify the task of learning to read. If word identification were mediated by letter analysis, readers would have to learn to identify each word only once, even though type fonts and writing styles for letters differ greatly. Indeed, the fact that people can identify words learned in one font when presented in another font (McClelland 1976, 1977) suggests that hierarchical models do capture at least one way people can identify words. Nevertheless, since the first experimental studies of the reading process, many have argued that hierarchical models are wrong, and that words must be identified directly from their visual configurations (Cattell, 1886; Cosky, 1976; Erdman & Dodge, cited in Huey, 1908; Johnson, 1975; Smith, 1971; Woodworth, 1938).

A major obstacle to acceptance of hierarchical models of word identification has long been the existence of the Word-Superiority Effect. As Cattell (1886) first demonstrated, people viewing brief displays are more accurate in reporting letters in words than letters in equally long strings

The idea for the experiments reported here was conceived independently by each of the authors, who collaborated on its execution. This research was supported by Bell Laboratories, by NSF Grant BNS76-14830 to the second author and by Grant PHS MH15828 to the Center for Human Information Processing, UCSD. We thank C. Whitman, J. Mannas, and B. Reisman for their assistance in conducting this research, and C. S. Harris, M. Jackson, T. K. Landauer, D. Noreen, B. Reisman, D. Scarborough, and S. Sternberg for helpful comments on earlier versions of this paper. Requests for reprints should be sent to James C. Johnston, Bell Laboratories, 2D-441, Murray Hill, New Jersey 07974.

0022-5371/80/050503-22$02.00/0

of unrelated letters. This finding has posed a paradox for hierarchical models: If performance on unrelated-letter strings measures the accuracy of letter identification, and letter identification provides the input to word identification, how can the output of the word identification process contain more information about the stimulus than the input?

Two attractive resolutions of this paradox have been available for some time:

1) *Redundancy*. One can deny that the output of the word identification process contains more information than the input. Although people are more accurate in reporting letters in words than in unrelated-letter strings, letters in words might represent less stimulus information, since they are partially redundant (Miller, Brunner, & Postman, 1954), and hence more easily inferred from partial cues.

2) *Forgetting*. One can deny that reports of unrelated-letter strings provide a valid measure of the number of identified letters that serve as input to the word identification process. People may fail to retain some of the letters they have identified in unrelated-letter strings until they can report them (Baddeley, 1964).

Redundancy effects may well contribute to the Word-Superiority Effect under some conditions (e.g., Rumelhart & Siple, 1974; Smith, 1969). However, Reicher (1969) obtained a strong Word-Superiority Effect using a paradigm intended to eliminate redundancy effects. Reicher measured performance with a forced-choice test between two alternative letters, whose probability of occurrence cannot be predicted from the remaining letters in the string. For instance, choice alternatives could be "D" vs "K" for testing the fourth letter in the word "WORD" (where "WORK" is equally likely to have been the stimulus) and for testing the fourth letter in the unrelated-letter string "OWRD" (where "OWRK" is equally likely to have been the stimulus). Although the Reicher paradigm rules out guessing from context during the selection of an overt response, it does not rule out the use of redundancy during earlier perceptual processing of stimulus information (Bjork & Estes, 1973; Massaro, 1973; Rumelhart & Siple, 1974; Smith, 1971; Thompson & Massaro, 1973; Wheeler, 1970). Johnston (1978) has devised a more comprehensive experimental test for redundancy effects. Johnston argues that if redundancy were being used at *any* stage of processing it should improve performance more for letters that are highly predictable from context than for letters that are not. Johnston (1978) found no such trend under conditions (similar to Reicher's) that produced a strong Word-Superiority Effect. Thus it appears that redundancy does not provide a valid way to reconcile hierarchical models with the Word-Superiority Effect.

Forgetting has been shown to be substantial during reporting of letters from long unrelated-letter strings (Baddeley, 1964). Forgetting therefore doubtless plays a role in some demonstrations of the Word-Superiority Effect. Reicher (1969) however, obtained the Word-Superiority Effect after taking three precautions to minimize forgetting: (1) use of short, 4-letter stimuli; (2) use of a forced-choice test that probed perception of only one letter in each string; and (3) use of a single letter alone as a control condition, in addition to the usual unrelated-letter-string control condition. The first two measures, while helpful, would hardly convince a determined forgetting theorist that no residual problem remained. However, Reicher's finding that performance can be better for a letter in a word than for even a *single letter* alone (the "Word-Letter Phenomenon," Johnston & McClelland, 1973) presents greater difficulties. A forgetting theorist must argue that people have difficulty remembering a single identified letter for a second or so, when they have nothing else they have to remember. This has not appeared to be an attractive proposition to embrace.

Upon reflection, however, the problem

which the Word–Letter Phenomenon poses for a forgetting theorist can be seen to depend on the assumption that an identified letter is represented in short-term memory, or something akin to it. It is possible, however, that an identified letter (or word) is first represented in some more peripheral form of storage, from which even one item can be quickly "forgotten." Relying on this idea, we have recently proposed (Johnston, 1978; McClelland, 1976; McClelland & Johnston, 1977) that a hierarchical model of word identification *can* be formulated that is consistent with what is presently known about the Word–Letter Phenomenon. The purpose of this paper is to explore the viability of this model. We will first explain the model and show how it can account for the Word–Letter Phenomenon, and then derive some new predictions from the model and test them.

The Hierarchical Model

Our model postulates a hierarchical network of detectors for features, letters, and words (see Figure 1). We assume that a word target is first preprocessed so that each letter in it is allocated to a position-specific letter-processing channel. Within each of these channels, information is analyzed for the presence of different letter properties or features. We do not know the features of letters for which detectors actually exist; plausible candidates include line segments in appropriate orientations and positions (Rumelhart, 1970). For instance, a detector for a vertical bar at the left of a character might be excited by presentation of the letter "E."

The outputs of the feature detectors for a given letter-channel serve as inputs to letter detectors for that channel. An active feature detector will provide excitatory input to detectors for all letters that are consistent with the feature, and provide inhibitory input to detectors for all letters that are inconsistent with the feature. (In Figure 1 excitatory inputs are represented by solid

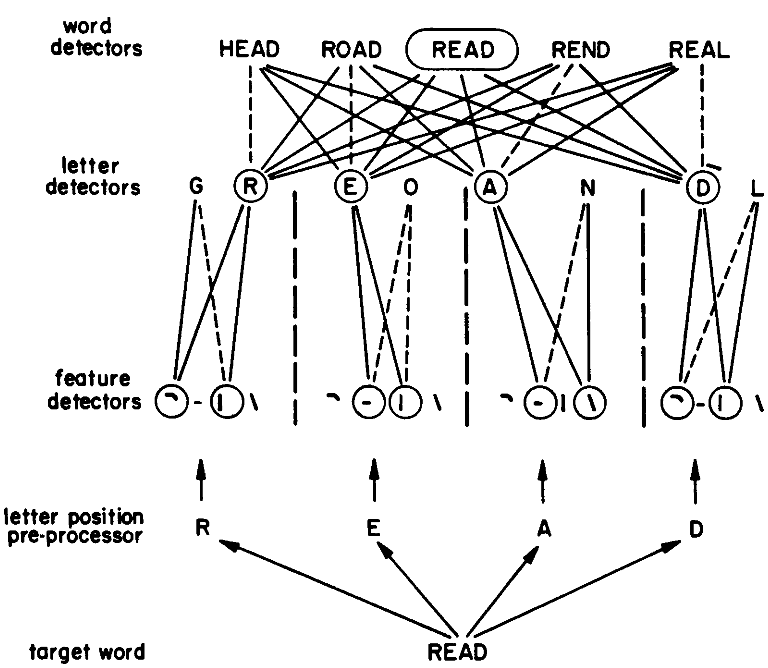

FIG. 1. Schematic diagram of a fragment of the detector network hypothesized to mediate word and letter identification. A possible set of active detectors after presentation of target word READ is circled. To guide overt responses, active detectors must be located by central attentional processes (not shown).

lines, and inhibitory inputs by dashed lines; active detectors are circled.)

The output of letter detectors serves in turn as input to word detectors. An active letter detector excites detectors for words consistent with that letter (in the appropriate position), and inhibits detectors for words inconsistent with that letter (in the appropriate position). Thus activation of the detector for "R" in the first position would excite the detector for "READ" and inhibit the detector for "HEAD." (Also inhibited would be detectors for words such as "DEAR," that had an "R" elsewhere in the word but had some other letter in the first position.)

Within our model, inhibitory input has a more extreme effect than the mere absence of excitatory input. In the absence of excitatory input, an activated detector is assumed to remain active for a considerable period of time before decaying back to resting level. Inhibitory input, however, is assumed to drive the activity in a detector down rapidly. Thus an activated detector will remain active longer when receiving no input than when receiving inhibitory input.

Finally, we assume that mere momentary activation of a detector is not sufficient for production of an appropriate overt response. Under most conditions further processes are necessary to locate an active detector and encode an appropriate representation in a holding buffer, whose contents are consulted during response selection. We will refer to these additional processes as "central attentional processes," but will not specify in detail how they operate. For present purposes it is only critical that these processes be relatively slow, so that there will be a substantial probability that a detector activated for only a brief interval (but sufficiently long to provide substantial excitatory and inhibitory input to the next level of detectors) will not be located before its activity ceases.

This model is similar in spirit to models described by Estes (1975) and LaBerge and Samuels (1974). It differs from them, however, in stressing the importance of inhibitory as well as excitatory connections between detector levels. Although such inhibitory connections have been hypothesized in physiological models of hierarchical feature detector systems (e.g., Hubel & Weisel, 1963), they have rarely been postulated in models of processes operating at higher levels of analysis. The distinction which we make between the rapid spread of activation within a detector network and slower encoding of information from the network is shared by many recent cognitive theories (e.g., Neisser, 1967; Posner & Snyder, 1975; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977).

The model we have presented includes only aspects of the word identification process relevant to our interpretation of the Word-Superiority Effect and the Word–Letter Phenomenon. Aspects of word identification not directly involved in these phenomena (e.g., role of syntactic and semantic context, Morton, 1969; semantic-relatedness activation effects, Jacobson, 1973; Meyer & Schvaneveldt, 1976) will not be considered. In addition, the role of several other complicating aspects of word processing has been minimized. These include: (1) the role of orhtographic structure or pronounceability in perceptual encoding (Baron & Thurston, 1973; Carr, Davidson, & Hawkins, 1978; McClelland & Johnston, 1977; Spoehr & Smith, 1975); (2) positional uncertainty at the feature and letter levels (Estes, 1975); (3) possible "level-skipping" connections from features to words (LaBerge, 1976; McClelland, 1977); and (4) possible within-level interactions between detectors (e.g., inhibition between detectors for the same features in nearby locations; Bjork & Murray, 1977; Estes, 1972). The last possibility will be considered explicitly in testing our experimental predictions. It turns out, however, that our results can be explained without regard to any of these potential complications.

Explaining the Word–Letter Phenomenon

The way in which our hierarchical model

explains the Word–Letter Phenomenon will be presented at two levels. In this section of the text a verbal explanation will be provided. To demonstrate that no sleight of hand is involved in this account, we present in the Appendix a more specific version of the model that can be demonstrated to perform appropriately. The specific model treats detectors as discrete two-state (on–off) devices, a convenient theoretical simplification. We argue in the Appendix that this simplification should provide a reasonable approximation under the visual conditions used to obtain the Word–Letter Phenomenon. We do not, however, wish to tie the validity of our general model to the validity of the special assumptions made in the Appendix version. We have therefore emphasized the more general version of the model in the text of this paper. Many readers may find that the Appendix provides an easier introduction to the model, because the simplifying assumptions allow a more concrete presentation, including a state diagram (Figure 6).

Now to proceed to how our hierarchical model explains the Word–Letter Phenomenon. We have already described in the previous section how presentation of a target letter or word activates appropriate units in the detector network: feature detectors corresponding to features present in the target, letter detector(s) corresponding to the target letter(s), and a word detector corresponding to the target word (if any). Our explanation of the Word–Letter Phenomenon depends critically on how these detectors are affected by presentation of a patterned mask following the target.

Presentation of the mask is hypothesized to propagate a new wave of activity up the detector network. At the feature level, presentation of the mask results in the rapid activation of detectors for features contained in the mask, and the rapid deactivation of previously active detectors inconsistent with mask features. Active detectors for mask features then send excitation to letter detectors consistent with those features and inhibition to letter detectors inconsistent with those features. Patterned masks are normally chosen to contain some features inconsistent with each possible letter (i.e., Reicher's mask of overstruck X's and O's), so any active letter detector should receive some inhibition. We assume that any inhibition has a powerful effect, so that even if excitatory input is also received, inhibitory input will usually be decisive (cf. Rumelhart & Siple, 1974) and the letter detector will rapidly be deactivated. Thus shortly after presentation of the mask, no letter detectors at all should remain active. Since word detectors receive input only from letter detectors, and inactive letter detectors send out no signals at all, any active word detector will receive no inhibitory input. In the absence of inhibition any word detector already active will remain active for a relatively long period of time (until passive decay occurs). We have now arrived at our explanation of the Word–Letter Phenomenon: performance is better for a letter in a word than for a letter alone because the word is represented by an additional higher-level detector that remains active longer, and is therefore more likely to be discovered by central attentional processes.

In our earlier discussion about how to reconcile hierarchical models of word identification with the Word-Superiority Effect and Word–Letter Phenomenon, we noted that reconciliations tended to rely on one of two principles: redundancy or forgetting. Our model can be considered to be an exotic forgetting model: we are suggesting how people can, after all, "forget" one identified letter before it can guide overt responses. However, we are proposing what is essentially a distributed processing system; letter identities are lost from the detector network rather than from short-term memory (which they have never reached). Thus we do not need to make the dubious assumption that forgetting *from short-term memory* is the problem for one letter—an assumption that would deviate dramatically from the usual capacity estimates for such a store (e.g., Miller, 1956).

In the model presented here, the patterned masking display clearly plays a key role in producing the Word–Letter Phenomenon: it activates feature detectors which in turn inhibit detectors for letters but not for words. For our purposes the important property of the masking display is that it contains some letter features that are inconsistent with any letters activated by the target. Of the reported experiments showing a clear Word–Letter Phenomenon (Hawkins, Reicher, Rogers, & Patterson, 1976; Holender, 1979; Johnston, 1978; Johnston & McClelland, 1973, 1974; McClelland & Johnston, 1977; Reicher, 1969; Taylor & Chabot, 1978; Thompson & Massaro, 1973; Wheeler, 1970) all but one (Wheeler, 1970) have used masks that meet our requirements (overstruck X's and O's, other overstruck letter combinations, haphazard patterns of line segments, or a collage of letter fragments). Wheeler (1970) has reported an advantage for a letter in a word over a single letter using a mask containing only random arrays of dots. Excluding one condition of this experiment, whose results Wheeler attributed to an apparent-movement artifact, only a small Word–Letter Phenomenon was obtained (5 vs 16% for Johnston & McClelland, 1973). Furthermore, even Wheeler's dot masks may have frequently produced effective contours that activated detectors for some letter features (letters made of small numbers of dots can, after all, be rapidly identified).

If a patterned posttarget mask is omitted, then there should be no new mask input to inhibit detectors for either letters or words. If target letters or words were displayed at the same duration and intensity as in patterned-masking experiments, we assume that detectors would remain active for a substantial period of time, and central attentional processes would have little or no trouble encoding information from them. (In fact, when patterned masks are omitted, either the duration or the intensity of the target must be drastically reduced before any substantial number of errors are made; cf. Johnston & McClelland, 1973.) Without a patterned mask, our model provides no reason why activity in word detectors should persist longer than activity in letter detectors. Thus, according to our model, the main cause of the Word–Letter Phenomenon should be absent when patterned masking is omitted.[1] To our knowledge only four studies have attempted to measure the Word–Letter Phenomenon, omitting the patterned mask (Johnston & McClelland, 1973; Juola, Leavitt, & Choe, 1974; Massaro & Klitzke, 1979; Taylor & Chabot, 1978). All four found a small advantage for single letters over letters in words, reversing the usual Word–Letter Phenomenon. The more traditional Word-Superiority Effect, using a letter in a UL string rather than a single letter as a control for a letter in a word, has been demonstrated without a patterned mask (e.g., Rumelhart & Siple, 1974). Such results can reasonably be attributed to differences in short-term memory retention for several letters vs a single word.

New Predictions from the Hierarchical Model

Let us adopt the following conventions for naming types of masks: A *Feature Mask* contains letter features, but no letters and no word. A *Letter Mask* contains letters (and, necessarily, letter features) but no word. A *Word Mask* contains a word (and, necessarily, letters and letter features).

[1] Although, without patterned masking, a word detector should not persist any longer than letter detectors, it might provide an additional opportunity for central attentional processes to find at least one active detector that indicates the correct response (see General Discussion). The absence of a Word–Letter Phenomenon without patterned masking might mean that an additional active detector provides little benefit under these conditions. (Without patterned masking, all active detectors might remain available for a relatively long time, leading to very few "misses" in locating active letter detectors.) Alternatively, this possible advantage for a word target might simply be outweighed by other disadvantages (e.g., lateral inhibition among letters, or the need for multiple letter-detectors to be activated on the same trial in order to activate a word detector).

According to our hierarchical model, the Word–Letter Phenomenon is obtained with a feature mask because the mask actively inhibits letter detectors but not word detectors. The reason is that detectors are actively inhibited only by input from the next lowest level. A feature mask activates detectors for new features inconsistent with the target letters, but does not activate detectors for new letters inconsistent with the target word. Suppose, however, that a letter mask were used instead. If the mask letters were properly chosen (so that no letter occurred in the same position in both the target and the mask), they should activate detectors for letters inconsistent with the target word. Active detectors for mask letters should then send inhibitory input to the detector for the target word. With a letter mask, then, the disruptive effects of upward inhibition should tend to deactivate detectors for target words as well as detectors for target letters. We would therefore expect a letter mask to produce a much smaller Word–Letter Phenomenon than a feature mask. There might still be a small residual Word–Letter Phenomenon if central attentional processes have a greater chance of encountering either a briefly activated word detector or a briefly activated letter detector (for a target word) rather than only a briefly activated letter detector (for a target letter).

The expectation that use of a letter mask would reduce the size of the Word–Letter Phenomenon depends critically upon the absence of interference effects between mask letters and target letters. Such "within-level" interference effects might occur if there were inhibitory connections between detectors for different letters, or if the activation of detectors for mask letters interfered with the encoding of information from detectors for target letters. If either kind of interference between mask letters and target letters did occur, use of a letter mask would reduce performance on target letters. This would tend to increase the Word–Letter Phenomenon, offsetting the decrease predicted from our model. We would prefer to find a test of the model that would be less likely to go astray if within-level interferences effects do indeed exist.

Fortunately an alternative manipulation is available: substitute word masks for letter masks, and measure the Word–Letter Phenomenon as before. If within-level interference effects do no exist, the prediction of our hierarchical model remains unchanged. A word mask, like a letter mask, contains letters, and these should provide inhibitory input to the detector for the target word; thus the Word–Letter Phenomenon should be greatly reduced or eliminated. If within-level interference effects do exist, interference effects between detectors for mask words and target words should now occur as well as interference effects between detectors for mask letters and target letters. Although there is no reason why the former should be exactly the same magnitude as the latter, at least they would tend to have offsetting effects on the size of the Word–Letter Phenomenon measured.

Therefore, we decided that the first prediction from our hierarchical model to be tested would be that a word mask should reduce or eliminate the Word–Letter Phenomenon obtained with a feature mask. One previous study (Taylor & Chabot, 1978) has measured performance on word and letter targets with a word mask and what we interpret as a type of feature mask (actually it consisted of heavily overstruck, overlapping letters). Taylor and Chabot's data (Figure 2, middle and right panels, intermediate SOAs) appear to show a smaller Word–Letter Phenomenon with word masking. Taylor and Chabot were not primarily interested in this aspect of the data, and the relevant statistical comparison was not reported. The results were, however, at least in the direction predicted by our model. Unfortunately, the word masks and what we interpret as feature masks differed greatly on superficial low-level properties (number and density of segments, average luminance, etc.) and overall thresholds were far apart (approximately twice as high for the word masks) so that the authors themselves found it difficult to interpret between-mask differences in performance. In order to minimize this problem in the present study, the properties of feature masks and work masks were carefully matched.

EXPERIMENT 1

Method

Subjects. Subjects, who were paid for their participation, were 32 students in the Murray Hill, N.J. area (grades 9–13) with normal or corrected-to-normal vision. Ten subjects were replaced because their performance averaged over all conditions was outside of a prespecified band of 64 to 86% correct forced-choice responses. One subject, whose exposure threshold (124 milliseconds) was well outside the range of the rest of the subjects was also replaced.

Materials. The alphabet was divided into two sets: 14 target letters (C,D,E,F,H,J,L, M,O,P,S,T,U,Z) and 11 mask letters (A,B, G,I,K,N,R,V,W,X,Y); Q was not used. The two sets were chosen so that a sufficient number of words could be spelled using only the letters within each set. The character font was designed using a restricted set of component line segments on a 17 × 25 dot grid (see Figure 2). The target and mask letters were chosen so that each mask letter contained at least one segment that was inconsistent with each target letter. For example, the bottom horizontal segment in the mask letter I was inconsistent with the target letter T.

Eleven "feature characters" were next designed for use in feature masks. We attempted to match the set of feature characters and the set of mask letters for number and type of letter features present, while at the same time minimizing the extent to which any particular feature character resembled any particular letter. The 11 mask letters were first divided into five groups (A,N), (I,Y,R), (X,V), (B,W), (K,G).

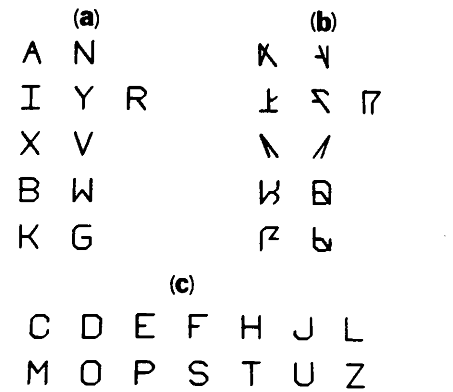

FIG. 2. Character sets used in experiments. Within each of the top five rows, each group of feature-mask characters contains the same component segments as the matched set of letter-mask characters.

Within each group, component segments were interchanged to create five matching groups of feature characters (see Figure 2). Thus the entire set of feature characters contained the same component segments in the same locations as the entire set of mask letters. In addition, the character sets were exactly matched for number of free segment ends, and reasonably well matched for number of segment junctions and angles (see Table 1).

A set of 18 4-letter word masks was constructed from the mask letters with the special property that over the entire set, the characters within each of the five groups (e.g., A and N) were used equally often. A set of 18 4-character feature masks was

TABLE 1

NUMBER OF OCCURRENCES OF POSSIBLE FEATURES IN MASK CHARACTER SETS

	Letter characters	Feature characters
Segment ends	27	27
Segment junctions	37	41
2-ray junctions	(25)	(32)
3-ray junctions	(11)	(7)
4-ray junctions	(1)	(2)
Total angles	54	55
Acute	(12)	(15)
Right	(17)	(16)
Obtuse	(25)	(24)

then constructed so that the feature characters in each group appeared with the same frequency as the corresponding mask letters from which they were derived. Corresponding groups of feature characters and mask letters appeared in the same serial positions with almost equal frequency (±1 occurrence). The result was that each component segment from which characters were composed appeared with exactly the same frequency in the 18 word masks and the 18 feature masks. Furthermore the frequency of occurrence of each component segment in each of the four letter positions was very closely matched for the set of word masks and the set of feature masks.

Word targets consisted of 72 pairs of 4-letter words composed from the 14 target letters. Each pair differed in only one critical letter (e.g., MUST, MOST); the critical letter occurred in each of the four possible letter positions equally often. Seventy-two pairs of single-letter stimuli were formed by deleting the three noncritical letters from each word, leaving the critical letter in the same position. For practice trials, an additional 48 word-pairs and matched single letter pairs were constructed in the same way.

Procedure. The sequence of events for a trial was as follows. The subject sat facing a CRT display. Each trial began with a fixation field consisting of a rectangle with vertical marks above and below positions where letters would appear. Next the stimulus was presented, followed immediately by a 4-character mask. The mask was displayed for 750 milliseconds, followed by two choice alternatives. One alternative was the target word or letter actually presented. The other alternative was the pairmate of the target, differing only in the critical letter position. The subject indicated his choice by pressing either a left or a right button. A subsequent confidence rating was also made, but rating data will not be presented here.

Visual Conditions. Background illumination of the CRT screen was .3 ftL. A filled matrix of dots with the same spacing and intensity as the target items displayed had a luminance of 2.5 ftL (5 ftL for the masks). At the viewing distance of approximately 160 centimeters individual characters subtended an angle of .25° in width and .37° in height. An entire 4-character string was approximately 1.3° wide. The mean exposure duration of targets was 61.5 milliseconds.

Design. The target items were divided into two lists each containing one member of each pair; 16 subjects viewed each list. Each subject served for one session consisting of 12 20-trial blocks. Each block was homogeneous, consisting of one of the four possible combinations of word or letter targets, and feature or word masks. Subjects cycled three times through these four conditions, alternating between word and letter targets every block, and between word and feature masks every two blocks. The order of target and mask block rotation was counterbalanced across subjects. The first cycle (four blocks) was used for practice and estimation of the exposure duration which would produce about a 75% performance level averaged across the different conditions. Data were collected on the last 18 trials of each block in the next two cycles (Blocks 5–12). (The first two trials contained more practice items and were discarded in data analysis.) Exposure duration was adjusted between the second and third cycles (after Block 8) if overall performance deviated substantially from the desired 75% level. Thus data from all four conditions were collected at the same exposure duration within each cycle. Each subject saw each target word and the corresponding target single letter once; each subject saw each word mask and feature mask four times (six times including practice trials). The assignment of masks to targets was arranged so that pooled over all subjects, each mask was paired equally often with a word target and its corresponding letter target.

Results

The results of Experiment 1 are shown in

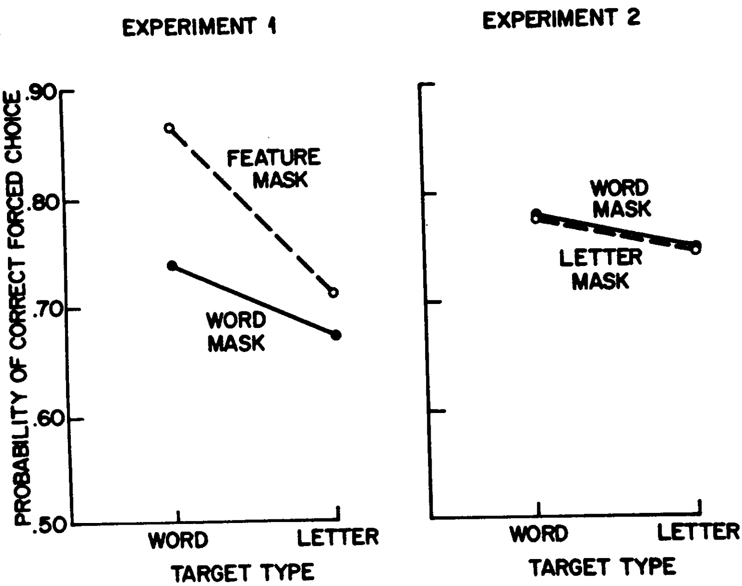

FIG. 3. Results of Experiment 1 (left panel), comparing performance with feature masks and word masks, and Experiment 2 (right panel), comparing performance with letter masks and word masks.

the left panel of Figure 3. Statistical analyses were carried out using the arcsine transformation of probability correct (Winer, 1962) as the dependent variable. Analyses of variance were carried out testing separately the reliability of trends measured across three random variables: subjects (statistics subscripted *s*); target items (statistics subscripted *i*), and masking arrays (statistics subscripted *m*).

Performance on word targets was 10.8% more accurate than on letter targets [$F_s(1,31) = 31.9, p < .001$; $F_i(1,68) = 52.4, p < .001$; $F_m(1,16) = 65.6, p < .001$]. Overall performance with feature masks was 7.6% more accurate than with word masks [$F_s(1,31) = 26.3, p < .001$; $F_i(1,68) = 36.4, p < .001$; $F_m(1,48) = 47.6, p < .001$]. However, mask type and target type interacted significantly [$F_s(1,31) = 13.4, p < .05$; $F_i(1,68) = 18.9, p < .001$; $F_m(1,16) = 10.6, p < .01$]. The direction of the interaction was as predicted. With feature masks the Word–Letter Phenomenon was quite large: 15.6%, which is comparable in size to the results found by Johnston and McClelland (1973), Johnston (1978), and McClelland and Johnston (1977) with other types of feature masks. With word masks, however, the size of the Word–Letter Phenomenon was much smaller (6.2%). Newman–Keuls tests were performed to see which pairs of the four cells (word vs letter targets by feature vs word masks) differed from each other significantly. Performance on word targets with feature masks was significantly higher than for all three other cells ($p < .01$ for tests across subjects, items and masks). The difference between word and letter targets for word masks was not quite significant at the .05 level (for each test). No other differences between pairs of cells approached significance.

Of particular interest is the surprisingly small (3.3%) difference between single letter performance with word masks and feature masks (cf. the large difference obtained by Taylor & Chabot, 1978, when no attempt was made to match the feature properties of the two types of masks). According to our model, this result supports two conclusions: (a) Enough letter features were incorporated into our feature masks for them to have nearly the full upward inhibitory effects of actual letters; and (b) Within-level interference (at least between mask letters and target letters) is not a major factor in determining performance under those experimental conditions. If the small, nonsignificant, difference found is real, it might be attributed either to incomplete success in incorporating appropriate features in the feature-mask characters or to small within-level interference between mask letters and target letters.

The analyses of variance turned up significant differences in performance across the positions in which the critical letter could appear. There was a significant main effect of serial position of the critical letter tested [$F_s(3,93) = 4.5, p < .01$; $F_i(3,68) = 2.9, p < .05$; $F_m(3,48) = 4.7, p < .01$] and a significant interaction of target type with serial position [$F_s(1,310) = 13.4, p < .005$; $F_i(1,68) = 18.9, p < .001$; $F_m(1,16) = 10.6, p < .01$]. Examination of serial position curves (see left panel of Figure 4) showed that single letter performance was roughly flat across serial position, while performance on word targets was higher in the left-most position. The residual Word–Letter Phenomenon remaining with the word mask was heavily concentrated in the left-most letter position (13% for that position alone).

Experiment 2

Our hierarchical model stresses the role of upward inhibitory interactions in the detector network. According to our model, the word masks used in Experiment 1 greatly reduced the Word–Letter Phenomenon because they contained letters, not because the letters spelled words. The mask letters were arranged as words so that any mask letter-target letter interference would tend to be cancelled out by mask word-target word interference, leaving a less biased measurement of the Word–Letter Phenomenon. Since Experiment 1 failed to find evidence of substantial within-level inhibitory effects between mask letters and target letters, it seemed reasonable that within-level inhibitory effects of mask words on target words might also be negligible. If so, a mask made up of letters

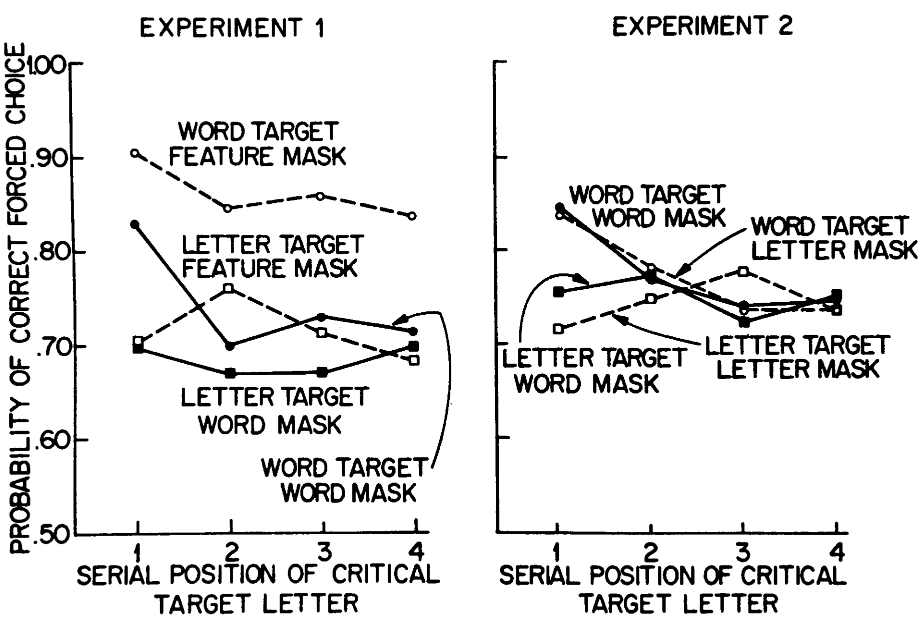

FIG. 4. Performance by serial position of the critical letter tested for Experiment 1 (left panel) and Experiment 2 (right panel).

would produce only a small Word–Letter Phenomenon whether it spelled a word or not. This expectation was tested in Experiment 2.

Method

Experiment 2 used exactly the same procedure, design and target items used in Experiment 1. Only the masks were different. A new set of 18 word masks and a matching set of 18 letter masks (four letters per mask) were constructed from the same pool of 11 mask letters used in Experiment 1. All of the letter masks were unpronounceable and orthographically irregular strings of letters. Word masks and letter masks were balanced so that the frequencies of each letter in each serial position were identical. Thirty-two new subjects were used (five were replaced because the average error rate was outside the prespecified limits). The mean exposure duration for this experiment was 72.5 milliseconds.

Results

The results of Experiment 2 are shown in the right panel of Figure 3. Analysis of the data was carried out as in Experiment 1.

Results show virtually identical performance with word and letter masks: a very small Word–Letter Phenomenon in each case (2.4% with word masks and 2.2% with letter masks). The main effect for target type failed to reach significance tested across subjects ($F < 1$) although it was significant across items ($F_i(1,68) = 4.7$, $p < .05$) and masks ($F_m(1,16) = 5.2$, $p < .05$). As one would expect from Figure 3, the main effect for mask type and the interaction of mask type and target type did not approach significance ($F < 1$ for each test). The data thus confirm the prediction that word masks and letter masks should have almost identical effects.

Although serial position of the critical letter had no overall significant effect ($F < 1$ for each test) the interaction of stimulus type and serial position was again significant [$F_s(3,93) = 6.4$, $p < 0.01$; $F_i(3,18) = 4.0$, $p < .05$; $F_m(3,48) = 5.5$, $p < .005$]. As in Experiment 1, this interaction (see right panel of Figure 4) was mainly due to elevation in performance for word targets tested in the first position for both mask types.

EXPERIMENT 3

Experiment 1 showed that the large Word–Letter Phenomenon produced with feature masks could be greatly reduced by using word masks. Experiment 2 showed that letter masks produced virtually the same performance as word masks. One would therefore expect that, if compared directly, letter masks would reduce the Word–Letter Phenomenon found with feature masks. Although this would seem to be a sound inference, we felt it was important to verify it in a third experiment directly comparing the Word–Letter Phenomenon with letter masks vs the Word–Letter Phenomenon with feature masks.

Experiment 3 was also designed to clear up a possible concern about the mask-type-by-target-type interaction found in Experiment 1. Even though performance in all conditions was well above the 50% chance guessing level, it remains true that overall performance with word masks was below overall performance with feature masks. One cannot therefore totally rule out the possibility that it was the lower level of performance with word masks that reduced the size of the Word–Letter Phenomenon, not the use of word masks per se. This concern is partially alleviated by the small size of the Word–Letter Phenomenon found with word (and letter) masks in Experiment 2 where the overall performance level was considerably higher. Experiment 3 was designed to entirely eliminate the problem by equalizing overall performance with the two types of masks used. An interaction of mask type and target type, if found, would then be a "crossover" interaction, which would be difficult indeed to attribute to a scaling artifact.

Method

The method for Experiment 3 was the same as for the first two experiments with several simple modifications. The procedure on a trial and the target stimuli were kept entirely unchanged. Letter masks were exactly the same as in Experiment 2; feature masks were composed from the feature characters used in Experiment 1. Each component segment appeared in the same letter position equally often in feature masks and letter masks. A pilot experiment showed that overall performance levels could be equated by setting exposure durations 10-milliseconds longer with the letter masks than with the feature masks. A 10-millisecond differential was therefore used throughout Experiment 3, with the yoked pair of durations adjusted for each subject to produce approximately 75% correct performance overall. Sixteen new subjects were run; one was replaced because the average error rate was outside the prespecified limits. Exposure durations averaged 51.5 milliseconds with the feature mask and 61.5 milliseconds with the letter mask.

Results

Results are shown in Figure 5. The attempt to equate performance levels on each type of mask at about the 75% performance level was very successful: letter-mask performance averaged 75.4% and feature mask performance averaged 75.1%. As hypothesized, results showed a much larger Word–Letter Phenomenon for the feature mask (20.7%) than for the letter mask (7.5%). Analysis of variance showed no main effect of mask type ($F < 1$ in each test), a significant overall advantage for word targets [$F_s(1,15) = 36.8, p < .001$; $F_i(1,32) = 38.2, p < .001$; $F_m(1,16) = 107.4, p < .001$], and a significant interaction between mask type and target type [$F_s(1,15) = 14.4, p < .005$; $F_i(1,32) = 12.9, p < .005$; $F_m(1,16) = 16.3, p < .001$]. This interaction was a crossover interaction as predicted: performance on word and letter targets with the letter mask

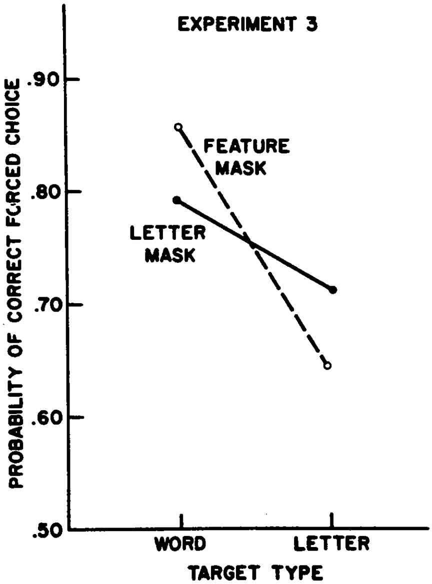

FIG. 5. Results of Experiment 3, comparing performance with feature masks and letter masks. In order to equalize overall performance with each mask type, exposure durations were 10-milliseconds longer with letter masks than with feature masks.

fell between performance on word and letter targets with the feature mask. Because different mask types were run at different exposure durations, the only meaningful comparisons among the four individual conditions are the comparisons within mask type. Newman–Keuls tests showed that the Word–Letter Phenomenon with feature masks was highly significant ($p < .01$ for each test) and the Word–Letter Phenomenon with letter masks was marginally significant ($p < .05$ tested across items and across masks, but $p > .05$ tested across subjects). Serial position curve data for Experiment 3 were qualitatively similar to data from the corresponding conditions in the first two experiments. Performance on word targets showed a left-to-right drop-off. This drop-off was slightly steeper with letter masks than with feature masks. Letter targets showed higher performance on the middle positions for each mask type. Although these data are consistent with the data from earlier experiments, fewer trials per cell were collected, and analysis of variance showed that only the main effect of critical-letter position reached significance [$F_s(3,45) = 10.3, p < .001; F_i(3,32) = 5.7, p < .01; F_m(3,48) = 6.9; p < .001$]; no interaction of position with any other variable was significant.

General Discussion

Testing the Hierarchical Model

Results of these experiments support our hierarchical model of word identification. According to the model, the major cause of the Word–Letter Phenomenon is that mask features inhibit detectors for target letters but not detectors for target words. The key prediction derived was that masks containing letters should deactivate detectors for target words as well as for target letters, thus reducing the size of the Word–Letter Phenomenon. This prediction was confirmed. Experimental results showed a large and significant reduction in the size of the Word–Letter Phenomenon when letter masks or word masks were used rather than feature masks (the Word–Letter Phenomenon was reduced by 9.6% in Experiment 1, and by 13.0% in Experiment 3).

In each of the four conditions where a mask containing letters was tested, a small residual Word–Letter Phenomenon was found. This residual reached 7.5% and was marginally significant in Experiment 3. The existence of a residual Word–Letter Phenomenon with masks containing letters is consistent with our hierarchical model. With masks containing letters, our model provides no reason to expect that word detectors will remain active longer than letter detectors. It is the case, however, that a word target should briefly activate detectors at three levels (feature, letter, word), while a letter target should only activate detectors at two levels (feature, letter). The additional word detector activated by a word target should provide central attentional processes with an *additional* opportunity to locate at least one detector specifying the identity of the critical letter. This additional opportunity should help performance, even if this opportunity in itself is no more likely to be capitalized upon than the others. (Performance would fail to improve only if success in locating active word detectors were totally correlated with success in locating active letter or feature detectors, an unlikely possibility.)

The multiple-opportunity interpretation provides a sensible explanation of the finding in each experiment that the residual Word–Letter Phenomenon with masks containing letters was largest when the critical letter appeared in the left-most position. Performance with a word target should be a mixture of trials mediated by locating the word detector and trials mediated by locating the detector for the critical letter. The probability of locating the word detector should not depend on the position of the critical letter. However, the probability of locating the detector for the critical letter might depend on position if subjects encoded information from the letter level using a left-to-right scan. In this case, the left-most letter would have the greatest probability of being encoded during the brief moment that letter detectors were active. It is worth noting that the serial-position curves for word targets with masks containing letters—which in fact show steep left-to-right drop-offs in performance—resemble the serial-position curves we have obtained in experiments using unrelated-letter strings as targets (unpublished data, McClelland & Johnston, 1977). If our model is correct in the claim that a mask containing letters greatly reduces the availability of information from word detectors, it should not be surprising that, with such a mask, subjects often revert to processing a displayed word as a string of unrelated letters.

An especially striking aspect of the present results is the lack of evidence for interference effects within the letter level and within the word level. In Experiment 1 performance on letter targets was only slightly (and not significantly) worse with masks that contained letters (word masks) than with masks that did not contain letters (feature masks). In Experiment 2 performance on word targets was virtually identical with masks that contained a word and with masks that did not contain a word. Thus there is no need to add within-level connections to the hierarchical model in order to explain the present data. We are reluctant to conclude, however, that within-level interference effects simply do not exist at all because of evidence from other paradigms (e.g., Bjork & Murray, 1977; Estes, 1972; Eriksen & Eriksen, 1974; Gardner, 1973; Wolford & Hollingsworth, 1974). It is possible that bottom-up inhibitory effects in the present paradigm occur so rapidly and completely that within-level inhibition is irrelevant. If a letter detector activated by a target letter were deactivated very rapidly by bottom-up inhibition from feature detectors activated by the mask, that letter detector might already be inactive before the arrival of within-level inhibition from letter detectors activated by the mask. Analogous reasoning applies to possible word–word inhibition. This explanation is consistent with the very rapid and decisive effects of bottom-up inhibition built into the specific version of the our model presented in the appendix.

Alternative Models

The specific combination of results found here appears to rule out several alternative explanations. The results cannot be explained by the hypothesis that detectors for similar types of units inhibit each other most strongly (Jacobson, 1974; Mayzner & Tresselt, 1970; Smith, Haviland, Reder, Brownell, & Adams, 1976). If this hypothesis were correct, then letter masks should have especially disrupted performance on letter targets, producing a larger Word–Letter Phenomenon than normal, instead of the smaller Word–Letter Phenomenon actually obtained. Furthermore, word masks should have depressed performance on word targets relative to letter masks and they did not. The pattern of results obtained is equally incompatible with interference effects among similar types of units in encoding or retrieval from short-term memory (e.g., Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). In our data, it is the "upward" relation between mask properties and target properties that produces interference: it is the presence of letter features in masks that interferes with performance on single letter targets, producing the Word–Letter Phenomenon found in previous experiments and in the present feature-mask condition; it is the presence of letters themselves in masks that interferes with performance on word targets, reducing the size of the Word–Letter Phenomenon found in letter-mask and word-mask conditions. Our model relies straightforwardly on "upward" inhibition to account for these effects of different types of masks.

The pattern of results we obtained is difficult to explain by models which do not suppose that word identification is based on component letters. If letter identification and word identification occur in separate channels, why is it that the presence of letters in a mask differentially interferes with target words? The problem is particularly clear for models (e.g., Smith, 1971) which suppose that the same visual features provide the input to both letter identification and word identification. Since the only upward connections are feature to letter and feature to word, upward inhibitory effects should be the same for all types of masks that contain features (e.g., our feature, letter, and word masks). Even if our letter masks contained more letter features than our feature masks, there would be no reason to expect these features to have inhibited detectors for target words any more than for target letters. An advocate of nonhierarchical word identification is free, of course, to rely on other kinds of interference besides the upward inhibition on which we rely. But we have already seen that the obvious alternative sources of interference (letter–letter or word–word interference, either in the detector network or in later processing) also would not operate in the appropriate direction to explain our data.

Further Development of the Hierarchical Model

We believe that further development of the hierarchical model is warranted not only by its success in generating the predictions confirmed in this paper, but also by its success in accounting for other recent data on word and letter-string processing. The hierarchical model has been used by McClelland (1976) to account for people's ability to read mixed-case words (see also Adams, 1979), and to transfer new vocabulary words from one calligraphy to another (McClelland, 1977). Johnston (1978) has used the hierarchical model to reconcile the existence of the Word-Superiority Effect and the Word–Letter Phenomenon with the surprising absence of effects of within-word redundancy on the accuracy and speed of word identification (Colthart, Davelear, Jonasson, & Besner, 1977; Johnston, 1974, 1978).

One direction in which further work is needed is the development of more complex and realistic quantitative versions of the model. The discrete-state version of the model presented in the Appendix is very unlikely to be literally true. Detectors presumably have a continuous range of activation levels, and the rate at which detectors at one level influence detectors at the next level presumably varies continuously as well (McClelland, 1979). The Appendix discusses the extent to which the discrete-state version may be a reasonable approximation under the favorable viewing conditions usually employed to obtain the Word–Letter Phenomenon, and suggests some avenues toward more sophisticated formulations of the hierarchical model.

Further work would also be desirable to clarify the relationship of the hierarchical model to two other phenomena:

1) With patterned masking, letters in orthographically regular, pronounceable pseudowords, like letter is words, show a perceptual advantage over letters alone (Johnston & McClelland, 1977). Perhaps the most obvious way to adapt the hierarchical model to account for this phenomenon would be to postulate the existence of another level of detectors for intermediate-size letter-cluster units between the letter and word levels. There are good reasons to believe that this proposal is not adequate. If detectors for familiar letter-clusters provide a major input to word detectors then masks which activate more letter-cluster detectors should produce more inhibitory input to word detectors, reducing the Word–Letter Phenomenon. In fact the present data show that unrelated-letter masks, which contain fewer familiar letter-clusters than word masks, are equally effective in reducing the Word–Letter Phenomenon. Other evidence also casts doubt on the letter-cluster detector hypothesis. McClelland and Johnston (1977) found no effect of letter-cluster frequency on performance in a study which did find evidence for effects of position-specific letter frequency and word frequency. They argued that pseudoword representations are not formed by activating letter-cluster detectors, but rather by a mere flexible, rule guided, constructive process. In order to explain the perceptual advantage of a letter in a pseudoword over a letter alone, our model requires only that pseudoword representations be responsive to both confirming and disconfirming input from letter detectors. If that is true, pseudowords, like words, would be shielded from the effects of feature masks, but not from the effects of masks containing letters (see Johnston, in press, for further discussion).

2) Although we have assumed for convenience that the position-specificity of the detector network is perfect, there is considerable evidence that people frequently report the letters from a tachistoscopic display correctly, but in the wrong order (e.g., Estes, 1975; Johnston, 1978; McClelland &

Johnston, 1977). The fact that these position errors are much more common for unrelated-letter strings than for words, suggests that much of the problem may occur not in the network, but in the process of encoding information from it and making an overt report. (Alternatively, as Estes argues, it may be that, if the target is a word, many position errors at the letter level can be corrected using lexical knowledge.) To the extent that letter position errors do actually occur within the detector network, an improved version of the hierarchical model should reflect this fact. Such a revision would lead to the prediction of a bias against word targets in all conditions of our experiments (for single letter targets a mistake in localization should not produce an error on our forced-choice test).

Methodological Implications

These experiments demonstrate that tachistoscopic masking methods can be used to test specific predictions of processing models. However, masking is frequently used simply as a way to increase the difficulty of the perceptual task by "degrading input" or by "stopping processing" (e.g., Kahneman, 1968; Lupker, 1979). If models such as the one advocated here are correct, interpretation of these traditional types of experiments will become quite complicated. Posttarget masks may well degrade peripheral representations of the target, but they may also produce new activity which will interact with activity produced by the target in complex ways (e.g., our conclusion that detectors for mask letters inhibit detectors for a target word). As for "stopping processing," the same mask may have quite different effects at different levels. For instance if a feature mask inhibits detectors for target letters, it may indeed prevent central attentional processes from encoding them. However, processing may continue for target word representations which survive the mask.

Our analysis suggests that extreme caution is required in interpreting reaction time experiments that measure the total time to perform a task. Of particular concern are periodic attempts (from Cattell, 1886, to Johnson, 1975) to show that word identification is as fast as letter identification. The implication is that, if this is so, the latter cannot mediate the former. However, according to the present model, the time it takes to make any overt response indicating identification will be a function of both how soon an appropriate detector is activated *and* how soon central attentional processes can locate, retrieve, and use the output of the detector. Even if word detectors take several milliseconds longer to be activated than letter detectors, no prediction can be made about the total time to make overt identification responses to letters and words without some independent knowledge about the process of retrieving information from detectors. This situation is even more complex if one attempts to compare the time to identify letters *in* a word to the time it takes to identify the word (Johnson, 1975). Depending on the search procedures used by central attentional processes, an active detector for the displayed word might sometimes be the first detector located. This event could easily delay the process of coding information from active letter detectors, increasing the mean response time for letters in words. Thus the measured time to "identify" a word and the measured time to "identify" one of its component letters could well be entirely misleading as to the timing of activation of detectors for the letter and for the word.

CONCLUSION

The hierarchical model presented here, and the experimental evidence corroborating it, serve to reconcile the existence of the Word–Letter Phenomenon with other evidence that word identification is mediated by letter identification (McClelland, 1976, 1977). The current account is consistent with what is presently known about the

Word–Letter Phenomenon, and with several recent models of cognitive processes which distinguish between the spread of activation within a detector network, and the encoding of information from the network by central attentional processes.

Appendix

Generation of the Word–Letter Phenomenon by a Discrete-State Version of the Hierarchical Model of Word Identification

The text presents our hierarchical model of word identification in a general form and derives from it an account of the Word–Letter Phenomenon. In this Appendix we will present a specific, more detailed version of our hierarchical model, treating detectors as discrete two-state (on–off) devices. Although the specific model is almost certainly an oversimplification, we will be able to use it to demonstrate more concretely that the Word–Letter Phenomenon does indeed follow from at least one version of our model. Some readers may also find this account easier to follow than the text because a state diagram can be provided (Figure 6).

The specific version of the hierarchical model makes the following assumptions: (1) Word identification is accomplished by a system consisting of a preprocessor, a network of detectors for features, letters and words, and a central processor which transfers information from detectors to a response buffer. (2) The preprocessor segregates display information into characters, and sends each character to the appropriate letter-position channel of the detector network. (3) Detectors are wired together as shown in Figure 1. Feature detectors and letter detectors are segregated into channels for each letter position, as shown in Figure 1. Feature detectors have output connections only to letter detectors, which in turn have output connections only to word detectors. A detector for unit A has an excitatory connection to the detector for unit B at the next level if the presence of A would be evidence for the presence of B. Otherwise the detector for A has an inhibitory connection to the detector for B (except that there are no feature-to-letter connections across position-channel boundaries; see Figure 1). (4) Each detector behaves as an all-or-none, two-state device (either active or inactive). (5) To be activated a detector must simultaneously receive an excitatory signal from all potentially excitatory detectors at the previous level. (6) Any inhibitory signal is decisive. That is, any inhibitory signal will prevent a detector from being activated, and any inhibitory signal is sufficient to deactivate a previously active detector. (7) Since feature detectors are the first level of detectors in the network, their functioning is assumed to depend directly on display properties. All features consistent with the display will be activated, and all others will be deactivated. (8) Detectors are activated or deactivated with a constant transmission delay after input conditions are met. Thus detectors for features are activated or deactivated with a delay of t_f after display onset (the value of t_f will depend on display conditions). Detectors for letters are activated or deactivated with a delay of t_l after feature detectors are activated. Detector for words are activated or deactivated with a delay of t_w after letter detectors are activated. (9) If an active detector receives no new input, either excitatory or inhibitory, for an extended period of time, we assume that it is subject to a probabilistic decay back to the inactive state. The precise distribution of decay probability over time need not be specified, but we assume that decay takes substantially longer than deactivation by inhibition. (10) We also do not need to specify in detail how the central processor works. For our purposes it is sufficient to assume that the probability of finding an active detector and encoding its identity in the response buffer is an increasing function of the length of time that a detector is active.

We will now show how the specific ver-

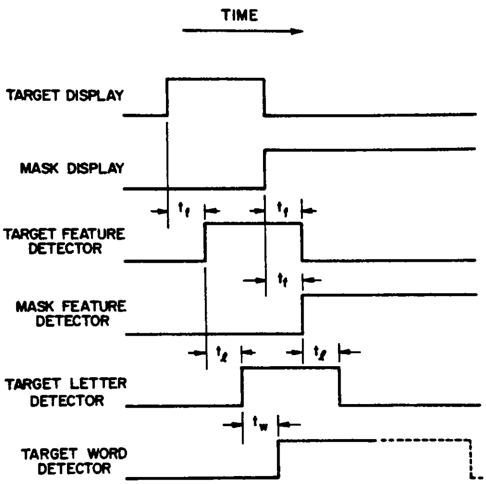

FIG. 6. Timing diagram representing events hypothesized to account for the Word–Letter Phenomenon. Parameters shown represent time to activate or deactivate a feature detector (t_f), a letter detector (t_l), or a word detector (t_w). Dotted line for word detector offset represents slow probabilistic decay.

sion of the hierarchical model just presented accounts for the Word–Letter Phenomenon produced with a traditional patterned mask. Figure 6 provides a diagram of hypothesized state changes in the network over time. Trials begin with warning signals followed by display of a target letter or word. After a delay of t_f following target onset, detectors for features consistent with the target are activated. After a further delay of t_l, detectors for letters specified by active feature detectors are activated. Finally, after a further delay of t_w a word detector (if any) specified by active letter detectors is activated. When the target display is terminated and the mask display begins, another chain of events is initiated. After a delay of t_f, detectors for features present in the mask are activated, and any active detectors for features not present in the mask are deactivated. Thus at the feature level a representation of the mask replaces the representation of the target. The mask presented is normally chosen to contain some features that are inconsistent with all target letters. Thus, after a further delay of t_l, all letter detectors previously activated by the target are deactivated by inhibitory input from feature detectors activated by the mask. If a word detector was previously activated by the target, it receives no further excitatory input. However, it also receives no inhibitory input from the mask, since a feature mask does not activate any letter detectors. Thus if any word detector was activated by the target, it will be subject only to passive decay, hypothesized to be a slower process than deactivation by inhibition. We have now shown that when the target is a word rather than a letter, information about the target is represented in the network by an additional active word detector, which is active for a longer period of time than letter or feature detectors. According to Assumption 10, central processes will thus have a higher probability of successfully encoding information from an active word detector, leading to higher performance on word targets than on letter targets. The *additional* encoding opportunity provided by an active word detector provides a secondary cause of the Word–Letter Phenomenon (assuming that success in encoding information from a word detector is less than perfectly correlated with success in encoding information from lower level detectors). This secondary cause of the Word–Letter Phenomenon should operate even with masks containing letters, where word detectors do not remain active longer than lower-level detectors.

We believe that the specific version of the hierarchical model presented in this Appendix provides useful insights about how the general model operates. However, the specific model relies on a number of oversimplifications. We would like to mention three of these and discuss how they might be remedied in more advanced versions of the general hierarchical model.

(1) Detectors are characterized as two-state (on–off) devices with all-or-none output levels (Rule 4). Clearly it would be more realistic to treat detectors as devices with continuous activation levels, and continuous output levels. Cascade models (McClel-

land, 1979) might provide a useful parametric treatment for these aspects of the functioning of the detector network. We believe that the discrete treatment provided here will probably turn out to provide a fairly close first approximation for the visual conditions normally used to produce the Word–Letter Phenomenon. Target letters and words are usually presented to a relatively light-adapted visual system, with a high contrast ratio, using a standard type font, and subtending a visual angle close to that for normal reading. Under these conditions the output of detectors could reasonably be expected to rise rapidly to near-asymptotic saturation levels. Under less favorable viewing conditions, detectors could reasonably be expected to perform more often at intermediate levels, and the continuous treatment would deviate more substantially from the discrete treatment.

(2) In the general hierarchical model inhibition clearly plays a very important role. In the specific model any inhibition is treated as absolutely decisive, so that no amount of concurrent excitation can override it (Rule 6). Actually this assumption is no more severe than the essentially equivalent assumption in recent models of character and word identification that only alternatives *consistent* with feature information are considered by decision processes (e.g., Rumelhart, 1971; Rumelhart & Siple, 1974). It is critical to the hierarchical model that inhibition must, at the least, have a very strong, if not necessarily overwhelming, weight relative to excitation. Otherwise detectors for units that were not present in a display, but shared many properties with the display, might frequently be activated. In particular, patterned masks not containing letters would be likely to activate some letter detectors, (especially for letters made up of a subset of the segments in the mask). These active letter detectors would in turn inhibit word detectors. Word detectors would thus be less well shielded from disruption by a patterned mask, and the Word–Letter Phenomenon would be reduced.

(3) For a detector to be activated, Rule 5 requires that all potentially excitatory units from the next lower level provide excitation (e.g., detectors for all letters in a short word must be active before the word detector can be activated). Clearly this simplifying assumption is unrealistically severe; a continuous version of the hierarchical model would rely on some function for integrating inputs of differing strengths over time to produce a resultant activation level. The present severe assumption serves importantly to highlight the difference between our account of the Word–Letter Phenomenon and the redundancy account (see introduction). Unlike the redundancy account, our account allows for a Word–Letter Phenomenon even if activation of a word detector was *always* accompanied by activation of the detector for the component letter critical to the forced choice.

Rule 5 serves to prevent the simultaneous activation of several different words sharing many common letters with the target word. If a relatively weak criterion for activating detectors were adopted, then multiple word detectors would be more likely to be activated by the target. One might expect that the probability of correctly identifying a word would then vary inversely with the number of similarly spelled "neighboring" words (to be "competed" against). Johnston (1978) has found no measurable trend of this kind using visual conditions that produce a strong Word–Letter Phenomenon. In order to account for this fact, it appears that other versions of the hierarchical model which relax the severe "input from all potentially excitatory units" rule must either use a close approximation to the rule, or else rely on some alternative means of preventing "competition" from detectors for similarly spelled words. (One possibility is to assume that the target almost always stimulates an abundance of feature detectors, with only rare "misses," so that strong inhibition from disconfirming feature detectors can be relied upon to prevent multiple competing activations.)

References

ADAMS, M. J. Models of word recognition. *Cognitive Psychology,* 1979, 11, 133–176.

BADDELEY, A. D. Immediate memory and the "perception" of letter sequences. *Quarterly Journal of Experimental Psychology,* 1964, 16, 364–367.

BARON, J., & THURSTON, I. An a analysis of the word-superiority effect. *Cognitive Psychology,* 1973, 4, 207–228.

BJORK, E., & ESTES, W. K. Letter identification in relation to linguistic context and masking conditions. *Memory and Cognition,* 1973, 1, 217–223.

BJORK, E., & MURRAY, J. T. On the nature of input channels in visual processing. *Psychological Review,* 1977, 84, 472–484.

CARR, T. H., DAVIDSON, B. J., & HAWKINS, H. L. Perceptual flexibility in word recognition: Strategies affect orthographic computation but not lexical access. *Journal of Experimental Psychology: Human Perception and Performance,* 1978, 4, 674–690.

CATTELL, J. M. The time taken up by cerebral operations. *Mind,* 1886, 11, 220–242.

COLTHART, M., DAVELEAR, E., JONASSON, J. T., & BESNER, D. Access to the internal lexicon. In S. Dornic (ed.), *Attention and performance, VI.* Hillsdale, N.J.: Erlbaum, 1977.

COSKY, M. The role of letter recognition in word recognition. *Memory and Cognition,* 1976, 4, 207–214.

ERIKSEN, B. A., & ERIKSEN, C. W. Effects of noise letters upon the identification of a target letter in a non-search task. *Perception & Psychophysics,* 1972, 12, 201–204.

ESTES, W. K. Interactions of signal and background variables in visual processing. *Perception & Psychophysics,* 1972, 12, 278–286.

ESTES, W. K. The locus of inferential and perceptual processes in letter identification. *Journal of Experimental Psychology: General,* 1975, 104, 122–145.

GARDNER, G. T. Evidence for independent parallel channels in tachistoscopic perception. *Cognitive Psychology,* 1973, 4, 130–155.

GEYER, J. J. Models of the perceptual process in reading. In H. Singer & R. Rudell, (eds.), *Theoretical models and processes in reading.* Newark, Delaware: International Reading Association, 1970.

GOUGH, P. B. One second in reading. In J. F. Kavanaugh & I. G. Mattingly (eds.) *Language by ear and by eye.* Cambridge, Mass: M.I.T. Press, 1972. Pp. 331–358.

HAWKINS, H. L., REICHER, G. M., ROGERS, M., & PATERSON, L. Flexible coding in word recognition. *Journal of Experimental Psychology: Human Perception and Performance,* 1976, 2, 380–385.

HENDERSON, L. Do words conceal their component letters? A critique of Johnson (1975) on the visual perception of words. *Journal of Verbal Learning and Verbal Behavior,* 1975, 14, 648–650. (a)

HENDERSON, L. Word recognition. In N. S. Sotherland (ed.), *Tutorial essays in experimental psychology.* Hillsdale, N.J.: Erlbaum, 1975. (b)

HOLENDER, D. Identification of letters in words and of single letters with pre- and postknowledge vs. postknowledge of the alternatives. *Perception & Psychophysics,* 1979, 25, 313–318.

HUBEL, D. H., & WIESEL, T. N. Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology,* 1968, 195, 215–243.

HUEY, E. B. *The psychology and pedagogy of reading.* New York: Macmillan, 1908.

JACOBSON, J. Z. Effects of association upon masking and reading latency. *Canadian Journal of Psychology,* 1973, 27, 58–69.

JACOBSON, J. Z. Interaction of similarity to words of visual masks and targets. *Journal of Experimental Psychology,* 1974, 102, 431–434.

JOHNSON, N. F. On the function of letters in word identification: Some data and a preliminary model. *Journal of Verbal Learning and Verbal Behavior,* 1975, 14, 17–29.

JOHNSTON, J. C. *The role of contextual constraint in the perception of letters in words.* Unpublished doctoral dissertation, University of Pennsylvania, 1974.

JOHNSTON, J. C. A test of the sophisticated guessing theory of word perception. *Cognitive Psychology,* 1978, 10, 123–153.

JOHNSTON, J. C. Understanding word perception: Clues from studying the word-superiority effect. In O. Tzeng & H. Singer (eds.), *Perception of print: Reading research in experimental psychology,* in press.

JOHNSTON, J. C., & MCCLELLAND J. L. Visual factors in word perception. *Perception & Psychophysics,* 1973, 14, 365–370.

JOHNSTON, J. C., & MCCLELLAND J. L. Perception of letters in words: Seek not and ye shall find. *Science,* 1974, 184, 1192–1194.

JUOLA, J. F., LEAVITT, D. D., & CHOE, C. S. Letter identification in word, nonword, and single letter displays. *Bulletin of the Psychonomic Society,* 1974, 4, 278–280.

KAHNEMAN, D. Methods, findings, and theory in studies of visual masking. *Psychological Bulletin,* 1968, 70, 404–425.

LABERGE, D. Perceptual learning and attention. In W. K. Estes (ed.), *Handbook of learning and cognitive processes, Vol. 4,* Hillsdale, N.J.: Erlbaum, 1976.

LABERGE, D., & SAMUELS, J. Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 1974, **6**, 293–321.

LUPKER, S. J. On the nature of perceptual information during letter perception. *Perception & Psychophysics*, 1979, **25**, 303–312.

MASSARO, D. Perception of letters, words and nonwords. *Journal of Experimental Psychology*, 1973, **100**, 349–353.

MASSARO, D. W., & KLITZE, D. The role of lateral masking and orthographic structure in letter and word recognition. *Acta Psychologica*, 1979, **43**, 413–426.

MAYZNER, M. S., & TRESSELT, M. E. Visual information with sequential inputs: A general model for sequential blanking, displacement, and overprinting phenomena. *Annals of the New York Academy of Sciences*, 1970, **169**, 599–618.

MCCLELLAND, J. L. Preliminary letter identification in the perception of words and nonwords. *Journal of Experimental Psychology: Human Perception and Performance*, 1976, **2**, 80–91.

MCCLELLAND, J. L. Letter and configuration information in word identification. *Journal of Verbal Learning and Verbal Behavior*, 1977, **16**, 137–150.

MCCLELLAND, J. L. On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, 1979, **86**, 287–330.

MCCLELLAND, J. L., & JOHNSTON J. C. The role of familiar units in perception of words and nonwords. *Perception & Psychophysics*, 1977, **22**, 249–261.

MEYER, D. E., & SCHVANEVELDT, R. W. Meaning, memory structure, and mental processes. *Science*, 1976, **192**, 27–33.

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 1956, **63**, 81–97.

MILLER, G., BRUNNER, J. S., & POSTMAN, L. Familiarity of letter sequences and tachistoscopic identification. *Journal of General Psychology*, 1954, **50**, 129–139.

MORTON, J. The interaction of information in word recognition. *Psychological Review*, 1969, **76**, 165–178.

NEISSER, J. *Cognitive psychology*. New York: Appleton-Century-Crofts, 1967.

POSNER, M. I., & SNYDER, C. R. R. Attention and cognitive control. In R. L. Solso (ed.). *Information processing and cognition: The Loyola Symposium*. Hillsdale, N.J.: Erlbaum, 1975.

REICHER, G. M. Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, 1969, **81**, 274–280.

RUMELHART, D. E. A multicomponent theory of perception of briefly exposed visual displays. *Journal of Mathematical Psychology*, 1970, **7**, 191–218.

RUMELHART, D. E., & SIPLE, P. Process of recognizing tachistoscopically presented words. *Psychological Review*, 1974, **81**, 99–118.

SCARBOROUGH, D. L., CORTESE, C., & SCARBOROUGH, H. S. Frequency and reptition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance*. 1977, **3**, 1–17.

SCHNEIDER, W., & SHRIFFIN, R. M. Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 1977, **84**, 1–66.

SHIFFRIN, R. M., & SCHNEIDER, W. Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Pshychological Review*, 1977, **84**, 127–191.

SMITH, E. E., HAVILAND, S. E., REDER, L. M., BROWNELL, H., & ADAMS, N. When preparation fails: Disruptive effects of prior information on perceptual recognition. *Journal of Experimental Psychology: Human Perception & Performance*, 1976, **2**, 151–161.

SMITH, F. *Understanding reading*. New York: Holt, Rinehart and Winston, 1971.

SMITH, F. The use of featural dependencies across letters in the visual identification of words. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 215–218.

SPOEHR, K. T., & SMITH, E. E. The role of orthographic and phonotactic rules in perceiving letter patterns. *Journal of Experimental Psychology: Human Perception and Performance*, 1975, **104**, 21–34.

TAYLOR, G. A., & CHABOT, R. J. Differential backward masking of words and letters by masks of varying orthographic structure. *Memory & Cognition*, 1978, **6**, 629–635.

THOMPSON, M. C., & MASSARO, D. W. Visual information and redundancy in reading. *Journal of Experimental Psychology*, 1973, **98**, 49–54.

WHEELER, D. Processes in word recognition. *Cognitive Psychology*, 1970, **1**, 59–85.

WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

WOLFORD, G., & HOLLINGSWORTH, S. Lateral masking in visual information processing. *Perception & Psychophysics*, 1974, **16**, 315–320.

WOODWORTH, R. S. *Experimental psychology*. New York: Holt, 1938.

(Received December 17, 1979)