AD-A086 946 CALIFORNIA UNIV SAN DIEGO LA JOLLA CENTER FOR HUMAN --ETC F/G 5/10
AN INTERACTIVE ACTIVATION MODEL OF THE EFFECT OF CONTEXT IN PER--ETC(U)
MAY 80 J L MCCLELLAND, D E RUMELHART N00014-79-C-0323
UNCLASSIFIED 8002 NL

1 of 2
AD
A086946

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 8002
2. GOVT ACCESSION NO.: AD-A086 946
3. RECIPIENT'S CATALOG NUMBER:
4. TITLE (and Subtitle): An Interactive Activation Model of the Effect of Context in Perception. Part I.
5. TYPE OF REPORT & PERIOD COVERED: Technical Report · Apr[redacted]79— Mar[redacted]80
6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): James L. McClelland, David E. Rumelhart
8. CONTRACT OR GRANT NUMBER(s): N00014-79-C-0323, NSF-BNS76-14830
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Center for Human Information Processing, University of California, San Diego, La Jolla, California 92093
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR 157-437
11. CONTROLLING OFFICE NAME AND ADDRESS: Personnel and Training Research Programs, Office of Naval Research (Code 458), Arlington, Virginia 22217
12. REPORT DATE: May [redacted]80
13. NUMBER OF PAGES: 95
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): 103
15. SECURITY CLASS. (of this report): Unclassified
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:
16. DISTRIBUTION STATEMENT (of this Report): Approved for public release; distribution unlimited.
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report):

DTIC ELECTE JUL 18 1980 S D A

18. SUPPLEMENTARY NOTES
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

perception, visual information processing, reading, pattern recognition, context effects, feedback, activation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report is the first part of a two-part series introducing an interactive activation model of context effects in perception. In this part, ~~we develop~~ is developed the model for the perception of letters in words and other contexts and ~~apply it~~ is applied to a number of experiments in the recent literature. The model is used to account for the perceptual advantage for letters in words compared to single letters and letters in unrelated strings. In the model, these word-superiority effects are produced by feedback. The visual input produces partial activations of letters, which in turn produce partial activations of words. These activations then pro-

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102 LF 014 6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

408267

20. (continued)

duce feedback to the letter level, reinforcing letter sequences which spell words. The model can account for the basic findings on the perception of pronounceable nonwords as well as words. The account is based on the idea that pseudowords can also activate representations of words, even though they do not match any word perfectly. As with word displays, feedback from the activated words reinforces the letters presented, thereby increasing their perceptibility. The model also accounts for the role of masking in determining the magnitude of the various effects, the fact that expectations influence perception of letters in words, and for the fact that effects of contextual constraint and letter cluster frequency are obtained under some conditions and not others.

An Interactive Activation Model

of the

Effect of Context in Perception

Part I

James L. McClelland & David E. Rumelhart

University of California, San Diego

Preparation of this paper was supported by NSF grant BNS-76-14830 to J. L. McClelland, by NSF grant BNS-76-15024 to D. E. Rumelhart, and by the Office of Naval Research under contract N00014-79-C-0323. We would like to thank Don Norman, James Johnston, and members of the LNR Perception research group for helpful discussions of much of the material covered in this paper.

Abstract

This paper is the first part of a two-part series introducing an interactive activation model of context effects in perception. In this part we develop the model for the perception of letters in words and other contexts and apply it to a number of experiments in the recent literature. The model is used to account for the perceptual advantage for letters in words compared to single letters and letters in unrelated strings. In the model, these word superiority effects are produced by feedback. The visual input produces partial activations of letters, which in turn produce partial activations of words. These activations then produce feedback to the letter level, reinforcing letter sequences which actually spell words. The model can account for the basic findings on the perception of pronounceable nonwords as well as words. The account is based on the idea that pseudowords can also activate representations of words, even though they do not match any word perfectly. As with word displays, feedback from the activated words reinforces the letters presented, thereby increasing their perceptibility. The model also accounts for the role of masking in determining the magnitude of the various effects, the fact that expectations influence perception of letters in pseudowords more than letters in words, and for the fact that effects of contextual constraint and letter cluster frequency are obtained under some conditions and not others.

As we perceive, we are continually extracting sensory information to guide our attempts to determine what is before us. In addition, we bring to perception a wealth of knowledge about the objects we might see or hear and the larger units in which these objects co-occur. As one of us has argued for the case of reading (Rumelhart, 1977) our knowledge of the objects we might be perceiving works together with the sensory information in the perceptual process. Exactly how does the knowledge which we have interact with the input? And, how does this interaction facilitate perception?

In this two-part article we have attempted to take a few steps toward answering these questions. We consider one specific example of the interaction between knowledge and perception -- the perception of letters in words and other contexts. In Part I we examine the main findings in the literature on perception of letters in context, and develop a model called the interactive activation model to account for these effects. In Part II (Rumelhart & McClelland, forthcoming) we extend the model in several ways. We present a set of studies introducing a new technique for studying the perception of letters in context, independently varying the duration and timing of the context and target letters. We show how the model fares in accounting for the results of these experiments and discuss how the model may be extended to an account of the pronunciation of nonwords. We also explore the influence of higher-level (semantic and syntactic) inputs to the perceptual process, not only for the case of visual word perception but for the perception of speech as well. Finally, we consider how the mechanisms developed in the course of exploring our model of perception might be used in other sorts of processes, such as categorization, memory search, and retrieval.

Basic Findings on the Role of Context in Perception of Letters

The notion that knowledge and familiarity play a role in perception has often been supported by experiments on the perception of letters in words or word-like letter strings (Bruner, 1957; Neisser, 1967). It has been known for nearly 100 years that it is possible to identify letters in words more accurately than letters in random letter sequences under tachistoscopic presentation conditions (Cattell, 1886; see Huey, 1908, and Neisser, 1967 for reviews). However, until recently such effects were obtained using whole reports of all of the letters presented. These reports are subject to guessing biases, so that it was possible to imagine that familiarity did not determine how much was seen but only how much could be inferred from a fragmentary percept. In addition, for longer stimuli, full reports are subject to forgetting. We may see more letters than we can actually report in the case of nonwords, but when the letters form a word we may be able to retain the item as a single unit whose spelling may simply be read out from long-term memory. Thus, despite strong arguments to the contrary by proponents of the view that familiar context really did influence perception, it has been possible until recently to imagine that the context in which a letter was presented only influenced the accuracy of post-perceptual processes, and not the process of perception itself.

The perceptual advantage of letters in words. The seminal experiment of Reicher (1969) seems to suggest that context does actually influence perceptual processing. Reicher presented target letters in words, unpronounceable nonwords, and alone, following the presentation of the target display with a presentation of a patterned mask. The subject was then tested on a single letter in the display, using a forced choice between two alternative letters. Both alternatives fit the context to form an item of the type presented, so that, for example, in the case of a word presentation, the alternative would also form a word in the context.

Forced choice performance was more accurate for letters in words than for letters in nonwords or even for single letters. Since both alternatives made a word with the context, it is not possible to argue that the effect is due to post-perceptual guessing based on equivalent information extracted about the target letter in the different conditions. It appears that subjects actually come away with more information relevant to a choice between the alternatives when the target letter is a part of a word. And, since one of the control conditions was a single letter, it is not reasonable to argue that the effect is due to forgetting letters that have been perceived. It is hard to see how a single letter, once perceived, could be subject to a greater forgetting than a letter in a word.

Reicher's finding seems to suggest that perception of a letter can be facilitated by presenting it in the context of a word. It appears, then, that our knowledge about words can influence the process of perception.

Our model presents a way of bringing such knowledge to bear. The basic idea is that the presentation of a string of letters results in partial activation of representations of letters consistent with the visual input. These activations in turn produce partial activations of representations of words consistent with the letters, if there are any. The activated representations of words then produce feedback which serves to reinforce the activations of the representations of letters. As a result, letters in words are more perceptible, because they receive more activation than representations of either single letters or letters in unrelated context.

Reicher's basic finding has been investigated and extended in a large number of studies, and there now appears to be a set of important related findings that must also be explained. Here follows a brief discussion of several further results which seem to be both basic and well established.

Irrelevance of word shape. The perceptual advantage for letters in words does not depend on presenting words in visually distinctive, or even familiar, forms. Typically, the effects are obtained using words typed in all upper case type, which minimizes configurational aspects of words as visual forms. In addition, the word advantage over nonwords can be obtained using stimuli presented in mixed upper and lower case type (Adams, 1979; McClelland, 1976). Although performance is affected by mixing upper and lower case letters in the same string, the disruption is of about the same magnitude for letters in nonwords as it is for letters in words, as long as both types of items are tested at comparable performance levels (Adams, 1979). It is therefore clear that the word advantage depends on presenting the target letter in the context of an item which together with the target forms a familiar arrangement of letters, independent of its actual visual form.

Dependence on masking. The word advantage over single letters and nonwords appears to depend upon the visual conditions used (Johnston & McClelland, 1973; Massaro & Klitzke, 1979; see also Juola, Leavitt & Choe, 1974; and Taylor & Chabot, 1978). The word advantage is quite large when the target appears in a distinct, high-contrast display followed by a patterned mask of similar characteristics. However, the word advantage over single letters is actually reversed, and the word advantage over nonwords becomes quite small when the target is indistinct, low in contrast and followed by a blank, nonpatterned field. Recently, it has also been shown that the word advantage over single letters is greatly reduced if the patterned mask contains letters instead of nonletter patterns (Johnston & McClelland, in press; Taylor & Chabot, 1978).

<u>Extension to pronounceable nonwords</u>. The word advantage also applies to pronounceable nonwords, such as REET or MAVE. A large number of studies (Aderman & Smith, 1971; Baron & Thurston, 1973; Carr, Davidson & Hawkins, 1978; Spoehr & Smith, 1975) have shown that letters in pronounceable nonwords (also called pseudowords) have a large advantage over letters in unpronounceable nonwords (also called unrelated letter strings), and three studies (Carr, et al, 1978; Massaro & Klitzke, 1979; McClelland & Johnston, 1977) have obtained an advantage for letters in pseudowords over single letters.

It now appears that the pseudoword advantage depends on the subjects' expectations (Aderman & Smith, 1971; Carr, et al, 1978). Carr, et al (1978) found that if subjects are under the impression that pseudowords might be shown, performance on pseudowords is almost as accurate as performance on letters in words. But if they do not expect any pseudowords, performance on these items is not much better than performance on unpronounceable nonwords. Interestingly, Carr, et al (1978) found that the word advantage did not depend on expectations. There was a sizable advantage for letters in words over letters in unrelated context whether the subject expected words or only unrelated letter strings.

Another important fact about performance on pseudowords is that differences in letter cluster frequency do not appear to influence accuracy of perception of letters in either words or pseudowords (McClelland & Johnston, 1977).

Absence of constraint effects. One important finding which rules out several of the models which have been proposed previously is the finding that letters in highly constraining word contexts have little or no advantage over letters in weakly constraining contexts under the distinct target/patterned mask conditions which produce a large word advantage (Johnston, 1978; see also Estes, 1975). For example, if the set of possible stimuli contains only words, the context _HIP constrains the first letter to be either an S, a C, or a W, whereas the context _INK is compatible with 12 to 14 letters (the exact number depends on what counts as a word). We might expect that the former, more strongly constraining context, would produce superior detection of a target letter, but, in a very carefully controlled and executed study, Johnston (1978) found a non-significant effect in the reverse direction. Although there are some findings suggesting that constraints do influence performance under other conditions, they do not appear to make a difference under the distinct target/patterned mask conditions of the Johnston study.

To be successful, any model of word perception must provide an account not only for Reicher's basic effect, but for the separate and joint effects (or lack thereof) due to visual conditions, stimulus structure, expectations, and constraints on the perception of letters in context. Our model provides an account for all of these effects. We begin by presenting the model in abstract form, then focus in on the details of the model, and present an example of the working of the model in a hypothetical experimental trial. Subsequently, we turn to a detailed consideration of the findings discussed in this section. In the final section of Part I, we also consider a few other facts about the perception of letters in context and suggest how our model might be extended to account for these effects as well.

The Interactive Activation Model

We approach the phenomena of word perception with a number of basic assumptions which we want to incorporate into the model. First, we assume that visual perception takes place within a system in which there are several levels of processing, each concerned with forming a representation of the input at a different level of abstraction. For visual word perception, we assume that there is a visual feature level, a letter level, and a word level, as well as higher levels of processing which provide "top-down" input to the word level.

Second, we assume that visual perception involves parallel processing. There are two different senses in which we view perception as parallel. We assume that visual perception is spatially parallel. That is, we assume that information covering a region in space at least large enough to contain a four-letter word is processed simultaneously. In addition, we assume that visual processing occurs at several levels at the same time. Thus, our model of word perception is spatially parallel, (i.e. capable of processing several letters of a word at one time) and involves processes which operate simultaneously at several different levels. Thus, for example, processing at the letter level presumably occurs simultaneously with processing at the word level, and with processing at the feature level.

Thirdly, we assume that perception is fundamentally an interactive process. That is, we assume that "top-down" or "conceptually driven" processing works simultaneously and in conjunction with "bottom-up" or "data driven" processing to provide a sort of multiplicity of constraints which jointly determine what we perceive. Thus, for example, we assume that knowledge about the words of the language interacts with the incoming featural information in co-determining the nature and time course of the perception of the letters in the word.

Finally, we wish to implement these assumptions using a relatively simple method of interaction between sources of knowledge whose only "currency" is simple "excitatory" and "inhibitory" activations of a neural type.

Figure 1 shows the general conception of the model. Perception is assumed to consist of a set of interacting levels, each level communicating with several others. Communication proceeds through a spreading activation mechanism in which activation at one level "spreads" to neighboring levels. The communication can consist of both excitatory and inhibitory messages. Excitatory messages increase the activation level of their recipients. Inhibitory messages decrease the activation level of their recipients. The arrows in the diagram represent excitatory connections and the circular ends of the connections represent inhibitory connections. The intra-level inhibitory loop represents a kind of lateral inhibition in which incompatible units at the same level compete. For example, since a string of, say, four letters can be interpreted as at most one four-letter word, the various possible words mutually inhibit one another and in that way compete as possible interpretations of the string.

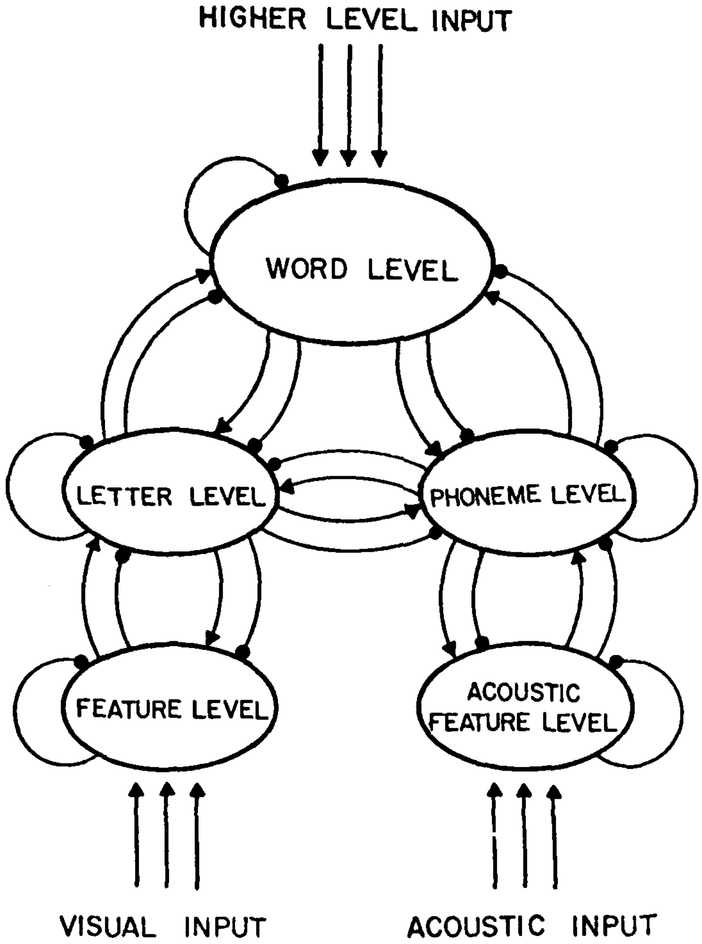

Figure 1. A sketch of some of the processing levels involved in visual and auditory word perception, with interconnections.

It is clear that there are many levels which are important in reading and perception in general and the interactions among these levels are important for many phenomena. However, a theoretical analysis of all of these interactions introduces an order of complexity which obscures comprehension. For this reason, we have restricted the present analysis to an examination of the interaction between a single pair of levels, the word and letter levels. We have found that we can account for the phenomena reviewed above by considering only the interactions between letter level and word level elements. Therefore, for the present we have elaborated the model only on these two levels, as illustrated in Figure 2. We have delayed consideration of the effects of higher-level processes and/or phonological processes, and we have ignored the reciprocity of activation which may occur between word and letter levels and any other levels of the system. We consider aspects of the fuller model including these influences in Part II.

Specific Assumptions

Representation assumptions. For every relevant unit in the system we assume there is an entity called a *node*. We assume that there is a node for each word we know, and that there is a node for each letter in each position.

The nodes are organized into levels. There are *word level* nodes, and *letter level* nodes. Each node has connections to a number of other nodes. The set of nodes to which a node connects are called its *neighbors*. Each connection is two way. There are two kinds of connections: *excitatory* and *inhibitory*. If the two nodes suggest each other's existence (in the way that the node for the word 'the' suggests the node for an initial 't' and vice versa) then the connections are excitatory. If the two nodes are inconsistent with

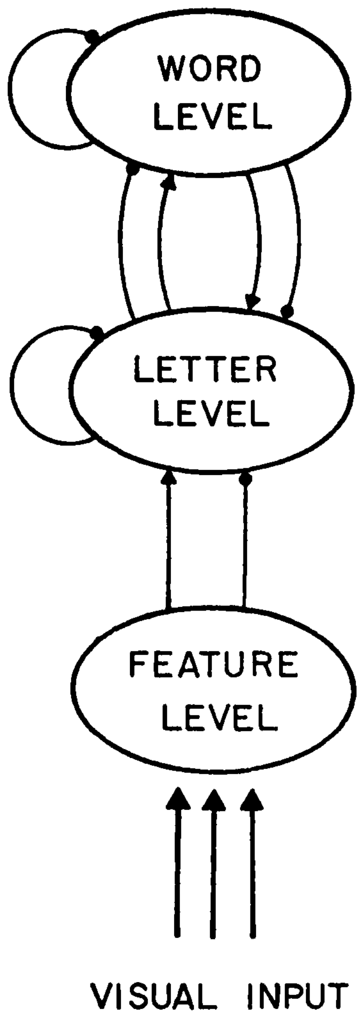

Figure 2. The simplified processing system considered in Part I.

one another (in the way that the node for the word 'the' and the node for the word 'boy' are inconsistent) then the relationship is inhibitory. (Note that we identify nodes by the units they detect, placing them in quotes: Stimuli presented to the system are typed in uppercase letters).

Connections may occur within levels or between adjacent levels. There are no connections between non-adjacent levels. Connections within the word level are mutually inhibitory since only one word can occur at any one place at any one time. Connections between the word level and letter level may be either inhibitory or excitatory (depending on whether or not the letter is a part of the word in the appropriate letter position). We call the set of nodes with excitatory connections to a given node its _excitatory neighbors_. We call the set of nodes with inhibitory connections to a given node its _inhibitory neighbors_.

A subset of the neighbors of the letter 't' are illustrated in Figure 3. Again, excitatory connections are represented by arrows ending with points and inhibitory connections are represented by arrows ending with dots. We emphasize that this is a small subset of the neighborhood of the initial 't'. The picture of the whole neighborhood, including all the connections among neighbors and their connections to their neighbors, is much too complicated to present in a two-dimensional figure.

Activation assumptions. There is, associated with each node, a momentary level of activation. This level of activation is a real number, and for node i we will represent it by $a_i(t)$. Any node with a positive degree of activation is said to be _active_. In the absence of inputs from its neighbors, all nodes are assumed to decay back to an inactive state; that is, to an

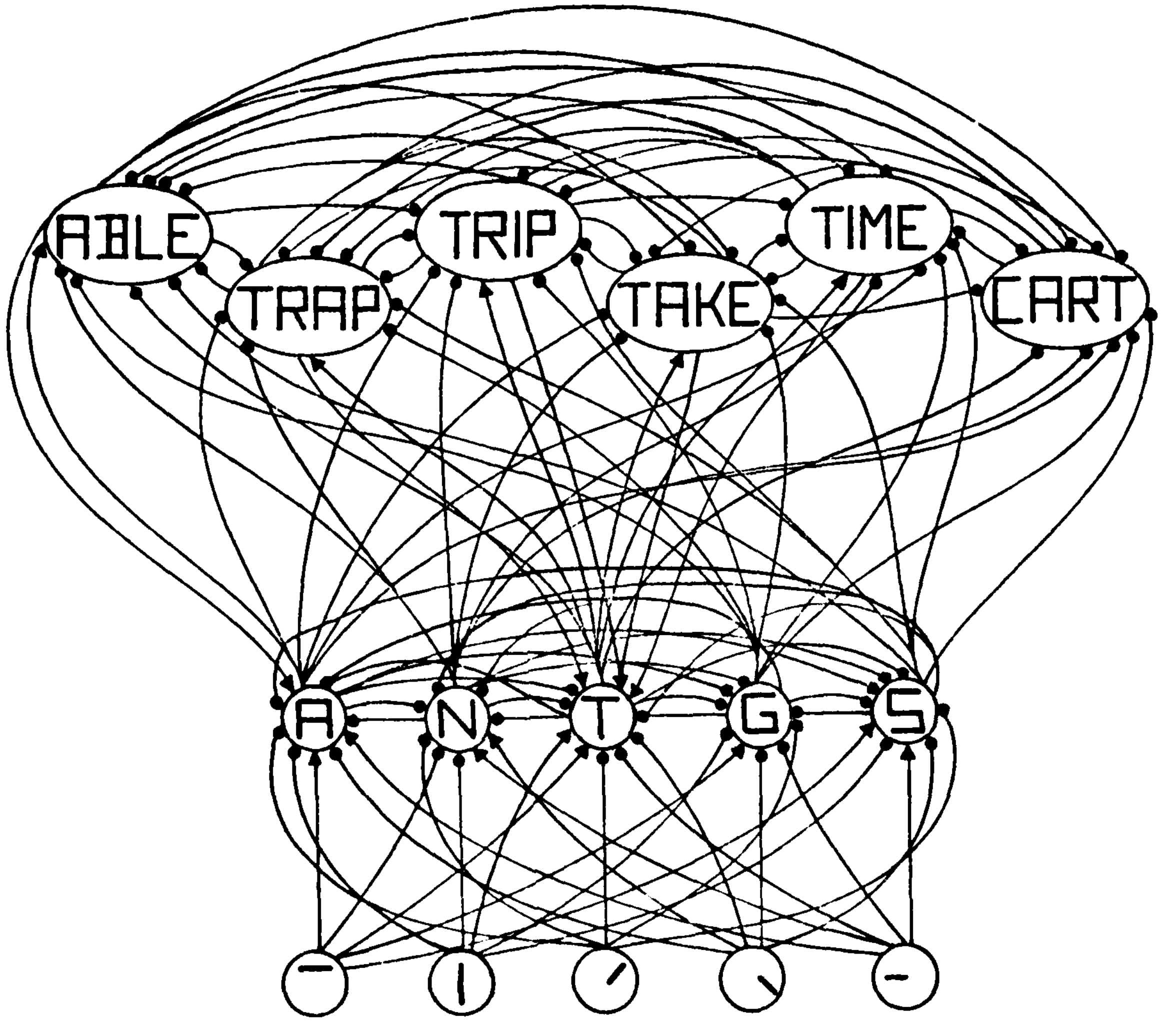

Figure 3. A few of the neighbors of the node for the letter 't' in the first position in a word, and their interconnections.

activation value at or below zero. This resting level may differ from node to node, and corresponds to a kind of a priori bias (Broadbent, 1967), determined by frequency of activation of the node over the long term. Thus, for example, the nodes for high frequency words have resting levels higher than those for low frequency words. In any case, the resting level for node i is represented by r_i. For units not at rest, decay back to the resting level occurs at some rate Θ_i.

When the neighbors of a node are active they influence the activation of the node by either excitation or inhibition, depending on their relation to the node. These excitatory and inhibitory influences combine by a simple weighted average to yield a net input to the unit, which may be either excitatory (greater than zero) or inhibitory. In mathematical notation, if we let $n_i(t)$ represent the net input to the unit, we can write the equation for its value as

$$n_i(t) = \sum_j \alpha_{ij} e_j(t) - \sum_k \gamma_{ik} i_k(t), \qquad (1)$$

where the $e_j(t)$s are the activations of the active excitatory neighbors of the node, the $i_k(t)$s are the activations of the active inhibitory neighbors of the node, and the α_{ij}s and γ_{ik}s are associated weight constants. Inactive nodes have no influence on their neighbors. Only nodes in an active state have any effects, either excitatory or inhibitory.

The net input to a node drives the activation of the node up or down depending on whether it is positive or negative. The degree of the effect of the input on the node is modulated by the node's current activity level, to keep the input to the node from driving it beyond some maximum and minimum values (Grossberg, 1978). When the net input is excitatory ($n_i(t)>0$), the effect on the node is given by

$$\varepsilon_i(t) = n_i(t)(M - a_i(t)) \; . \tag{2}$$

where M is the maximum activation level of the unit. The modulation has the desired effect because as the activation of the unit approaches the maximum, the effect of the input is reduced to zero.

In the case where the input is inhibitory ($n_i(t)<0$), the effect of the input on the node is given by

$$\varepsilon_i(t) = n_i(t)(a_i(t) - m) \; , \tag{3}$$

where m is the minimum activation of the unit.

The new value of the activation of a node at time $t+\delta t$ is equal to the value at time t, minus the decay, plus the influence of its neighbors at time t:

$$a_i(t+\delta t) = a_i(t) - \Theta_i(a_i(t) - r_i) + \varepsilon_i(t). \tag{4}$$

<u>Input assumptions</u>. Upon presentation of a stimulus a set of featural inputs are assumed to be made available to the system. During each moment in time each feature has some probability <u>p</u> of being detected. Upon being detected, the feature begins sending activation to all letter level nodes

which contain that feature. All letter level nodes which do not contain the extracted feature are inhibited. The probability of detection and the rate at which the feature excites or inhibits the relevant letter nodes are assumed to depend on the clarity of the visual display. It is assumed that features are binary and that we can extract either the presence or absence of a particular feature. So, for example, when viewing the letter R we can extract among other features the presence of a diagonal line segment in the lower right corner and the absence of a horizontal line across the bottom.

Presentation of a new display following an old one results in the probabilistic extraction of the set of features present in the new display. These features, when extracted, replace the old ones in corresponding positions. Thus, the presentation of an O following the R described above would result in the replacement of the two features described above with their opposites.

The Operation of the Model

Now, consider what happens when an input reaches the system. Assume that at time t_0 all prior inputs have had an opportunity to decay, so that the entire system is in its quiescent state and each node is at its resting level. The presentation of a stimulus initiates a chain in which certain features are extracted and excitatory and inhibitory pressures begin to act upon the letter level nodes. The activation levels of certain letter nodes are pushed above their resting levels. Others receive predominately inhibitory inputs and are pushed below their resting levels. These letter nodes, in turn, begin to send activation to those word level nodes they are consistent with and inhibit those word nodes they are not consistent with. In addition, the various letter level nodes attempt to suppress each other with the strongest ones getting the upper hand. As word level nodes become active they in turn compete with one another and send excitation and inhibition back down to the letter level nodes. If the input features were close to those for one particular set of letters and those letters were consistent with those forming a particular word, the positive feedback in the system will work to rapidly converge on the appropriate set of letters, and the appropriate word. If not, they will compete with each other and perhaps no single set of letters or single word will get enough activation to dominate the others and their inhibitory relationships might strangle each other. The exact details of this process depend on the values of the various parameters of the model in ways which we will explore as we proceed.

Simulations

In the following example, as in the remainder of the paper, we illustrate the properties of the model with computer simulations. For purposes of these simulations we have made a number of other simplifying assumptions. These additional assumptions fall into four classes:

(1) discrete rather than continuous time,
(2) simplified feature analysis of the input font,
(3) restrictions of the parameter space, and
(4) a limited lexicon.

The simulation of the model operates in discrete time slices or ticks, updating the activations of all of the nodes in the system once each cycle on the basis of the values on the previous cycle. Obviously, this is simply a matter of computational convenience, and not a fundamental assumption. We have endeavored to keep the time slices "thin" enough so that the model's behavior is continuous for all intents and purposes.

Any simulation of the model involves making explicit assumptions about the appropriate featural analysis of the input font. We have, for simplicity, chosen the font and featural analysis employed by Rumelhart (1971) and by Rumelhart and Siple (1974) and illustrated in Figure 4. Although the experiments we have simulated employed different type fonts, presumably the basic results do not depend on the particular font used. The simplicity of the present analysis recommends it for the simulations.

We have endeavored to find a single set of parameter values for our model which would allow us to account for all of the basic findings reviewed above. In order to keep the search space to an absolute minimum, we have adopted various restrictive simplifications. We have assumed that the weight parameters, α_{ij} and γ_{ij} depend only on the levels of nodes i and j and on no other characteristics of their identity. This means, among other things, that the excitatory connections between all letter nodes and all of the relevant word nodes are equally strong, independent of the identity of the words. Thus, for example, the degree to which the node for an initial 't' excites the node for the word 'tock' is exactly the same as the degree to which it excites the node for a word like 'this,' in spite of a substantial difference in frequency of usage. To further simplify matters, two types of influences have been set to zero, namely the word to letter inhibition and the letter to letter inhibition. We have also assigned the same resting value to all of the letter nodes, simply giving each node the value of zero. The resting value of nodes at the word level has been set to a value between -.05 and 0, depending on

Figure 4. The features used to construct the letters in the font assumed by the simulation program, and the letters themselves (from Rumelhart & Siple, 1974).

word frequency. The values of the remaining parameters have been fixed at the values given in Table 1. In the simulations which follow, all parameters are fixed at the values indicated in the table. The table also includes a brief statement of the significance or rationale for the particular value assigned. In some cases, fuller discussions are warranted, and are given in the context of a discussion of the model's behavior in accounting for one effect or another.

In order to account for the dependence of the phenomena of letter perception on visual conditions and expectations, it is necessary to assume that some parameters depend on these factors. The quality of the visual display is assumed to influence the system in two ways. First of all, it may not be possible for the visual system to extract all the features of the display if it becomes too degraded. To capture this possibility, we allow the probability of feature extraction to vary with the quality of the display. Once the quality is sufficiently good for perfect feature extraction, the strength of the effect exerted by the features is assumed to depend on such things as the brightness, contrast, size, and retinal position of the display. The parameters which reflect the differential strength of the effect of the input are the feature to letter excitation parameters. It is assumed that these parameters increase and decrease together as visual quality increases or decreases, but stay in the same ratio. To accommodate the fact that performance depends in some conditions on the subjects' expectations, we have found it sufficient to assume that one of the internal parameters of the model is under subject control. As we shall see below, we are able to provide a straightforward account of the effects of expectations about whether pronounceable nonwords will be shown if we assume that subjects have control over the strength of the

Table 1

Parameter Values Used in the Simulations

Parameter	Value	Remarks
Basic node characteristics		
decay rate	.07	Scales time. Low value ensures adequacy of approximation of continuity.
maximum activation	1.00	Scales activations.
minimum activation	.20	Small negative value allows rapid re-activation of inhibited units.
Resting levels		
letter level	0	Simplifying assumption.
word level	<0	Depends on frequency. (range: 0 to -.05)
Input		
p of feat detection	var.	Depends on visual conditions.
feat-let excitation	var.	Depends on visual conditions.
feat-let inhibition	var.	Inhibition much stronger than excitation so
E/I ratio	1/30	that one feature incompatible with a letter results in net bottom-up inhibition.
Letter-word influences		
excitation	.07	
inhibition	.04	Low value allows letter level to excite words
	or	with some letters incompatible with input.
	.21	High value prohibits these activations.
Within-level inhibition		
word level	.21	Large inhibitory interactions allow correct word to dominate total activity at word level.
letter level	0	Simplifying assumption. Unnecessary because of strong inhibition from inappropriate features.
Word-letter feedback		
excitation	.30	
inhibition	0	Simplifying assumption.
Output		
integration rate	.05	Low rate lets units be quickly activated then inhibited without becoming accessible.
Output Exponentiation		
letter level	10	Scales relation of activation to p(correct).
word level	20	Larger value required to offset greater number of alternatives.

letter to word inhibition parameter. We will see why this is so below. In any case, the parameters which are assumed to be influenced by visual conditions or expectations are designated as variable in Table 1. As we go along we will explore the effects of variations in these parameters on the performance of the model.

Finally, our simulations have been restricted to four-letter words. We have equipped our simulation program with knowledge of 1179 four-letter words occurring at least 2 times per million in the Kucera and Francis word count (1967). Plurals, inflected forms, first names, proper names, acronyms, abbreviations, and occasional unfamiliar entries arising from apparent sampling flukes have been excluded. This sample appears to be sufficient to reflect the essential characteristics of the language and to show how the statistical properties of the language can affect the process of perceiving letters in words.

An example. For the purposes of this example, imagine that the word WORK has been presented to the subject and that the subject has extracted those features shown in Figure 5. In the first three-letter positions the features of the letters W, O and R have been completely extracted. In the final position a set of features consistent with the letters K and R have been extracted, with those features in a portion of the pattern unavailable. We wish now to chart the activity of the system resulting from this presentation. Figure 6 shows the time course of the activations for selected nodes at the word and letter levels respectively.

At the word level, we have charted the activity levels of the nodes for the words 'work', 'word', 'wear' and 'weak'. Note first, that 'work' is the

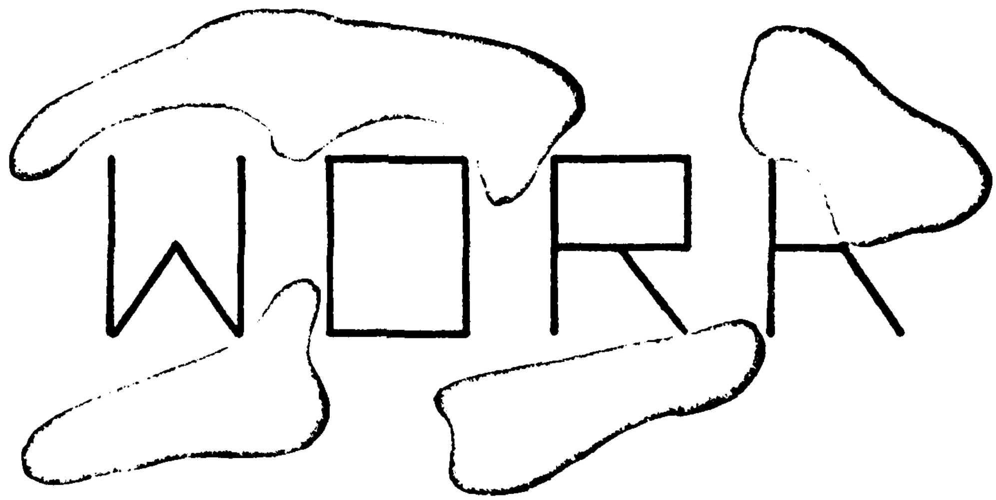

Figure 5. A hypothetical set of features which might be extracted on a trial in an experiment on word perception.

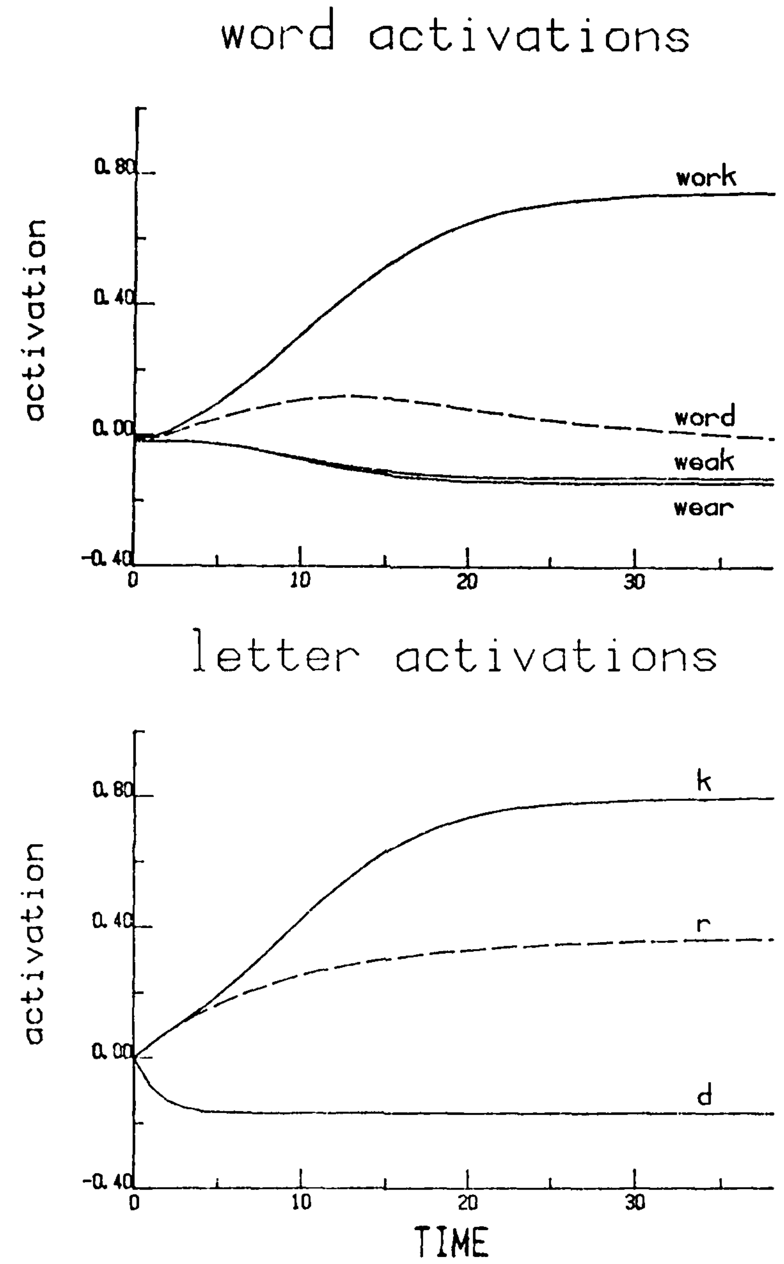

Figure 6. The time course of activations of selected nodes at the word and letter levels, after extraction of the features shown in Figure 5.

only word in the lexicon consistent with all the presented information. As a result, its activation level is the highest and reaches a value of .8 through the first 40 time cycles. The word 'word' is consistent with the bulk of the information presented and, as a result, first rises and later, as a result of competition with 'work' is pushed back down below its resting level. The words 'wear' and 'weak' are consistent with the information presented in the first and fourth letter positions, but inconsistent with the information in letter positions 2 and 3. Thus, the activations of these nodes drop to a rather low level. This level is not quite as low of course as the activation level of words such as 'gill' which contain nothing in common with the presented information. Although not shown in the figure these words attain near-minimum activation levels of about -.20 and stay there as the stimulus stays on. Returning to 'wear' and 'weak', we note that these words are equally consistent with the presented information and thus drop together for the first 9 or so time units. At this point, however, top-down information has determined that the final letter is K and not R. As a result, the word 'weak' becomes more similar to the pattern at the letter level than the word 'wear' and, as a result, begins to gain a slight advantage over 'wear.' This result occurs in the model because as the word 'work' gains in activation it feeds activation back down to the letter level to strengthen the 'k' over the 'r'. The strengthened 'k' continues to feed activation into the word level and strengthen consistent words. The words containing 'r' continue to receive activation from the words consistent with 'k', and are therefore ultimately weakened, as illustrated in the lower panel of the Figure.

One of the characteristics of the parameter set we have adopted is that feature to letter inhibition is 30 times stronger than feature to letter excitation (see Table 1). This ratio ensures that as soon as a feature is detected which is inconsistent with a particular letter, that letter receives relatively strong net bottom-up inhibition. Thus, in our example, the information extracted clearly disconfirms the possibility that the letter D has been presented in the fourth position, and thus the activation level of the 'd' node decreases quickly to near its minimum value. However, the bottom-up information from the feature level supports both 'k' and 'r' in the fourth position. Thus, the activation level for each of these nodes rises slowly. These activation levels, along with those for 'w', 'o' and 'r' push the activation level of 'work' above zero and it begins to feed back, and by about time cycle 4 it is beginning to push the 'k' above the 'r' (WORR is not a word). Note that this separation occurrs just before the words 'weak' and 'wear' separate. It is this feedback that causes them to separate. Ultimately, the 'r' reaches a level well below that of 'k' where it remains, and the 'k' pushes toward a .8 activation level. Remember that for purposes of simplicity the word to letter inhibition and the intra-letter level inhibition have both been set to 0. Thus, 'k' and 'r' both co-exist at moderately high levels, the 'r' fed only from the bottom-up and the 'k' fed from both bottom-up and top-down.

Although this example is not too realistic in that we assumed that only partial information was available in the input for the fourth letter position, whereas full information is available at the other letter positions, it does illustrate many of the important characteristics of the model. It shows how ambiguous sensory information can be disambiguated by top-down processes. Here we have a very simple mechanism capable of applying knowledge of words in the perception of their component letters.

On Making Responses

One of the more problematic aspects of a model such as this one is a specification of how these relatively complex patterns of activity might be related to the content of percepts and the sorts of response probabilities we observe in experiments. We assume that responses and perhaps the contents of perceptual experience depend on the temporal integration of the pattern of activation over all of the nodes. The integration process is assumed to occur slowly enough that brief activations may come and go, without necessarily becoming accessible for purposes of responding or entering perceptual experience. However, as the activation lasts longer and longer, the probability that it will be reportable increases. Specifically, we think of the integration process as taking a running average of the activation of the node averaged over the immediately preceding time interval:

$$\bar{a}_i(t) = \int_0^t a_i(t)e^{-(t-x)r}dt. \tag{5}$$

The parameter r represents the relative weighting given to old and new information. Larger values of r correspond to larger weight for new information. Response strength in the sense of Luce's choice model (Luce, 1959), is an exponential function of the running average activation:

$$s_i(t) = e^{\mu \bar{a}_i(t)} \tag{6}$$

The parameter μ determines how rapidly response strength grows with increases in activation. Following Luce's formulation, we assume that the probability of making a response based on node i is given by

$$p(R_i,t) = \frac{s_i(t)}{\sum_{j \in L} s_j(t)} \tag{7}$$

where L represents the set of nodes competing at the same level with node i.

Most of the experiments we will be considering test subject's performance on one of the letters in a word, or on one of the letters in some other type of display. In accounting for these results, we have adopted the assumption that responding is always based on the output of the letter level, rather than the output of the word level or some combination of the two. Thus, with regard to the previous example, it is useful to look at the "output values" for the letter nodes 'r', 'k' and 'd'. Figure 7 shows the output values for these simulations. The output value is the probability that, if a response was initiated at time t, the letter in question would be selected as the output or response from the system. As intended, these output values grow somewhat more slowly than the values of the letter activations themselves, but eventually come to reflect the activations of the letter nodes, as they reach and hold their asymptotic values.

Comments on Related Formulations

Before turning to the applications of the model, some comments on the relationship of this model to other models extant in the literature is in order. We have tried to be synthetic. We have taken ideas from our own previous work and from the work of others in the literature. In what follows, we

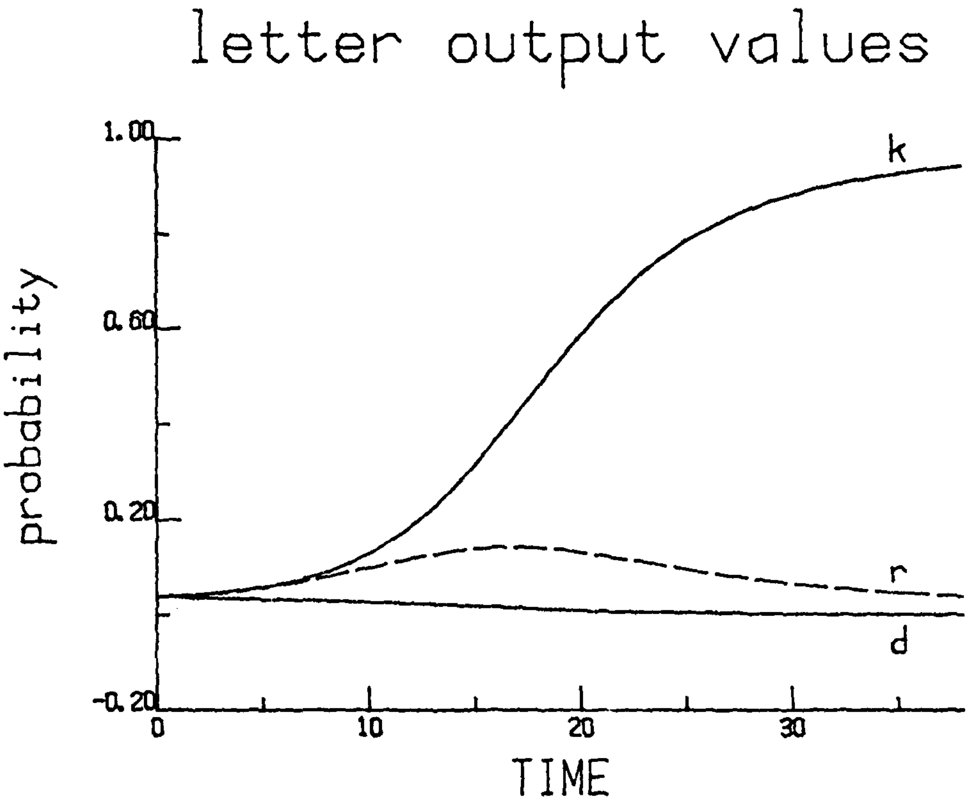

Figure 7. "Output values" for the letters 'r', 'k', and 'd', after presentation of the display shown in Figure 5.

have attempted to identify the sources of most of the assumptions of the model and to show in what ways our model differs from the models we have drawn on.

First of all, we have adopted the approach of formulating the model in terms which are similar to the way in which such a process might actually be carried out in a neural or neural-like system. We do not mean to imply that the nodes in our system are necessarily related to the behavior of individual neurons. We will, however, argue that we have kept the kinds of processing involved well within the bounds of capability for simple neural circuits. The approach of modeling information processing in a neural-like system has recently been advocated by Szentagothai and Arbib (1975), and is embodied in many of the papers presented in the forthcoming volume by Hinton and Anderson (in press) as well as many of the specific models mentioned below.

One case in point is the work of Levin and Eisenstadt (1975) and Levin (1976). They have proposed a parallel computational system capable of interactive processing which employed only excitation and inhibition as its "currency." Although our model could not be implemented exactly in the format of their system (called Proteus) it is clearly in the spirit of their model and could readily be implemented within a variant of the Proteus system.

In a recent paper McClelland (1979) has proposed a cascade model of perceptual processing in which activations on each level of the system drive those at the next higher level of the system. This model has the properties that partial outputs are continuously available for processing and that every level of the system processes the input simultaneously. The present model certainly embodies these assumptions. It also generalizes them, permitting information to flow in both directions simultaneously.

Hinton (1977) has developed a relaxation model for visual perception in which multiple constraints interact by means of incrementing and decrementing real numbered values associated with various interpretations of a portion of the visual scene in an attempt to attain a maximally consistent interpretation of the scene. Our model can be considered a sort of relaxation system in which activation levels are manipulated to get an optimal interpretation of an input word.

James Anderson and his colleagues (Anderson, 1977; Anderson, Silverstein, Ritz, & Jones, 1977) and Kohonen and his colleagues (Kohonen, 1977) have developed a sort of pattern recognition system which they call an associative memory system. Their system shares a number of commonalities with ours. One thing the models share is the scheme of adding and subtracting weighted excitation values to generate output patterns which represent cleaned up versions of the input patterns. In particular, our α_{ij} and γ_{ij} correspond to the matrix elements of the associative memory models. Our model differs in that it has multiple levels and employs a non-linear cumulation function similar to one suggested by Grossberg (1978), as mentioned above.

Our model also draws on earlier work in the area of word perception. There is, of course, a strong similarity between this model and the logogen model of Morton (1969). What we have implemented might be called a hierarchical, non-linear, logogen model with feedback between levels and inhibitory interactions among logogens at the same level. We have also added dynamic assumptions which are lacking from the logogen model.

The notion that word perception takes place in a hierarchical information processing system has, of course, been advocated by several researchers interested in word perception (Adams, 1979; Estes, 1975; LaBerge & Samuels, 1974; Johnston & McClelland, in press; McClelland, 1976). Our model differs from those proposed in many of these papers in that processing at different levels is explicitly assumed to take place in parallel. Many of the models are not terribly explicit on this topic, although the notion that partial information could be passed along from one level to the next so that processing could go on at the higher level while it was continuing at the lower level had been suggested by McClelland (1976). Our model also differs from all of these others, except that of Adams (1979), in assuming that there is feedback from the word level to the letter level. The general formulation suggested by Adams (1979) is quite similar to our own, although she postulates a different sort of mechanism for handling pseudowords (excitatory connections among letter nodes) and does not present a detailed model.

Our mechanism for accounting for the perceptual facilitation of pseudowords involves, as we will see below, the integration of feedback from partial activation of a number of different words. The idea that pseudoword perception could be accounted for in this way is similar to the assumptions of Glushko (1979), who suggested that partial activation and synthesis of word pronunciations could account for the process of constructing a pronunciation for a novel pseudoword.

The feature extraction assumptions and the bottom-up portion of the word recognition model are nearly the same as those employed by Rumelhart (1970, 1971) and Rumelhart and Siple (1974). The interactive feedback portion of the model is clearly one of the class of models discussed by Rumelhart (1977) and could be considered a simplified control structure for expressing the model proposed in that paper.

The Word Advantage, and the Effects of Visual Conditions

As we noted previously, word perception has been studied under a variety of different visual conditions, and it is apparent that different conditions produce different results. The advantage of words over nonwords appears to be largest under conditions in which a bright, high-contrast target is followed by a patterned mask with similar characteristics. The word advantage appears to be considerably smaller when the target presentation is dimmer or otherwise degraded and is followed by a blank white field.

Typical data demonstrating these points (from Johnston & McClelland, 1973) is presented in Table 2. Forced-choice performance on letters in words is compared to performance on letters imbedded in a row of #'s (e.g., READ vs #E##). The #'s serve as a control for lateral facilitation and/or inhibition. (The latter factor appears to be important under dim target/blank mask conditions).

Target durations were adjusted separately for each condition so that it is only the pattern of differences within display conditions which is meaningful. What the data show is that a 15% word advantage was obtained in the bright target/patterned mask condition, and only a 5% word advantage in the dim target/blank mask condition. Massaro and Klitzke (1979) obtained about the same size effects. Various aspects of these results have also been corroborated in two other studies (Juola, Leavitt & Choe, 1974; Taylor & Chabot,

Table 2

Effect of Display Conditions on

Probability Correct Forced Choices in

Word & Letter Perception, from Johnston & McClelland, 1973

	Display Type	
Visual Conditions	Word	Letter with #'s
Bright Target/Patterned Mask	.80	.65
Dim Target/Blank Mask	.78	.73

1978).

To understand the difference between these two conditions it is important to note that in order to get about 75 percent performance in the no-mask condition, the stimulus must be highly degraded. Since there is no patterned mask, the iconic trace presumably persists considerably beyond the offset of the presentation. The effect of the blank mask is simply to reduce the contrast of the icon by summating with it. Thus, the limit on performance is not so much the amount of time available in which to process the information as it is the quality of the information made available to the system. In contrast, when a patterned mask is employed, the mask interrupts the iconic trace and produces spurious inputs which can serve to disrupt the processing. Thus, in the bright target/pattern mask conditions, the primary limitation on performance is the time in which the information is available to the system rather than the quality of the information presented. This distinction between the way in which blank masks and patterned masks interfere with performance has previously been made by a number of investigators, including Rumelhart (1970) and Turvey (1973). We now turn to consider each of these sorts of conditions in turn.

Word Perception Under Conditions of Degraded Input

In conditions of degraded (but not abbreviated) input, the role of the word level is to selectively reinforce possible letters consistent with the visual information extracted which are also consistent with the words in the subject's vocabulary. Recall that the task requires the subject to choose between two letters which (on word trials) both make a word with the rest of the context. There are two distinct cases to consider. Either the featural information extracted about the to-be-probed letter is sufficient to distinguish between the alternatives, or it is not. Whenever the featural information is consistent with both of the forced-choice alternatives, any feedback will selectively enhance both alternatives, but will not permit the subject to improve his ability to distinguish between them. When the information extracted is inconsistent with one of the alternatives, there is nothing for the model to do if we assume that the subject can actually use the extracted feature information directly when it comes time to make the forced choice. However, the subject may not have direct access to this information. If we assume that forced-choice responses are based not on the feature information itself but on the subject's best guess about what letter was actually shown, then the model can produce a word advantage. The reason is that feedback from the word level will increase the probability of correct choice in those cases where the subject extracts information inconsistent with the incorrect alternative, but consistent with a number of other letters. Thus, feedback would have the effect of helping the subject select the actual letter shown from several possibilities consistent with the set of extracted features. Consider again, for example, the case of the presentation of WORD discussed above. In this case, the subject extracted incomplete information about the final letter consistent with both R and K. Assume that the forced choice the subject was to face on this trial was between a D and a K. The account supposes that the subject encodes a single letter for each letter position before facing the forced choice. Thus, if the features of the final letter had been extracted in the absence of any context, the subject would encode R or K equally often since both are equally compatible with the features extracted. This would leave him with the correct response some of the time. But if he chose R instead, he would enter the forced choice between D and K without knowing the correct answer directly. When the whole word display is shown, the feedback generated by the processing of all of the letters greatly strengthens the K, increasing the probability that it will be chosen over the R, and thus increasing the probability that the subject will proceed to the forced choice with the correct response in mind.

Our interpretation of the small word advantage in blank mask conditions is a specific version of the early accounts of the word advantage offered by Wheeler (1970) and Thompson & Massaro (1973), before it was known that the effect depends on masking. Johnston (1978) has argued that this type of account does not apply under patterned mask conditions. We are suggesting that it does apply to the small word advantage obtained under blank mask conditions like those of the Johnston and McClelland (1973) experiment. We will see below that the model offers a different account of performance under patterned mask conditions.

We simulated this interpretation of the small word advantage obtained in blank mask conditions in the following way. A set of 40 pairs of four-letter words differing by a single letter was prepared. From these words corresponding control pairs were generated in which the critical letters from the word pairs were presented in non-letter contexts (#'s). Because they are presented in non-letter contexts, we assume that these letters do not engage the word processing system at all. In fact we have run some simulations allowing such stimuli to interact with word-level knowledge and it makes little difference to the overall results.

Each member of each pair of items was presented to the model 4 times, yielding a total of 320 stimulus presentations of word stimuli and 320 presentations of single letters. On each presentation, the simulation sampled a random subset of the possible features to be detected by the system. The probability of detection of each feature was set at .45. The values of the feature to letter excitation and inhibition parameters were set at .005 and .15 respectively. As noted previously, these values are in a ratio of 1 to 30, so that if any one of the fourteen features extracted is inconsistent with a particular letter, that letter receives net inhibition from the features, and is rapidly driven into an inactive state.

For simplicity, the features were treated as a constant input which remained on while letter and word activations (if any) were allowed to take place. At the end of 50 processing cycles, output was sampled. Sampling results in the selection of one letter to fill each position; the selected letter is assumed to be the only thing the subject takes away from the target display.

The forced choice is assumed to occur as follows. The subject compares the letter selected for the appropriate position against the forced-choice alternatives. If the letter selected is one of the alternatives, then that alternative is selected. If it is not one of the alternatives, then one of the two alternatives is simply picked at random.

The simulation was run twice, once using the low value of letter to word inhibition listed in Table 1 and once using the high value. The results were different in the two cases. When the small letter to word inhibition value was used the letters embedded in words were 78% correct, whereas those in #'s were 68% correct -- a 10% difference. When the larger value of letter to word inhibition was used, the two conditions showed no difference. The reason for this difference is as follows. Under conditions in which incomplete feature information is extracted from the display, multiple letters become active in each position. When the letter to word inhibition is strong, these activations keep any word from becoming activated. For example, suppose that 'e', 'o', 'c' and 'q' were all partially activated in the second position after presentation of the word READ. Then the activations of 'o', 'c', and 'q' would inhibit the node for 'read', the activations of 'e', 'c' and 'q' would inhibit the node for 'road', etc. Other partial activations in other positions would have similar effects. Thus, few words ever receive net excitatory input, no feedback is generated, and little advantage of words over letters emerges. When the letter to word inhibition is weak, on the other hand, words which are consistent with one of the active letters in each position can become active, thereby allowing for facilitation by feedback. If, as we have assumed, the letter to word inhibition parameter is under the subject's control, then this would be a situation in which it would be advantageous for subjects to use a small value of this parameter. Thus, we would assume that under conditions of degraded input subjects would be inclined to adopt a low value of letter to word inhibition, with the effect that partial activation of multiple possible letters in each position would permit the activation of a set of possible words.

Apparently, the low value of letter to word inhibition produced a larger effect in the simulation than is observed in experiments. However, there are, as Johnston (1978) has pointed out, a number of reasons why an account such as the one we have offered would overestimate the size of the word advantage.

For one thing, subjects may occasionally be able to retain an impression of the actual visual information they have been able to extract. On such occasions, feedback from the word level will be of no further benefit. Second, even if subjects only retain a letter identity code, they may tend to choose a forced-choice alternative which is most similar to the letter encoded, instead of simply guessing when the letter encoded is not one of the two choice alternatives. Since the letter encoded will tend to be similar to the letter shown, this would tend to result in a greater probability correct and less of a chance for feedback to increase accuracy of performance. It is hard to know exactly how much these factors should be expected to reduce the size of the word advantage under these conditions, but they should reduce it some, bringing our simulation closely in line with the results.

Word Perception Under Patterned Mask Conditions

When a high quality display is followed by a patterned mask, we assume that the bottleneck in performance does not come in the extraction of feature information from the target display. Thus, in our simulation of these conditions, we assume that all of the features presented can be extracted on every trial. The limitation on performance comes from the fact that the activations produced by the target are subject to disruption and replacement by the mask before they can be translated into a permanent form suitable for overt report. This general idea was suggested by Johnston and McClelland (1973), and considered by a variety of other investigators, including Carr, et al (1978), Massaro and Klitzke (1979) and others. On the basis of this idea, a number of possible reasons for the advantage for letters in words have been suggested. One is that letters in words are for some reason translated more quickly into a non-maskable form (Johnston & McClelland, 1973; Massaro & Klitzke, 1979). Another is that words activate representations removed from the direct effects of visual patterned masking (Johnston & McClelland, 1973, in press; Carr et al, 1978; McClelland, 1976). In the interactive activation model, the reason letters in words fare better than letters in nonwords is that they benefit from feedback which can either drive then to higher activation levels or which can keep them active longer in the face of inhibitory influences of masking, or both. In either case, the probability that the activated letter representations will be correctly encoded is increased.

To understand how this account works in detail, consider the following example. Figure 8 shows the operation of our model for the letter E both in an unrelated letter context and in the context of the word READ for a visual display of moderately high quality. We assume that display conditions are sufficient *for complete feature extraction*, so that only the letters actually contained in the target receive net excitatory input on the basis of feature information. After some number of cycles have gone by, the mask is presented with the same parameters as the target. The mask simply replaces the target display at the feature level, resulting in a completely new input to the letter level. This input, because it contains features incompatible with the letter shown in all four positions, immediately begins to drive down the activations at the letter level. After only a few more cycles, these activations drop below resting level in both cases. Note that the correct letter was activated briefly, and no competing letter was activated. However, because of the sluggishness of the output process, these activations do not necessarily result in a high probability of correct report. As shown in the right half of the figure, the probability of correct report reaches a maximum

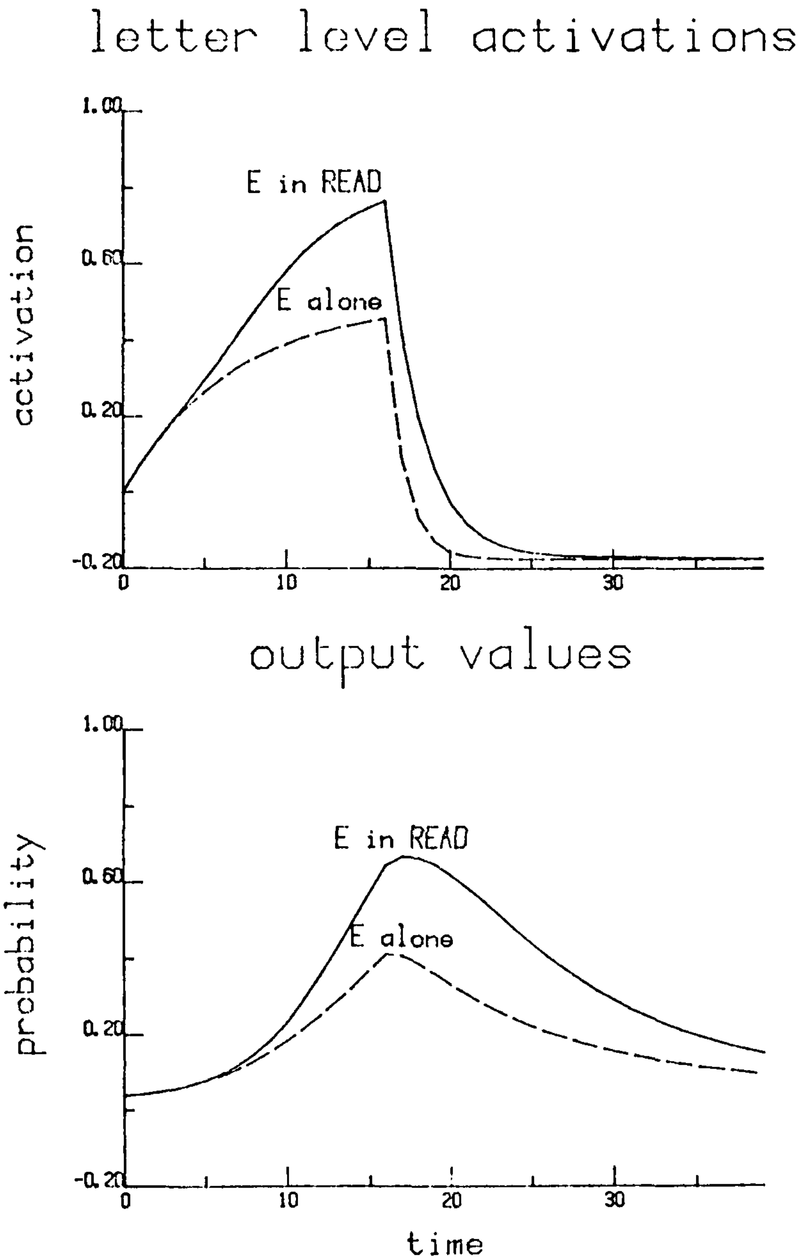

Figure 8. Activation functions (top) and output values (bottom) for the letter E, in unrelated context and in the context of the word READ.

after 16 cycles at a performance level far below the ceiling.

When the letter is part of a word (in this case, READ), the activation of the letters results in rapid activation of one or more words. These words, in turn, feed back to the letter level. This results in a higher net activation level for the letter embedded in the word. Moreover, since the letter embedded in a word has feedback from the word level to help sustain its activation, it is less readily displaced by the mask. This effect is not visible in the Figure. However, as the input strength is increased and the activations begin to level off, the difference between these two functions is increasingly in persistence and not in height of the activation curve.

We have carried out several simulations of the word advantage using the same stimulus list used for simulating the blank mask results. Since the internal workings of the model are completely deterministic as long as probability of feature extraction is 1.0, it was only necessary to run each item through the model once to obtain the expected probability that the critical letter would be encoded correctly for each item, under each variation of parameters tried.

One somewhat problematical issue involves deciding when to read out the results of processing and select candidate letters for each letter position. For simplicity, we have assumed that this occurs in parallel for all four letter positions and that the subject learns through practice to choose a time to read out in order to optimize performance. We have assumed that readout time may be set at a different point in different conditions, as long as they are blocked so that the subject knows in advance what type of material will be presented on each trial in the experiment. Thus, in simulating the Johnston and McClelland (1973) results, we assumed different readout times for letters in words and letters in unrelated context, with the different times selected on the basis of practice to optimize performance on each type of material. However, this is not a critical characteristic of the account. The word advantage is only reduced slightly if the same readout time is chosen for both single letters and letters in words, based on optimal performance averaged over the two material types.

Employing the parameter values given in Table 1 with the high value of the letter to word inhibition parameter and the moderate intensity input parameters employed in the figure, we get 81 percent correct on the letters embedded in words and 66 percent correct for letters in a # context or isolated single letters with a 15-cycle target presentation followed immediately by the mask. The results were hardly effected at all by using the lower value of letter to word inhibition, for reasons which will be clearer when we consider the effect of this parameter on activation at the word level in the section on the perception of pronounceable nonwords below. For either parameter value, the model provides a close account of the Johnston-McClelland data.

We have explored our model over a substantial range of input parameter values and have obtained large word advantages over single letters over much of the range. In the case of very high intensity inputs, however, we were forced to add an additional assumption to produce a reasonably large word advantage. As we already noted, when the input is very strong the effect of feedback is to increase the persistence, rather than the height of the letter activation curves. But as we increase the intensity of the display we also increase the potency of the mask. Eventually, the mask becomes so strong that it can drive activations for both single letters and letters embedded in words down so quickly that there is little difference between them. In order to get the advantage in this case, it was necessary to adopt the assumption that there is a maximum inhibitory effect that can be exerted from the feature to the letter level. A value of .55 works out well over a large range of stimulus intensities. Note that for low or moderate values of input strength this parameter does not come in to play, but it is quite important in the case of a very high quality display.

Such high quality input conditions represent a kind of upper extreme of the range we have explored. We have also explored what happens with low quality inputs in which the stimulus quality is so poor that some of the features may go undetected. These conditions produce a reasonable word advantage also, but only as long as a lower value of letter to word inhibition is adopted. As we saw before, with degraded input it is necessary to use a lower value of letter to word inhibition in order to allow words to become activated even when there are multiple letter possibilities active in some or all of the letter positions.

Effects of Masking with Letters and Words

Several studies in the recent literature examine the effects on word perception of following the target with a mask which is composed of letters or words, as opposed to a patterned stimulus containing nonsense squiggles or nonletter printing characters (Jacobson, 1973, 1974; Taylor & Chabot, 1978). In all three of these studies, it appears that performance on words is worse when the mask contains unrelated letters or words than it is when the mask contains nonletters, and there is little or no difference between words and unrelated letter strings as masks, as long as the word is unrelated to the target. One of us has recently collaborated in a study using the Reicher procedure which shows analogous results (Johnston & McClelland, in press). In addition, we find that the presence of letters in the mask hurts performance on single letter displays very little compared to the extent to which it hurts performance on letters in words. Thus, the word advantage over single letters is reduced when a mask containing letters is used, compared to non-letter patterned masks.

In these experiments, Johnston and McClelland (in press) compared performance on single letters and letters in words under three types of masking conditions: Masking with words, masking with random letter sequences, and masking with non-letter characters formed by recombining fragments of letters to make non-letters. One experiment compared perception of letters and words when the stimuli were masked with non-letter mask characters and when they were masked with words. Each condition was tested in a separate block of trials, to allow subjects to try to optimize their performance in each condition. As in most word perception experiments, target duration was varied between subjects to find a duration for each subject at which about 75% correct average performance over all material types was achieved. The results, shown in Table 3, indicate that there was a large word advantage with the non-letter masks. This replicates the typical finding in such studies. The interesting finding is that the word advantage is considerably reduced with word masks. This is true even though the non-letter character masks contain the same set of line segments occurring in the letters used in the word masks.

Table 3

Actual & Simulated Results

(Probability Correct Forced Choice)

Johnston & McClelland (in press)

	Target Type		
	Word	Letter	Difference
Experiment I			
Nonletter Mask	.86	.71	.15
Word Mask	.74	.68	.06
Experiment II			
Word Mask	.78	.75	.03
Letter Mask	.78	.75	.03
Experiment III			
Nonletter Mask	.86	.65	.21
Letter Mask	.79	.71	.08
Simulation			
Nonletter Mask	.90	.70	.20
Letter Mask	.76	.69	.06
Word Mask	.76	.69	.06

Note: In Experiment III, target duration was 10 msec longer with letter masks than with nonletter masks, in order to produce the observed cross-over interaction.

A second experiment compared performance on words and single letters using two kinds of masks containing letters. In one, the letters spelled words as in Experiment I; in the other they formed unrelated letter strings. Both types of material produced a very slight word advantage, and there was no difference between them.

The third experiment compared performance on words and single letters with the same non-letter masks used in the first experiment, and with masks containing four unrelated letters. Target duration was set slightly longer in the letter mask condition to achieve approximately the same overall percent correct performance level in each of the two mask conditions. That is, target duration was always set to be 10 msec longer with letter mask than with the feature mask. The manipulation was successful in eliminating the overall difference between feature and letter mask conditions, but did not eliminate the interaction of target and mask type. The size of the word advantage over nonwords was more than twice as great in the feature mask condition as in the letter mask condition.

Our model provides a simple account of the main findings as illustrated in Figure 9. In the case of word targets, the letters in the mask become active before the output reaches its maximum strength. These new activations compete with the old ones produced by the target to reduce the probability of correctly encoding the target letter. A secondary effect of the new letters is to inhibit the activation of the word (or words) previously activated by the mask. This indirectly results is an increase in the rate of decay of the target letters, because their top-down support is weakened. A tertiary effect of the mask, if it actually contains a word, is to begin activating a new word

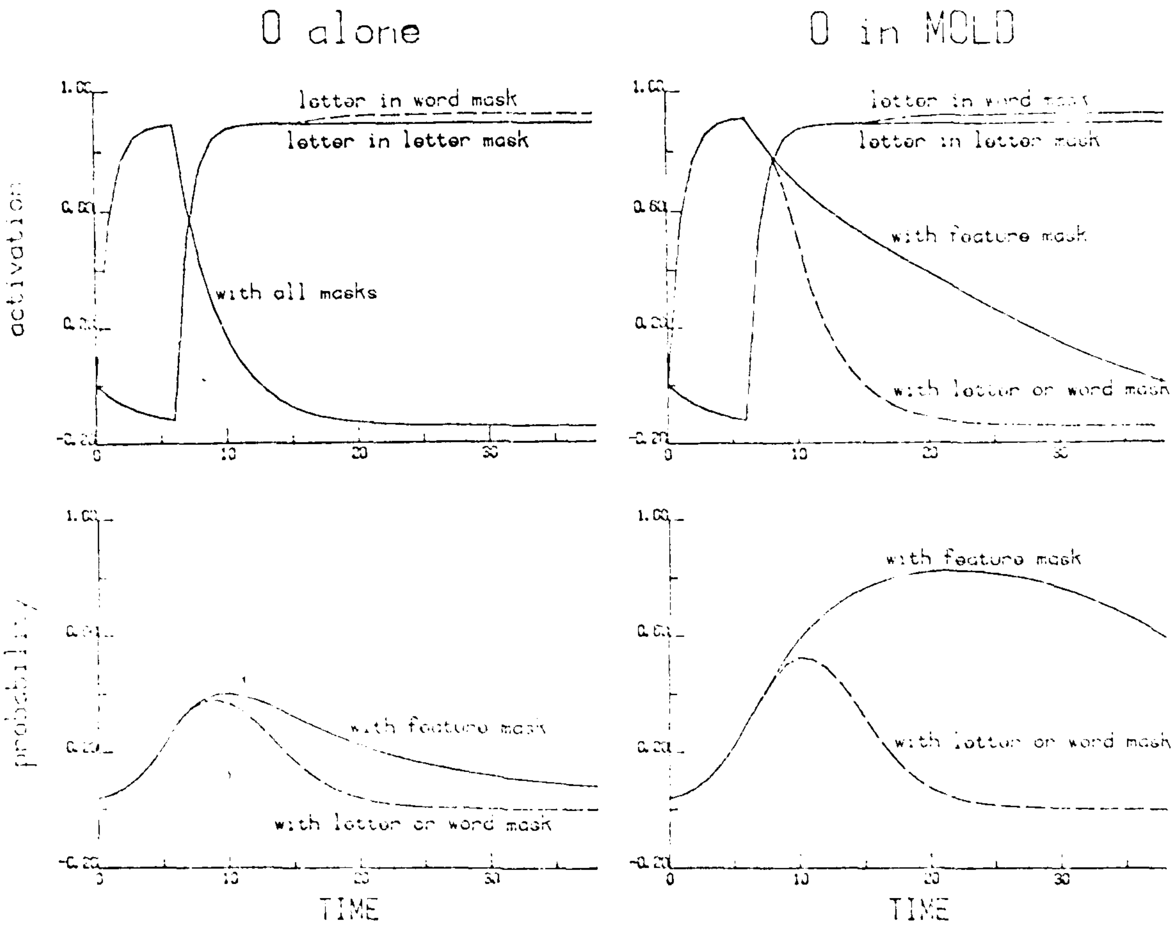

Figure 9. Activation functions (top) and output probability curves (bottom) for the letter O, both alone (left) and in the word MOLD (right), with feature, letter, and word masks.

at the word level. These later two effects do not actually come into play until after the peak of the output function has already passed, so they have no effect on performance.

According to this interpretation, the major role of letters in the mask is to compete at the letter level with the letters previously activated by the target. Competition of this sort also happens with single letter targets as well, but it has less of an effect in this case for the following reason. The activations for single letter targets are not reinforced by the word level, and so the bottom-up inhibition generated by the mask more quickly drives the old activations down. By the time the mask has a chance to activate new letters, the peak in the output function has already been reached. The new letters definitely have an effect on the tail of the output function, but we assume that subjects read out at or near the peak so these differences are irrelevant.

In preliminary attempts to simulate these results, we found that the model was quite sensitive to the similarity of the letters in the target and the feature-arrays (be they letters or non-letters) in the mask. We therefore tailored the non-letter mask characters to have the same number of features different from the target letter they were masking as the mask letters had. For this reason, it was not feasable to test a large number of different items. Instead, we tested all four letters in the word MOLD. The letter mask display was ARAT, and the four feature masks were constructed so that the first had the same number of features in common with M as the letter A did, the second had the same number of features in common with O as R did, etc. For the word mask, we simply altered the lexicon of the program so that ARAT "became" a word (if only such manipulations could be used on human subjects!). Thus, we have tests of four different letters (M,O,L, and D) at each joint level of target type (word, single letter) and mask type (feature, letter word), and all three masks types are exactly equated in their bottom-up potency.

The results of the simulation are summarized in the Table 3. In producing an interaction of this magnitude, we had to assume very high levels of feature to letter excitation and inhibition (.04 and 1.2, respectively). Under these conditions, the the bulk of the effect of feedback is to increase the persistence (rather than the height) of the activation function. The strong input values for the mask also permit the new letters in the mask to produce new activations very rapidly at the letter level, thus contributing to the size of the interaction.

The simulation results shown in the Table were produced using the strong value (.21) of letter to word inhibition. It seems appropriate to use the strong value since the subjects expected only words, as discussed in the next section (with this value, the fact that ARAT is pronounceable is irrelevant to the functioning of the model, as we shall see). In fact though, the simulation produces the interaction both with strong and weak letter to word inhibition, although it is somewhat weaker with weak letter to word inhibition. The reason for the difference has to do with the strength of the secondary effect of the mask letters in inhibiting the word(s) activated by the target, thereby removing the support of the activations of the letters in the target word. With stronger letter to word inhibition, this effect is stronger than when the letter to word inhibition is weak.

The Johnston & McClelland (in press) experiment was designed as a test of a hierarchical model of word perception. in which there was no feedback from the word level to the letter level. Instead, readout could occur from either the letter level or the word level. The greater effectiveness of letter masks was assumed to be due to activation of new letters which would provide disruptive input to the word level. In our model, the greater effectiveness of letter masks is also assumed to be due to activation of new letters, but for a slightltly different reason. Instead of interfering directly with the representation at the word-level, the new letters produce the bulk of their effect by interfering with the readout of old activations at the letter level which are being maintained by feedback. We have not been able to think of a way of distinguishing these views, since they differ mainly in the level of the system from which readout occurs, something which may be very difficult to assess directly. In any case, it is clear that our model provides an account of the effect of mask letters, in addition to its account of the basic effects of patterned and unpatterned masks.

Perception of Regular Nonwords

One of the most important findings in the literature on word perception is that an item need not be a word in order to produce facilitation with respect to unrelated letter or single letter stimuli. The advantage for pseudowords over unrelated letters has been obtained in a very large number of studies (Aderman & Smith, 1971; Baron & Thurston, 1973; Carr, et al, 1978; McClelland, 1976; Spoehr & Smith, 1975). The pseudoword advantage over single letters has been obtained in three studies (Carr, et al, 1978; Massaro & Klitzke, 1979; McClelland & Johnston, 1977).

As we have already noted, these effects appear to depend on subjects' expectations. When subjects know that the stimuli include pseudowords, both words and pseudowords have an advantage over unrelated letters (and single letters) and the difference between words and pseudowords is quite small. In some studies, no reliable difference is obtained (Spoehr & Smith, 1975; Baron & Thurston, 1973; McClelland & Johnston, 1977) whereas in others, a difference has been reported of up to about 6% (Carr, et al, 1978; Manelis, 1974; McClelland, 1976).

Interestingly, when subjects do not expect pseudowords to be shown, letters in these stimuli have no advantage over unrelated letters. Aderman and Smith (1971) found that this was true when the subjects expected only unrelated letters. Carr, et al (1978) replicated this effect, and added two very interesting facts (Table 4). First, the word advantage over unrelated letters can be obtained when subjects expect only unrelated letters, even though letters in pseudowords show no reliable advantage at all under these conditions. Second, when subjects expect only words they perform quite poorly on letters in pseudowords compared to unrelated letters.

At first glance, these data seem to suggest that there must be different processing mechanisms responsible for the word and pseudoword effects. There seems to be a word mechanism which is engaged automatically if the stimulus is a word, and a pseudoword mechanism which is brought into play only if pseudowords are expected. However, we will show that these results are completely consistent with the view that there is a single mechanism for processing both words and pseudowords, with a parameter which is under the subject's control determining whether the mechanism will produce a facilitation only for words

Table 4

Effect of Expected Stimulus Type

on the Word and Pseudoword Advantage over Unrelated Letters

(Difference in Probability Correct Forced Choice)

Carr, et al, 1978

	Expectation		
Target	Word	Pseudoword	Unrelated Letters
Word	.15	.15	.16
Pseudoword	.03	.11	-.02

or for both words and pseudowords. First, we will examine how the model accounts for the pseudoword advantage at all.

The Basic Pseudoword Advantage

The model produces the facilitation for pseudowords by allowing them to activate nodes for words which share more than one letter in common with the display. When they occur, these activations produce feedback, just as in the case of words, strengthening the letters which gave rise to them. These activations occur in the model if the strength of letter to word inhibition is reasonably small compared to the strength of letter to word excitation.

To see how this takes place in detail, consider a brief presentation of the pseudoword MAVE, followed by a patterned mask (the pseudoword is one used by Glushko, 1979, in developing the idea that partial activations of words are combined to derive pronunciations of pseudowords). For this example, the input parameters corresponding to the moderate quality display were used, in conjunction with low letter to word inhibition. As illustrated in Figure 10, presentation of MAVE results in the initial activation of 16 different words. Most of these words, like 'have' and 'gave', share three letters in common with MAVE. By and large, these words steadily gain in strength while the target is on, and produce feedback to the letter level, sustaining the letters which supported them.

Some of the words are weakly activated for a brief period of time before they fall back below zero. These, typically, are words like 'more' and 'many' which share only two letters with the target but are very high in frequency, so they need little excitation before they exceed threshold. But, soon after

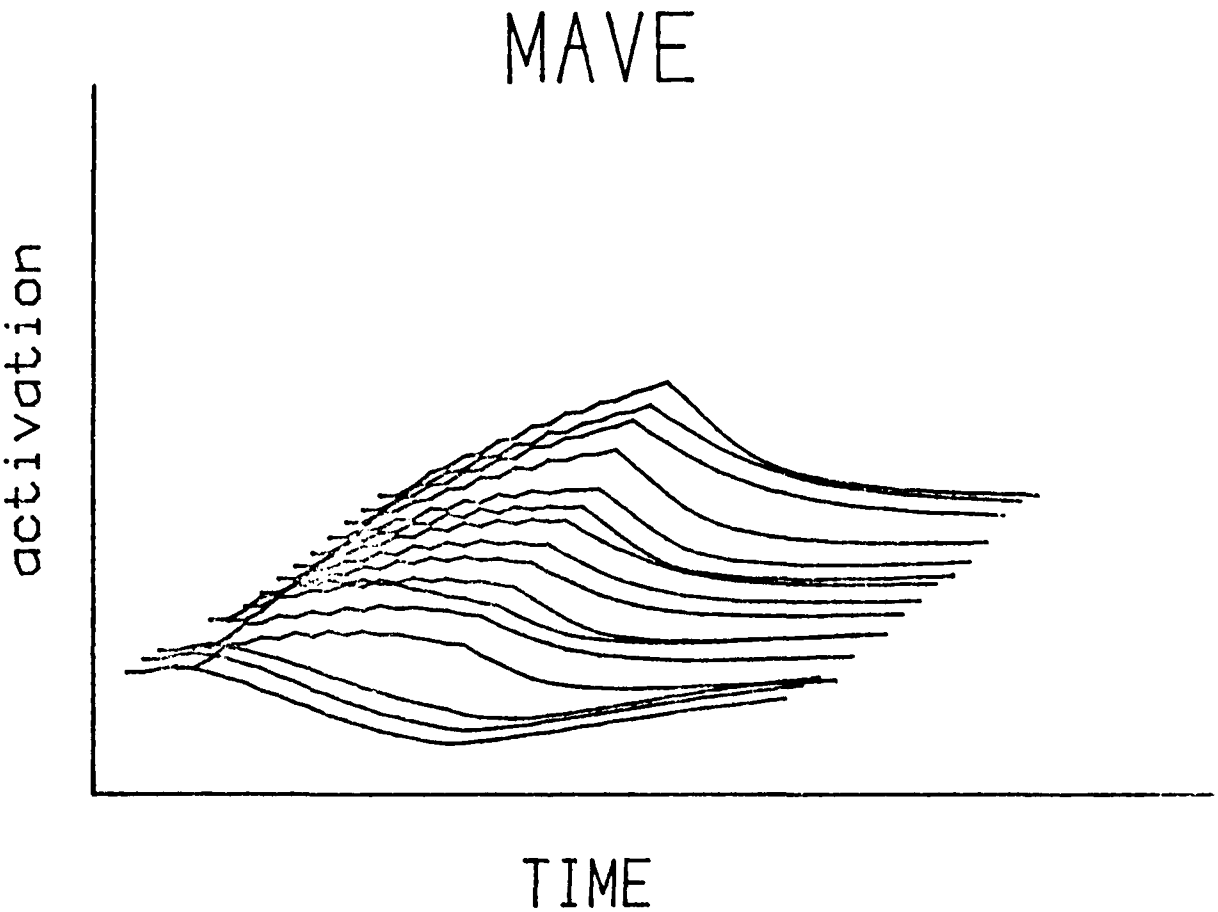

Figure 10. Activation at the word level upon presentation of the nonword MAVE.

they exceed threshold, the total activation at the word level gets strong enough to overcome the weak excitatory input, causing them to drop down just after they begin to rise. Less frequent words sharing two letters with the word displayed have a less exciting fate still. Since they start out initially at a lower value, they generally fail to receive enough excitation to make it up to threshold. Thus, words which share only two letters in common with the target tend to exert a rather minimal influence on the amount of feedback being generated. In general then, the amount of feedback, and hence the amount of facilitation, depends primarily on the activation of nodes for words which share three letters with a displayed pseudoword. It is the nodes for these words which primarily interact with the activations generated by the presentation of the actual target display, so in what follows we will use the word neighborhood to refer to the set of words which have three letters in common with the target letter string.

The amount of feedback a particular letter in a nonword receives depends, in the model, on two primary factors and two secondary factors. The two primary factors are the number of words in the entire nonword's neighborhood which include the letter, and the number of words which do not. In the case of the M in MAVE, for example, there are 7 words in the neighborhood of MAVE which begin with M, so the 'm' node gets excitatory feedback from all of these. These words are called the "friends" of the 'm' node in this case. Because of competition at the word level, the amount of activation which these words receive depends on the total number of words which share three letters in common with the target. Those which share three letters with the target but are inconsistent with 'm' (e.g., 'have') produce inhibition which tends to limit the activation of the friends of 'm', and can thus be considered the enemies of 'm'. These words also produce feedback which tends to activate letters which were not actually presented. For example, activation from 'have' produces excitatory input to 'h', thereby producing some competition with the 'm'. These activations, however, are usually not terribly strong. No one word gets very strongly active, and so letters not in the actual display tend to get fairly weak excitatory feedback. This weak excitation is usually insufficient to overcome the bottom-up inhibition acting on non-presented letters. Thus, in most cases, the harm done by top-down activation of letters which were not shown is minimal.

A part of the effect we have been describing is illustrated in Figure 11. Here, we compare the activations of the nodes for the letters in MAVE. Without feedback, the four curves would be identical to the one "single letter" curve included for comparison. So, although there is facilitation for all four letters, there are definitely differences in the amount, depending on the number of friends and enemies of each letter. Note that within a given pseudoword, the total number of friends and enemies (i.e., the total number of words with three letters in common) is the same for all the letters.

There are two other factors which affect the extent to which a particular word will become active at the word level when a particular pseudoword is shown. Although the effects of these factors are only rather weakly reflected in the activations at the letter level, they are nevertheless interesting to note, since they indicate some synergistic effects which emerge from the interplay of simple excitatory and inhibitory influences in the neighborhood. These are the <u>rich-get-richer</u> effect and the <u>gang</u> effect. The rich-get-richer effect is illustrated in Figure 12, which compares the activation curves for

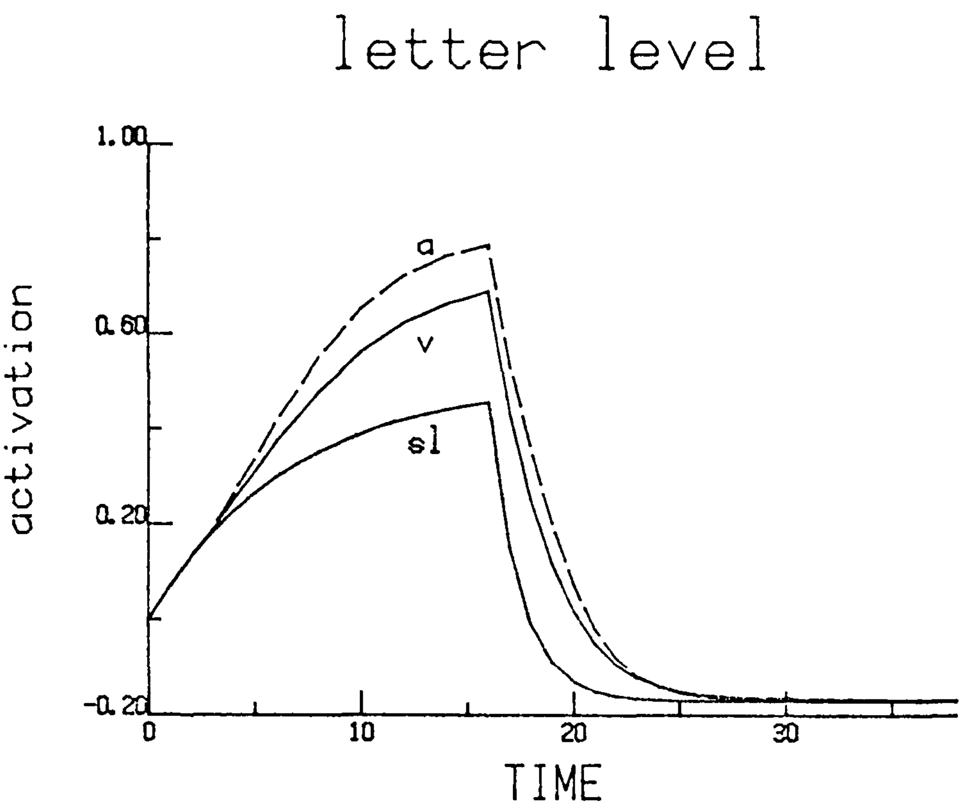

Figure 11. Activation functions for the letters 'a' and 'v' in on presentation of MAVE. Activation function for 'e' is indistinguishable from function for 'a', and that for 'm' is similar to that for 'v'. The activation function for a letter alone or in unrelated context is included for comparison.

the nodes for 'have', 'gave', and 'save' under presentation of MAVE. The words differ in frequency, which gives the words slight differences in base-line activation. What is interesting is that the difference gets magnified, so that at the point of peak activation there is a much larger difference. The reason for the amplification can be seen by considering a system containing only two nodes 'a' and 'b', starting at different initial positive activation levels, 'a' and 'b' at time t. Let's suppose that 'a' is stronger than 'b' at t. Then at t+1, 'a' will exert more of an inhibitory influence on 'b', since inhibition of a given node is determined by the sum of the activations of all units other than itself. This advantage for the initially more active nodes is compounded further in the case of the effect of word frequency by the fact that more frequent words creep above threshold first, thereby exerting an inhibitory effect on the lower frequency words when they are still too weak to fight back at all.

Even more interesting is the gang effect, which depends on the coordinated action of a related set of word nodes. This effect is depicted in Figure 13. Here, the activation curves for the 'move, 'make', and 'save' nodes are compared. In the language, 'move' and 'make' are of approximately equal frequency, so their activations start out at about the same level. But they soon pull apart. Similarly, 'save' starts out below 'move', but soon reaches a higher activation. The reason for these effects is that 'make' and 'save' are both members of gangs with several members, while 'move' is not. Consider first the difference between 'make' and 'move'. The reason for the difference is that there are several words which share the same three letters in common with MAVE as 'make' does. In the list of words used in our simulations, there are 6. These words all work together to reinforce the 'm', the 'a', and the

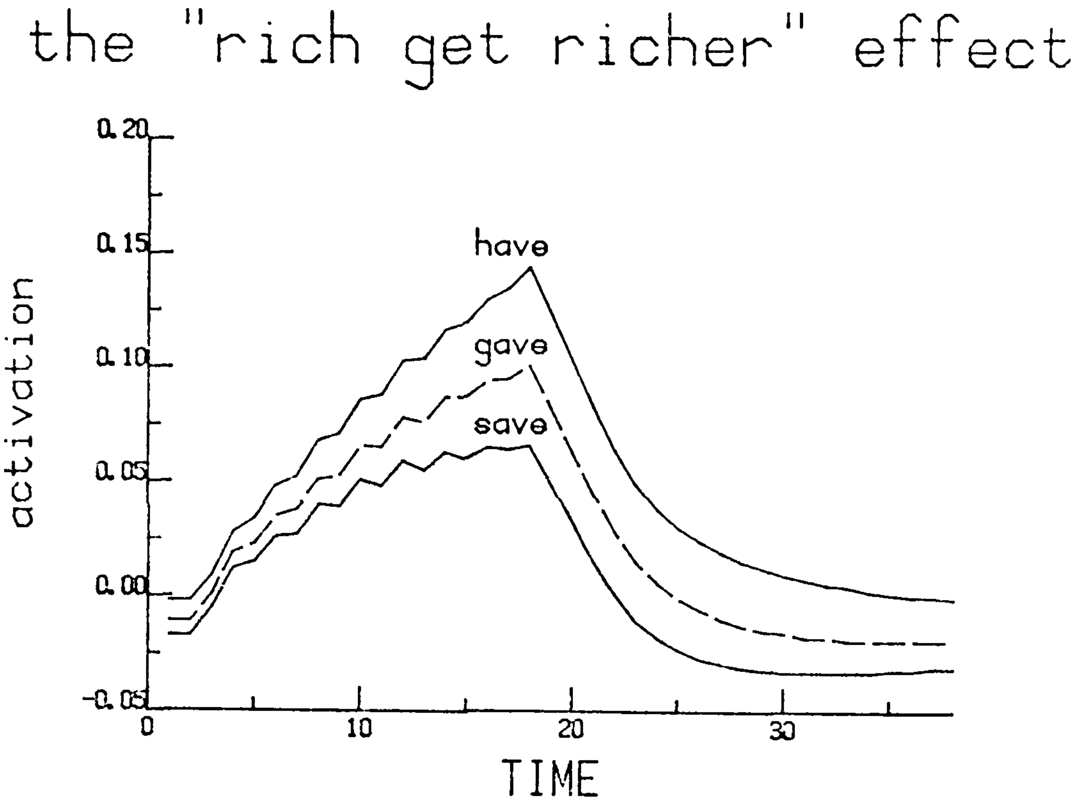

Figure 12. The rich-get-richer effect. Activation functions for the nodes for 'have', 'gave' and 'save', under presentation of MAVE.

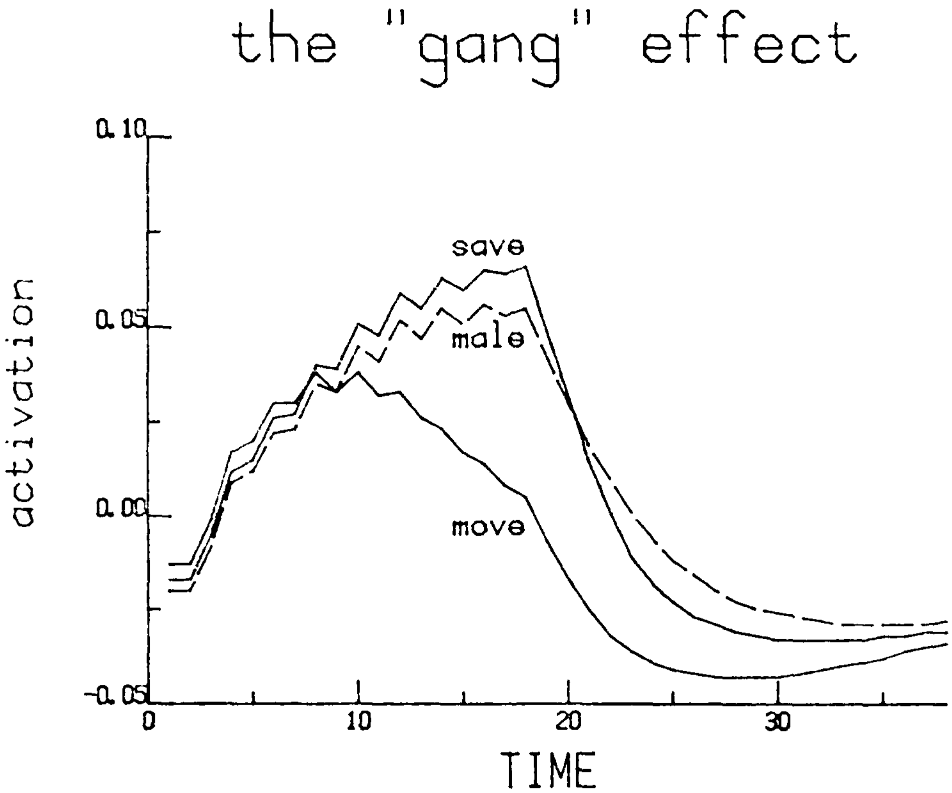

Figure 13. The gang effect. Activation functions for 'move', 'male' and 'save' under presentation of MAVE.

'e', thereby producing much stronger reinforcement for themselves. Thus, these words make up a gang called the 'ma_e' gang. In this example, there is also a '_ave' gang consisting of a different 6 words, of which 'save' is one. All of these work together to reinforce the 'a', 'v', and 'e'. Thus, the 'a' and 'e' are reinforced by two gangs, while the letters 'v' and 'm' are reinforced by only one each. Now consider the word 'move'. This word is a loner; there are no other words in its gang, the 'm_ve' gang. Although two of the letters in 'move' receive support from one gang each, and one receives support from both other gangs, the letters of 'move' are less strongly enhanced by feedback than the letters of the members of the other two gangs. Since continued activation of one word in the face of the competition generated by all of the other partially activated words depends on the activations of the component letter nodes, the words in the other two gangs eventually gain the upper hand and drive 'move' back below the activation threshold.

As our study of the MAVE example illustrates, the pattern of activation which is produced by a particular pseudoword is complex and idiosyncratic. In addition to the basic friends and enemies effects, there are also the rich-get-richer and the gang effects. These effects are primarily reflected in the pattern of activation at the word level, but they also exert subtle influences on the activations at the letter level. In general, though, the main result is that when the letter to word inhibition is low, all four letters in the pseudoword receive some feedback reinforcement. The result, of course, is greater accuracy reporting letters in pseudowords compared to single letters.

The Role of Expectations

It should now be clear that variation in letter to word inhibition produces different degrees of enhancement. When this parameter is small, the pseudoword advantage is large, and when the parameter is large, the advantage gets small. Indeed, if the letter to word inhibition is equal to three times the letter to word excitation, then no four-letter nonword can activate the node for any four-letter word. The reason is that it can have no more than three letters in common with a word. The inhibition generated by the letter which is different will cancel the excitation generated by the letters that are the same.

We can now account for Carr, et al's (1978) findings with pseudowords by simply assuming that when subjects expect only words they will adopt a large value of the letter to word inhibition parameter, but when they expect pseudowords they adopt a small value. Apparently, when they expect unrelated letter strings, at least of the type used in this experiment, they also adopt a large value of letter to word inhibition. Perhaps this is the normal setting, with a relaxation of letter to word inhibition only used if pseudowords are known to occur in the list or when the stimulus input is very degraded.

But we have still to consider what effects variation of letter to word inhibition might have for word stimuli. If relaxation of letter to word inhibition increases accuracy for letters in pseudowords, we might expect it to do the same thing for letters in words. However, in general this is not the case. Part of the reason is that the word shown still gets considerably more activation than any other word, and tends to keep the activations of other nodes from getting very strong. This situation is illustrated for the word CAVE in Figure 14. A second factor is that partial activations of other words are not an unmixed blessing. The words which receive partial activations all produce inhibition which keeps the activation of the node for the word shown

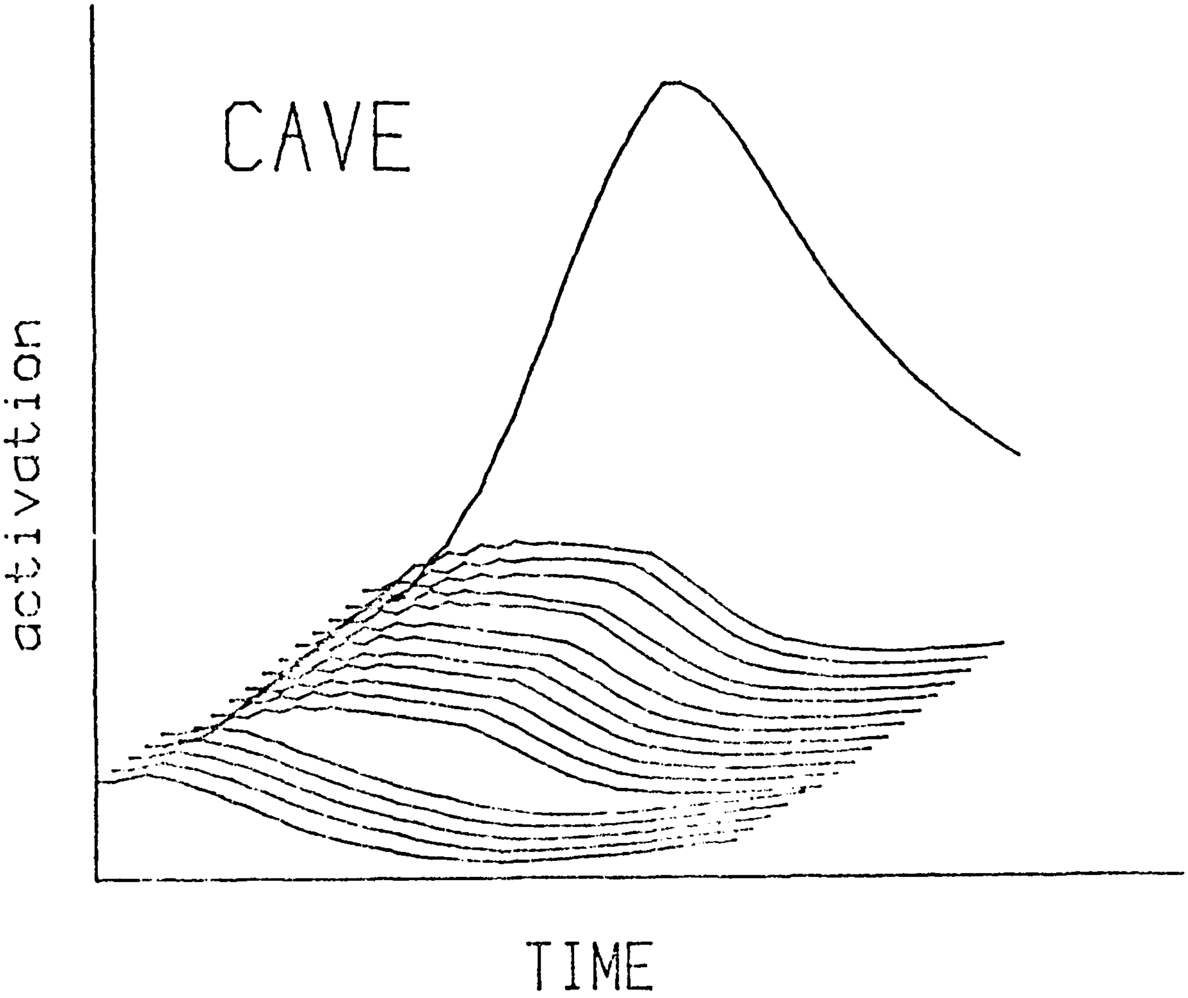

Figure 14. Activity at the word level upon presentation of CAVE, with weak letter to word inhibition.

from getting activated as strongly as it would be otherwise. The third factor is that the activations of any one word sharing three letters with the word shown only reinforce three of the four letters in the display. For these reasons, it turns out that the value of letter to word inhibition can vary from .04 to .21 with very little effect on word performance.

Comparison of Performance on Words and Pseudowords

Let us now consider the fact that the word advantage over pseudowords is generally rather small in experiments where the subject knows that the stimuli include pseudowords. Some fairly representative results, from the study of McClelland and Johnston (1977) are illustrated in Table 5. The visual conditions of the study were the same as those used in the patterned mask condition in Johnston and McClelland (1973). Trials were blocked, so subjects could adopt the optimum strategy for each type of material. The slight word-pseudoword difference, though representative, is not actually statistically reliable in this study.

Words differ from pseudowords in that they strongly activate one node at the word level. While we would tend to think of this as increasing the amount of feedback for words as opposed to pseudowords, there is the word-level inhibition which must be taken into account. This inhibition tends to equalize the total amount of activation at the word level between words and pseudowords. With words, the word shown tends to dominate the pattern of activity, thereby keeping all the words with three letters in common with it from achieving the activation level they would reach in the absence a node activated by all four letters. The result is that the sum of the activations of all the active units at the word level is not much different between the

Table 5

Actual and Simulated Results of the
McClelland & Johnston (1977) Experiments
(Probability Correct Forced Choice)

	Target Type		
	Word	Pseudoword	Single Letter
Data			
High BF	.81	.79	.67
Low BF	.78	.77	.64
Average	.80	.78	.66
Simulation			
High BF	.81	.79	.67
Low BF	.79	.77	.67
Average	.80	.78	.67

two cases. Thus, CAVE produces only slightly more facilitation for its constituent letters than MAVE as illustrated in Figure 15.

In addition to the mere leveling effect of competition at the word level, it turns out that one of the features of the design of most studies comparing performance on words and pseudowords would operate in our model to keep performance relatively good on pseudowords. In general, most studies comparing performance on words and pseudowords tend to begin with a list of pairs of words differing by one letter (e.g., PEEL-PEEP), from which a pair of nonwords is generated differing from the original word pair by just one of the context letters, thereby keeping the actual target letters and as much of the context as possible the same between word and pseudoword items (e.g., TEEL-PEEL). A previously unnoticed side-effect of this matching procedure is that it ensures that the critical letter in each pseudoword has at least one friend, namely the word from the matching pair which differs from it by one context letter. In fact, most of the critical letters in the pseudowords used by McClelland and Johnston tended to have relatively few enemies, compared to the number of friends. In general, a particular letter should be expected to have three times as many friends as enemies. In the McClelland and Johnston stimuli, the great majority of the stimuli had much larger differentials. Indeed, more than half of the critical letters had no enemies at all.

The Puzzling Absence of Cluster Frequency Effects

In the account we have just described, facilitation of performance on letters in pseudowords was explained by the fact that pseudowords tend to activate a large number of words, and these words tend to work together to reinforce the activations of letters. This account might seem to suggest that

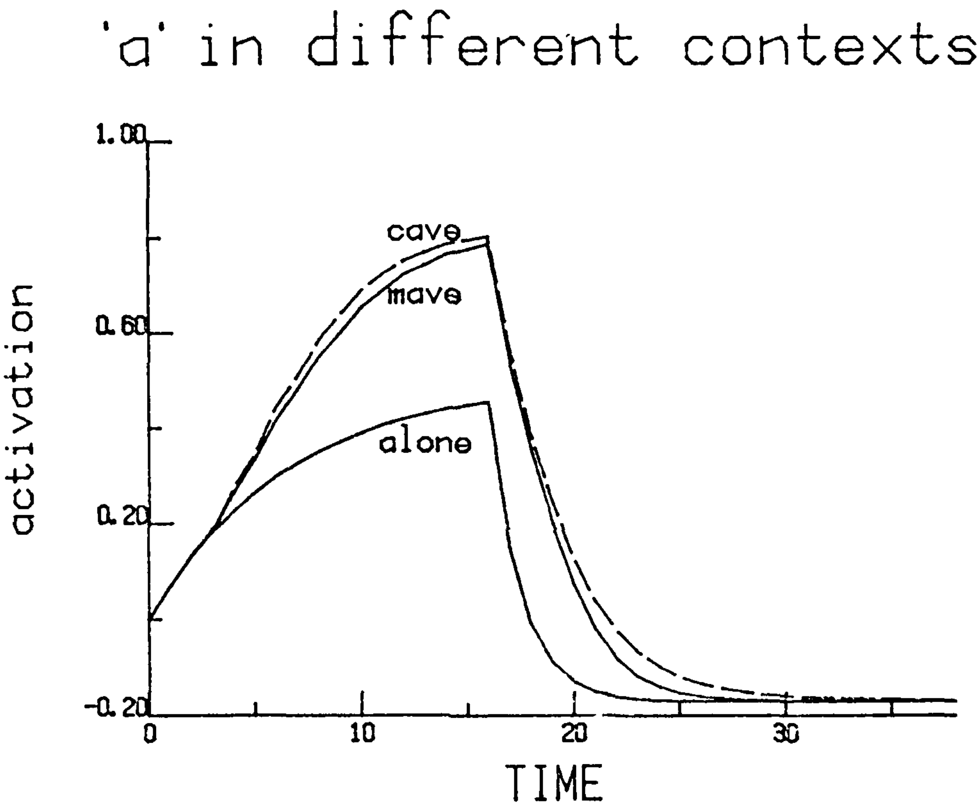

Figure 15. Activation functions for the letter 'a', under presentation of CAVE and MAVE, and alone.

pseudowords which have common letter-clusters, and therefore have several letters in common with many words, would tend to produce the greatest facilitation. However, this factor has been manipulated in a number of studies and little has been found in the way of an effect. The McClelland and Johnston study is one case in point. As the table illustrates, there is only a slight tendency for superior performance on high cluster frequency words. This slight tendency is also observed in single letter control stimuli, suggesting that the difference may be due to differences in perceptibility of the target letters in the different positions, rather than cluster frequency per se. In any case, the effect is very small. Others studies have likewise failed to find any effect of cluster frequency (Spoehr & Smith, 1975; Manelis, 1974). The lack of an effect is most striking in the McClelland and Johnston study, since the high and low cluster frequency items differed widely in cluster frequency as measured in a number of different ways.

In our model, the lack of a cluster frequency effect is due to the effect of mutual inhibition at the word level. As we have seen, this mutual inhibition tends to keep the total activity at the word level roughly constant over a variety of different input patterns, thereby greatly reducing the advantage for high cluster frequency items. Items containing infrequent clusters will tend to activate few words, but there will be less competition at the word level, so that the words which do become active will reach higher activation levels.

The situation is illustrated for the nonwords TEEL and HOEM in Figure 16. While TEEL activates many more words, the total activation is not much different in the two cases.

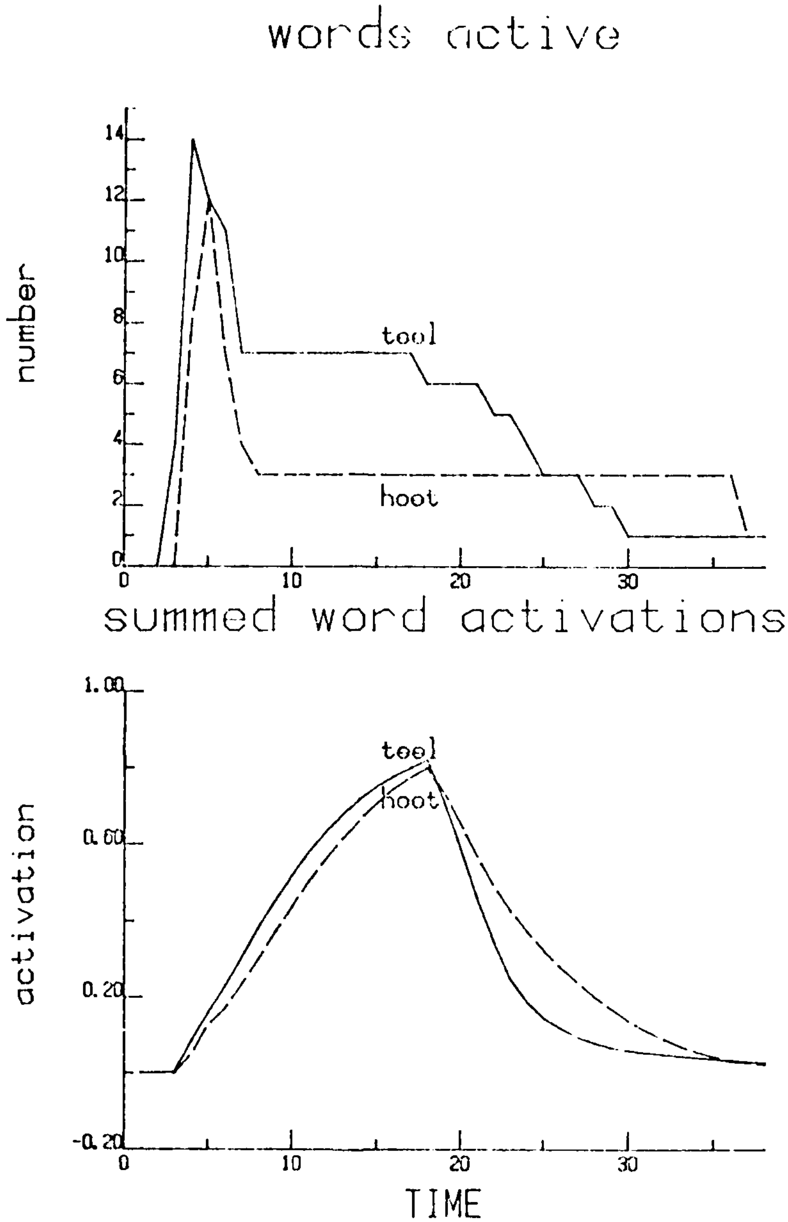

Figure 16. The number of words activated (top) and the total activation at the word level (bottom) upon presentation of the nonwords TEEL and HOEM.

The total activation is not, of course, the whole story. The ratio of friends to enemies is also important. And, it turns out that this ratio is working against the high cluster items more than the low cluster items. It turns out that in McClelland and Johnston's stimuli only one of the low cluster frequency nonword pairs had critical letters with any enemies at all! For 23 out of 24 pairs, there was at least one friend (by virtue of the method of stimulus construction), and no enemies. In contrast, for the high cluster frequency pairs, there was a wide range, with some items having several more enemies than friends.

To simulate the McClelland and Johnston results, we had to select a subset of their stimuli, since many of the words they used were not in our word list. Since the stimuli had been constructed in sets containing a word pair, a pseudoword pair, and a single letter pair differing by the same letters in the same position (e.g., PEEL-PEEP TEEL-TEEP; ___L-___P), we simply selected all those sets in which both words in the pair appeared in our list. This resulted in a sample of 10 high cluster frequency sets and 10 low cluster frequency sets. The single letter stimuli derived from the high and low cluster frequency pairs were also run through the simulation. Both members of each pair were tested.

Since the stimuli were presented in the actual experiment blocked by material type, we selected an optimal time for readout separately for words, pseudowords, and single letters. Readout time was the same for high and low cluster frequency items of the same type, since these were presented in a mixed list in the actual experiment. The run shown in the table used the following parameters: letter to word inhibition was set to the low value (.04);

the input parameters associated with the moderate quality display were used (feature to letter excitation = .005, inhibition = .15). The display was presented for a duration of 15 cycles.

The simulation shows the same general pattern as the actual data. As in the actual data, the magnitude of the pseudoword advantage over single letters is just slightly smaller than the word advantage, and the effect of cluster frequency is very slight. Qualitatively similar results are obtained when the input parameters associated with the very high quality display are used. For the word condition, it makes very little difference if the value of letter to word inhibition is high or low, except that the slight advantage for high cluster frequency words is eliminated.

We have yet to consider how the model deals with unrelated letter strings. This depends a little on the exact characteristics of the strings, and the value of letter to word inhibition. With high letter to word inhibition, unrelated letters fare no better than pseudowords: they fail to excite any words, and there is no feedback. When the value of letter to word inhibition gets low, there is some activity at the word level with many so-called unrelated letter strings. Generally speaking, however, these strings rarely have more than two letters in common with any one word. Thus, they only tend to activate a few words very weakly, and because of the weakness of the bottom-up excitation, competition among partially activated words keeps any one from getting very active. So, little benefit results. When we ran our simulation on randomly-generated consonant strings, there was only a 1% advantage over single letters.

Some items which have been used as unpronounceable nonwords or unrelated letter strings do produce a weak facilitation. We ran the nonwords used by McClelland and Johnston (1977) in their Experiment 2. These items contain a large number of vowels in positions which vowels tend to occupy in words, and they therefore tend to activate more words than, say, random strings of consonants. The simulation was run under the same conditions as the one reported above for McClelland and Johnston's first experiment. The experiment produced a slight advantage for letters in these nonwords, compared to single letters, as did the experiment. In both the simulation and the actual experiment, forced-choice performance was 4% more accurate for letters in these unrelated letter strings than in single letter stimuli.

On the basis of this characteristic of our model, the results of one experiment on the importance of vowels in reading may be reinterpreted. Spoehr and Smith (1975) found that subjects were more accurate reporting letters in unpronounceable nonwords containing vowels than in all consonant strings. They interpreted the results as supporting the view that subjects parse letter strings into "Vocalic Center Groups." However, an alternative possible account is that the strings containing vowels had more letters in common with actual words than the all consonant strings.

In summary, the model provides a good account of the perceptual advantage for letters in pronounceable nonwords but not unrelated letter strings. In addition, it accounts for the dependence of the pseudoword advantage on expectation and for the lack of an effect of expectation on the advantage for letters in words. Third, the model accounts for the small difference between performance on words and pseudowords when the subject is aware that the stimuli include pseudowords, and for the absence of any really noticeable cluster frequency effect.

Our examination of the model suggests that there are different ways interactive activation can influence perception. When letter to word inhibition is set to a high value, the system acts as a sharply tuned filter. In this mode, the system will reinforce activations only of those patterns which it has explicitly stored in particular nodes. When the same parameter is set to a small value, the system allows for nodes for stored patterns which are similar to the new input to become partially activated, thereby permitting it to reinforce activations of patterns which are not in fact stored. In this mode the model shows the capacity to apply knowledge explicitly encoded as spellings of particular words in such a way that it facilitates the processing of stimuli that are similar to several stored patterns, but not identical to any.

The Role of Lexical Constraints

The Johnston Experiment

Several models which have been proposed to account for the word advantage rely on the idea that the context letters in a word facilitate performance by constraining the set of possible letters which might have been presented in the critical letter position. Models of this class predict that contexts which strongly constrain what the target letter might be result in greater accuracy of perception than more weakly constraining contexts. For example, the context _HIP should facilitate the perception of an initial S more than the context _INK. The reason is that _HIP is more strongly constraining, since only three letters (S, C, and W) fit in the context to make a word, compared to _INK, where nine letters (D, F, K, L, M, P, R, S, and W) fit in the context to make a word. In a test of such models, Johnston (1978) compared accuracy of perception of letters occurring in high and low constraint contexts. The same target letters were tested in the same positions in both cases. For example, the letters S and W were tested in the high constraint _HIP context and the low constraint _INK context. Using bright target/patterned mask conditions, Johnston found no difference in accuracy of perception between letters in the high and low constraint contexts. The results of this experiment are shown in Table 6. Johnston measured letter perception in two ways. He not only asked the subjects to decide which of two letters had been presented (the forced-choice measure), but he also asked subjects to report the whole word and recorded how often they got the critical letter correct. No significant difference was observed in either case. In the forced choice there was a slight difference favoring low constraint items, but in the free report there was no difference at all.

Although our model does use contextual constraints (as they are embodied in specific lexical items), it turns out that it does not predict that highly constraining contexts will facilitate perception of letters more than weakly constraining contests under bright target/pattern mask conditions. Under such conditions, the role of the word level is not to help the subject select among alternatives left open by an incomplete feature analysis process, but rather to help maintain the activation of the nodes for the letters presented.

In Johnston's experiments, only words were shown, so on the basis of our interpretation of the Carr et al (1978) findings mentioned above, we would

Table 6

Actual & Simulated Results from Johnston (1978)

(Probability Correct)

	Constraint	
	High	Low
Actual Results		
Forced Choice	.768	.795
Free Report	.545	.544
Simulation		
Forced Choice	.773	.763
Free Report	.563	.544

Note: Simulation was run using low letter to word inhibition and moderate quality display parameters. Similar results are obtained using high quality display parameters. There is no effect of constraints when high letter to word inhibition is used.

expect that subjects would tend to adopt a large value of letter to word inhibition. If the .21 value were used, our model produces no difference whatsoever between high and low constraint items. The reason is simply that only the node for the word actually shown ever gets activated at all. The nodes for all other words receive either net inhibition or a net neutral input if they share three letters in common with the word shown.

If we assume that a small value of letter to word inhibition is used (.04 instead of .21), our model produces a very small advantage for high constraint items. In this case, the presentation of a target word results in the weak activation of the words which share three letters in common with the target. Some of these words are "friends" of the critical letter in that they contain the actual critical letter shown, as well as two of the letters from the context (e.g., 'shop' is a friend of the initial S in SHIP). Some of the words, however, are "enemies" of the critical letter, in that they contain the three context letters of the word, but a different letter in the critical letter position (e.g, 'chip' and From our point of view, Johnston's constraint manipulation is essentially a manipulation of the number of enemies the critical letter has in the given context. It turns out that Johnston's high and low constraint stimuli have equal numbers of friends, on the average, but (by design), the high constraint items have fewer enemies as shown in Table 7.

Using a low value for the letter to word inhibition results in the friends and enemies of the target word receiving some activation. Under these conditions (with either high or moderate quality input parameters) our model does produce a slight advantage for the high constraint items. The reason for the slight effect is that lateral interference at the word level lets the

Table 7

Friends and Enemies of the

Critical Letters in the

Stimuli Used by Johnston (1978)

	High Constraint			Low Constraint		
	friends	enemies	ratio	friends	enemies	ratio
pos 1	3.33	2.22	.60	3.61	6.44	.36
pos 2	9.17	1.00	.90	6.63	2.88	.70
pos 3	6.30	1.70	.79	7.75	4.30	.64
pos 4	4.96	1.67	.75	6.67	3.50	.66
ave	5.93	1.65		6.17	4.27	

enemies of the critical letter keep the node for the word presented and the nodes for the friends from getting quite as strongly activated as they would otherwise. The effect is quite small for two reasons. First, the node for the word presented receives four excitatory inputs from the letter level, and all other words can only receive at most three excitatory inputs, and at least one inhibitory input. As we saw in the case of the word CAVE, the node for the correct word dominates the activations at the word level, and is predominantly responsible for any feedback to the letter level. Second, while the high constraint items have fewer enemies, by more than a two to one margin, both high and low constraint items have, on the average, more friends than enemies. The friends of the target letter work with the actual word shown to keep the activations of the enemies in check, thereby reducing the extent of their inhibitory effect still further. The ratio of the number of friends over the total number of neighbors is not all that different in the two conditions, except in the first serial position.

This discussion may give the impression that contextual constraint is not an important variable in our model. In fact, it is quite powerful. But its effects are obscured in the Johnston experiment because of the strong dominance of the target word when all the features are extracted, and the fact that we are concerned with the likelihood of perceiving a particular letter rather than performance in identifying correctly what whole word was shown. We will now consider an experiment in which contextual constraints play a strong role, because the characteristics just mentioned are absent.

The Broadbent and Gregory Experiment

Up to now we have found no evidence that either bigram frequency or lexical constraints have any effect on performance. However, in experiments using the traditional whole report method these variables have been shown to have substantial effects. Various studies have shown that recognition thresholds are lower, or recognition accuracy higher at a fixed recognition threshold value, when relatively unusual words are used (Bouwhuis, 1979; Havens & Foote, 1963; Newbigging, 1961). Such items tend to be low in bigram frequency, and at the same time high in lexical constraint.

In one experiment, Broadbent and Gregory (1968) investigated the role of bigram frequency at two different levels of word frequency and found an interesting interaction. We now consider how our model can account for their results. To begin, it is important to note that the visual conditions of their experiment were quite different from those of McClelland and Johnston (1977) in which the data and our model failed to show a bigram frequency effect, and of Johnston (1978) in which the data and the model showed no constraint effect. The conditions were like the dim target/blank mask conditions discussed above, in that the target was shown briefly against an illuminated background, without being followed by any kind of mask. The dependent measure was the probability of correctly reporting the whole word. The results are indicated in Table 8. A slight advantage for high bigram frequency items over low bigram frequency was obtained for frequent words, although it was not consistent over different subsets of items tested. The main finding was that words of low bigram frequency had an advantage among infrequent words. For these stimuli, higher bigram frequency actually resulted in a lower percent

Table 8

Actual and Simulated Results of the
Broadbent & Gregory (1968) Experiment
(Probability Correct Whole Report)

	Word Frequency	
	High	Low
Actual Data		
High BF	.645	.431
Low BF	.637	.583
Simulation		
High BF	.414	.212
Low BF	.394	.371

correct.

Unfortunately, Broadbent and Gregory used 5 letter words, so we were unable to run a simulation on their actual stimuli. However, we were able to select a subset of the stimuli used in the McClelland and Johnston (1977) experiment which fit the requirements of the Broadbent and Gregory design. We therefore presented these stimuli to our model, under the presentation parameters used in simulating the blank mask condition of the Johnston and McClelland (1973) experiment above. The only difference was that the output was taken, not from the letter level, as in all of our other simulations, but directly from the word level. The low value of letter to word inhibition was used, since with a high value few words ever become activated on the basis of partial feature information. The results of the simulation, shown in the Table below the actual data, replicate the obtained pattern very nicely. The simulation produced a large advantage for the low bigram items, among the infrequent words, and produced a slight advantage for high bigram frequency items among the frequent words.

In our model, low frequency words of high bigram frequency are most poorly recognized because these are the words which have the largest number of neighbors. Under conditions of incomplete feature extraction, which we expect to prevail under these visual conditions, the more neighbors a word has the more likely it is to be confused with some other word. This becomes particularly important for lower frequency words. As we have seen, if both a low frequency word and a high frequency word are equally compatible with the detected portion of the input, the higher frequency word will tend to dominate. When incomplete feature information is extracted, the relative activation of the target and the neighbors is much lower than when all the features have been seen. Indeed, some neighbors may turn out to be just as compatible with the features extracted as the target itself. Under these circumstances, the word of the highest frequency will tend to gain the upper hand. The probability of correctly reporting a low frequency word will therefore be much more strongly influenced by the presence of a high frequency neighbor compatible with the input than the other way around.

But why does the model actually produce a slight reversal with high frequency words? Even here, it would seem that the presence of numerous neighbors would tend to hurt instead of facilitate performance. However, we have forgotten the fact that the activation of neighbors can be beneficial, as well as harmful. The active neighbors produce feedback which strengthens most or all of the letters, and these in turn increase the activation of the node for the word shown. As it happens, there turns out to be a delicate balance for high frequency words between the negative and positive effects of neighbors, which only slightly favors the words with more neighbors. Indeed, the effect only holds for some of these items. We have not yet had the opportunity to explore what all the factors are which determine whether the effect of neighbors will balance out to be positive or negative in individual cases.

Different Effects in Different Experiments

This discussion of the Broadbent and Gregory experiment indicates once again that our model is something of a chameleon. The model produces no effect of constraint or bigram frequency under the visual conditions and testing procedures used in the Johnston (1978) and McClelland and Johnston (1977) experiments, but we do obtain such effects under the conditions of the

Broadbent and Gregory (1968) experiment. This flexibility of the model, of course, is fully required by the data. While there are other models of word perception which can account for one or the other type of result, to our knowledge the model presented here is the only scheme that has been worked out to account for both.

Discussion

The interactive activation model does a good job accounting for the results of the literature we have reviewed on the perception of letters in words and nonwords. The model provides a unified account for the results of a variety of experiments, and provides a framework in which the effects of both physical and psychological manipulations of the characteristics of the experiments may be accounted for. In addition, as we shall see in Part II, the model readily accounts for a variety of additional phenomena of word perception. Moreover, as we shall also show, it can be readily extended beyond its current domain of applicability with substantial success. In Part II we will report a number of experiments demonstrating what we call "Context Enhancement Effects," and show how the model can account for the major findings in the experiments.

However, there are some problems which we have either ignored or failed to solve which remain to be resolved. First, we have ignored the fact that there is a high degree of positional uncertainty in reports of letters, particularly letters in unrelated strings, but also in reports of letters in words and pseudowords on occasion (Estes, 1975; McClelland, 1976; McClelland & Johnston, 1977). It is not entirely clear whether these uncertainty effects arise in the perceptual system itself, in the readout process, or both. It is quite possible that letters are kept well-organized by position in the activation system, but the process of reading them out is not easily restricted to a single position channel (cf. Eriksen & Eriksen, 1972). Of course, it is also quite possible that much of the problem arises from positional uncertainty within the activation system itself. Although we have not attempted to model these effects in this paper, our model could easily be modified to account for the rearrangements of letters and the fact that they occur more frequently in unrelated letters than in words and pseudowords. Suppose, for example, that the activations of letters were distributions of activation along a spatial dimension, instead of points of activation assigned to a particular point in an array. Then the activations for letters in adjacent positions would overlap, and if there was noise in the location of the mean of the distribution of activation produced by a letter presented in a particular position, order errors would be expected. Under these circumstances, feedback from the word level could serve to reinforce that portion of the distribution of activation in the correct spatial position, thereby shifting the mean of the distribution toward the right position.

Another thing that we have not considered very fully is the serial position curve. In general, it appears that performance is more accurate on the end letters in multi-letter strings, particularly the first letter. The effect is much more striking for unrelated letters than for pseudowords or words (McClelland & Johnston, 1977). While part of this effect may be due to reduced lateral masking of end letters and/or to a reduced opportunity for order error at the ends of the string, it seems likely that the first position advantage reflects some sort of processing priority given to the first letter. Some or all of this effect could be accommodated by our model by assuming that the strength of the effect exerted by the letter in a given position is influenced by the deployment of attention, and that attention is deployed preferentially to the first letter position.

A different possibility that we considered is that part of the serial position effect could be due to neighborhood effects. However, these would if anything tend to hurt the first letter position relative to other positions for the following reason. The first letter is, generally speaking, the letter which has the most enemies. That is, the largest gangs tend to be those consisting of the last three letters of the item and leaving out the first letter. Thus, the word level will tend to produce greater feedback for the second, third and fourth letter than for the first. In view of this, we can see that one reason for directing attention predominantly to the first letter would be to offset this gang effect.

There are some effects of set on word perception which we have not considered. Johnston and McClelland (1974) found that perception of letters in words was actually hurt if subjects focused their attention on a single letter position in the word (See also Holender, 1979, and Johnston, 1974). One possible interpretation of these effects would be that they result from the narrowing of the focus of attention so that visual information from the nontarget letters is simply not made available to the letter and word levels. Another possibility is that the focusing of attention on the contents of a single letter position disrupts the process of directing the letter information into the correct position-specific channels. It seems likely that either of these possibilities could be worked into our model.

In all but one of the experiments we have simulated, the primary (if not the only) data for the experiments were obtained from forced choices between pairs of letters, or strings differing by a single letter. In these cases, it seemed to us most natural to rely on the output of the letter level as the basis for responding. However, it may well be that subjects often base their responses on the output of the word level. Indeed, we have assumed that they do in experiments like the Broadbent and Gregory (1968) study, in which subjects were told to report what word they thought they had seen. This may also have happened in the McClelland and Johnston (1977) and Johnston (1978) studies, in which subjects were instructed to report all four letters before the forced choice on some trials. Indeed, both studies found that the probability of reporting all four letters correctly for letters in words was greater than we would expect given independent readout of each letter position. It seems natural to account for these completely correct reports by assuming that they often occurred on occasions where the subject encoded the item as a word. Even in experiments where only a forced choice is obtained, subjects may still come away with a word, rather than a sequence of letters on many occasions. In the early phases of the development of our model, we explicitly included the possibility of output from the word level as well as the letter level. We assumed that the subject would either encode a word, with some probability dependent on the activations at the word level or, failing that, would encode some letter for each letter position dependent on the activations at the letter level. However, we found that simply relying on the letter level permitted us to account equally well for the results. In essence, the reason is that the word-level information is incorporated into the activations at the letter level because of the feedback, so that the word level is largely redundant. In addition, of course, readout from the letter level is necessary to the model's account of performance with nonwords. Since it is adequate to account for all of the forced-choice data, and since it is difficult to know exactly how much of the details of free-report data should be attributed to perceptual processes and how much to such things as possible biases in the readout processes, etc., we have stuck for the present with readout from the letter level.

Another decision which we adopted in order to keep the model within bounds was to exclude the possibility of processing interactions between the visual and phonological systems. However, in the model as sketched at the outset (Figure 1), activations at the letter level interacted with a phonological level as well as the word level. As we will show in Part II, some of our Context Enhancement results with pseudowords are difficult to account for in the simplified framework applied in Part I. To accommodate the findings, it may be appropriate to incorporate interactions between the letter level and the phoneme level.

Another simplification we have adopted in Part I has been to consider only cases in which individual letters or strings of letters were presented in the absence of linguistic context. In Part II we will consider the effects of introducing contextual inputs to the word level, and we will explore how the model might work in processing spoken words in context as well.

Thus far we have commented in this discussion on the completeness of the interactive activation model to account for the data in the literature on word perception and related domains. But the model is also interesting for reasons quite apart from its success in accounting for the data obtained in particular experiments. It also illustrates the operation of a kind of mechanism which we believe deserves further exploration, not only for word perception but for other perceptual domains and other aspects of information processing as well. Our various simulations show a number of different ways an activation mechanism can be used to process information. It can fill in missing information in familiar words. It can act as a sharply tuned filter, focusing activation on a single word consistent with all of the information presented. Or it can synthesize novel percepts, making use of feedback from a number of partially relevant partial activations. In Part II we will consider a few of the ways such a mechanism might be used in such diverse tasks as categorization, memory search, and retrieval.

References

Adams, M. J. Models of word recognition. Cognitive Psychology, 1979, 11, 133-176.

Aderman, D., & Smith, E. E. Expectancy as a determinant of functional units in perceptual recognition. Cognitive Psychology, 1971, 2, 117-129.

Anderson, J. A. Neural models with cognitive implications. In D. LaBerge & S. J. Samuels (Eds), Basic processes in reading: Perception and comprehension. Hillsdale, N.J.: Erlbaum Associates, 1977.

Anderson, J.A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. Distinctive features, categorical perception, and probability learning: Some applications of a neural model. Psychological Review, 1977, 84, 413-451.

Baron, J., & Thurston, I. An analysis of the word-superiority effect. Cognitive Psychology, 1973, 4, 207-228.

Bouwhuis, D. G. Visual recognition of words. Eindhoven: Greve Offset B. V., 1979.

Broadbent, D. E. Word-frequency effect and response bias. Psychological Review, 1967, 74, 1-15.

Broadbent, D. E. & Gregory, M. Visual perception of words differing in letter digram frequency. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 569-571.

Bruner, J. S. On perceptual readiness. Psychological Review, 1957, 64, 123-152.

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.

Eriksen, B. A., & Eriksen, C. W. Effects of noise letters upon the identification of a target letter in a non-search task. Perception and Psychophysics, [illegible], [illegible], [illegible].

Estes, W. The locus of inferential and perceptual processes in letter identification. Journal of Experimental Psychology: General, [illegible], [illegible], [illegible]-[illegible].

Glushko, R. The psychology of phonography: Reading aloud by orthographic activation and phonological synthesis. Unpublished doctoral dissertation,

University of California, San Diego, 1979.

Grossberg, S. A theory of visual coding, memory, and development. In Leeuwenberg, E. L. J.& Buffart, H. F. J. M. (Eds.), Formal theories of visual perception. New York: John Wiley and Sons, 1978.

Havens, L. L. & Foote, W. E. The effect of competition on visual duration threshold and its independence of stimulus frequency. Journal of Experimental Psychology, 1963, 65, 5-11.

Hinton, G. E. Relaxation and its role in vision. Ph. D. Thesis, University of Edinburgh, 1977.

Hinton, G. E. & Anderson, J. A. (Eds.) Parallel models of associative memory. Hillsdale, N. J.: Lawrence Erlbaum Associates, in press.

Holender, D. Identification of letters in words and of single letters with pre- and postknowledge vs. postknowledge of the alternatives. Perception and Psychophysics, 1979, 25, 213-318.

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.

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, 123-154.

Johnston, J. C., & McClelland, J. L. Visual factors in word perception. Perception and 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.

Johnston, J. C. & McClelland, J. L. Experimental tests of a hierarchical model of word identification. Journal of Verbal Learning and Verbal Behavior, in press.

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.

Kohonen, T. Associative memory: A system-theoretic approach. Berlin: Springer-Verlag, 1977.

Kucera, H., & Francis, W. Computational analysis of present-day American English. Providence, R. I.: Brown University Press, 1967.

LaBerge, D., & Samuels, S. Toward a theory of automatic information processing in reading. Cognitive Psychology, 1974, 6, 293-323.

Levin, J. A. Proteus: An activation framework for cognitive process models (ISI/WP-2). Marina del Rey, California: Information Sciences Institute, 1976.

Levin, J. A., & Eisenstadt, M. Proteus: A control framework for processing. Unpublished manuscript, 1975.

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

Manelis, L. The effect of meaningfulness in tachistoscopic word perception. Perception and Psychophysics, 1974, 16, 182-192.

Massaro, D. W. & Klitzke, D. The role of lateral masking and orthographic structure in letter and word recognition. Acta Psychologica, 1979, 43, 413-426.

McClelland, J. Preliminary letter identification in the perception of words and nonwords. Journal of Experimental Psychology: Human Perception and Performance, 1976, 1, 80-91.

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., and Johnston, J. The role of familiar units in perception of words and nonwords. Perception and Psychophysics, 1977, 22, 249-261.

Morton, J. Interaction of information in word recognition. Psychological Review, 1969, 76, 165-178.

Neisser, U. Cognitive psychology. New York: Appleton-Century-Crofts, 1967.

Newbigging, P. L. The perceptual redintegration of frequent and infrequent words. Canadian Journal of Psychology, 1961, 15, 123-132.

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 the perception of briefly exposed visual displays. Journal of Mathematical Psychology, 1970, 7, 191-218.

Rumelhart, D. E. A multicomponent theory of confusion among briefly exposed alphabetic characters (Tech. Rep. 22). San Diego, Ca.: University of California, San Diego, Center for Human Information Processing, 1971.

Rumelhart, David E. Toward an Interactive Model of Reading. In S. Dornic

(Ed.), Attention and Performance VI. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Rumelhart, D. E., & Siple, P. The process of recognizing tachistoscopically presented words. Psychological Review, 1974, 81, 99-118.

Spoehr, K., and Smith, E. The role of orthographic and phonotactic rules in perceiving letter patterns. Journal of Experimental Psychology: Human Perception and Performance, 1975, 1, 21-34.

Szentagothai, J. & Arbib, M. A. Conceptual models of neural organization. Cambridge, Mass.: MIT Press, 1975.

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.

Turvey, M. On peripheral and central processes in vision: Inferences from an information-processing analysis of masking with patterned stimuli. Psychological Review, 1973, 80, 1-52.

Wheeler, D. Processes in word recognition. Cognitive Psychology, 1870, 1, 59-85

Navy

1 Dr. Arthur Bachrach
Environmental Stress Program Center
Naval Medical Research Institute
Bethesda, MD 20014

1 CDR Thomas Berghage
Naval Health Research Center
San Diego, CA 92152

1 Dr. Robert Blanchard
Navy Personnel R&D Center
Managment Support Department
San Diego, CA 92151

1 Dr. Jack R. Borsting
Provost & Academic Dean
U.S. Naval Postgraduate School
Monterey, CA 93940

1 Dr. Robert Breaux
Code N-711
NAVTRAEQUIPCEN
Orlando, FL 32813

1 Chief of Naval Education and Training
Liason Office
Air Force Human Resource Laboratory
Flying Training Division
WILLIAMS AFB, AZ 85224

1 DR. PAT FEDERICO
NAVY PERSONNEL R&D CENTER
SAN DIEGO, CA 92152

1 Dr. John Ford
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Richard Gibson
Bureau of medicine and surgery
Code 3C13
Navy Department
Washington, DC 20372

1 LT Steven D. Harris, MSC, USN
Code 6021
Naval Air Development Center
Warminster, Pennsylvania 18974

1 Commanding Officer
U.S. Naval Amphibious School
Coronado, CA 92155

1 Library
Naval Health Research Center
P. O. Box 85122
San Diego, CA 92138

1 Naval Medical R&D Command
Code 44
National Naval Medical Center
Betheada, MD 20014

1 Ted M. I. Yellen
Technical Information Office, Code 201
NAVY PERSONNEL R&D CENTER
SAN DIEGO, CA 92152

1 Library, Code P201L
Navy Personnel R&D Center
San Diego, CA 92152

1 Technical Director
Navy Personnel R&D Center
San Diego, CA 92152

6 Commanding Officer
Naval Research Laboratory
Code 2627
Washington, DC 20390

1 Psychologist
ONR Branch Office
Bldg 114, Section D
666 Summer Street
Boston, MA 02210

1 Psychologist
ONR Branch Office
536 S. Clark Street
Chicago, IL 60605

1 Office of Naval Research
Code 437
800 N. Quincy Street
Arlington, VA 22217

1 Mr. Arnold Rubenstein
Naval Personnel Support Technology
Naval Material Command (08T244)
Room 1044, Crystal Plaza #5
2221 Jefferson Davis Highway
Arlington, VA 20360

1 Dr. Worth Scanland
Chief of Naval Education and Training
Code N-5
NAS, Pensacola, FL 32508

1 Dr. Sam Schiflett, SY 721
Systems Engineering Test Directorate
U.S. Naval Air Test Center
Patuxent River, MD 20670

1 Dr. Robert G. Smith
Office of Chief of Naval Operations
OP-987H
Washington, DC 20350

1 Dr. Alfred F. Smode
Training Analysis & Evaluation Group
(TAEG)
Dept. of the Navy
Orlando, FL 32813

1 Dr. Richard Sorensen
Navy Personnel R&D Center
San Diego, CA 92152

1 W. Gary Thomson
Naval Ocean Systems Center
Code 7132
San Diego, CA 92152

1 Dr. Robert Wisher
Code 309
Navy Personnel R&D Center
San Diego, CA 92152

1 Mr John H. Wolfe
Code P310
U. S. Navy Personnel Research and
Development Center
San Diego, CA 92152

1 Dr. Patrick R. Harrison
Psychology Course Director
LEADERSHIP & LAW DEPT. (7b)
DIV. OF PROFESSIONAL DEVELOPMENT
U.S. NAVAL ACADEMY
ANNAPOLIS, MD 21402

1 Dr. Lloyd Hitchcock
Human Factors Engineering
Division (6022)
Naval Air Development Center
Warminster, PA 18974

1 CDR Charles W. Hutchins
Naval Air Systems Command Hq
AIR-340F
Navy Department
Washington, DC 20361

1 Dr. Norman J. Kerr
Chief of Naval Technical Training
Naval Air Station Memphis (75)
Millington, TN 38054

1 Dr. William L. Maloy
Principal Civilian Advisor for
Education and Training
Naval Training Command, Code 00A
Pensacola, FL 32508

1 Dr. Kneale Marshall
Scientific Advisor to DCNO(MPT)
OP01T
Washington DC 20370

1 CAPT Richard L. Martin, USN
Prospective Commanding Officer
USS Carl Vinson (CVN-70)
Newport News Shipbuilding and Drydock Co
Newport News, VA 23607

1 Dr. George Moeller
Head, Human Factors Dept.
Naval Submarine Medical Research Lab
Groton, CN 06340

1 Dr William Montague
Navy Personnel R&D Center
San Diego, CA 92152

1 Office of Naval Research
Code 441
800 N. Quincy Street
Arlington, VA 22217

5 Personnel & Training Research Programs
(Code 458)
Office of Naval Research
Arlington, VA 22217

1 Psychologist
ONR Branch Office
1030 East Green Street
Pasadena, CA 91101

1 Office of the Chief of Naval Operations
Research, Development, and Studies Branc
(OP-102)
Washington, DC 20350

1 Captain Donald F. Parker, USN
Commanding Officer
Navy Personnel R&D Center
San Diego, CA 92152

1 LT Frank C. Petho, MSC, USN (Ph.D)
Code L51
Naval Aerospace Medical Research Laborat
Pensacola, FL 32508

1 DR. RICHARD A. POLLAK
ACADEMIC COMPUTING CENTER
U.S. NAVAL ACADEMY
ANNAPOLIS, MD 21402

1 Dr. Gary Poock
Operations Research Department
Code 55PK
Naval Postgraduate School
Monterey, CA 93940

1 Roger W. Remington, Ph.D
Code L52
NAMRL
Pensacola, FL 32508

1 Dr. Bernard Rimland (03B)
Navy Personnel R&D Center
San Diego, CA 92152

Army

1 LTC Michael Plummer
Chief, Leadership & Organizational Effectiveness Division
Office of the Deputy Chief of Staff for Personnel
Dept. of the Army
Pentagon, Washington DC 20301

1 Dr. Robert Sasmor
U. S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Technical Director
U. S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 HQ USAREUE & 7th Army
ODCSOPS
USAAREUE Director of GED
APO New York 09403

1 Mr. J. Barber
HQS, Department of the Army
DAPE-ZBR
Washington, DC 20310

1 DR. RALPH DUSEK
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 Col Frank Hart
Army Research Institute for the Behavioral & Social Sciences
5001 Eisenhower Blvd.
Alexandria, VA 22333

1 Dr. Michael Kaplan
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 Dr. Milton S. Katz
Training Technical Area
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Director
U.S. Army Human Engineering Labs
Attn: DRXHE-DB
Aberdeen Proving Ground, MD 21005

1 Dr. Harold F. O'Neil, Jr.
Attn: PERI-OK
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Air Force

1 Air University Library
AUL/LSE 76/443
Maxwell AFB, AL 36112

1 Dr. Earl A. Alluisi
HQ, AFHRL (AFSC)
Brooks AFB, TX 78235

1 DR. T. E. COTTERMAN
AFHRL/ASR
WRIGHT PATTERSON AFB
OHIO 45433

1 Dr. Genevieve Haddad
Program Manager
Life Sciences Directorate
AFOSR
Bolling AFB, DC 20332

1 Dr. Ronald G. Hughes
AFHRL/OTR
Williams AFB, AZ 85224

1 Dr. Ross L. Morgan (AFHRL/LR)
Wright -Patterson AFB
Ohio 45433

1 Dr. Marty Rockway (AFHRL/TT)
Lowry AFB
Colorado 80230

1 Dr. Frank Schufletowski
U.S. Air Force
ATC/XPTD
Randolph AFB, TX 78148

2 3700 TCHTW/TTGH Stop 32
Sheppard AFB, TX 76311

1 Jack A. Thorpe, Maj., USAF
Naval War College
Providence, RI 02846

Marines

1 H. William Greenup
Education Advisor (E031)
Education Center, MCDEC
Quantico, VA 22134

1 Major Howard Langdon
Headquarters, Marine Corps
OTTI 31
Arlington Annex
Columbia Pike at Arlington Ridge Rd.
Arlington, VA 20380

1 Special Assistant for Marine Corps Matters
Code 100M
Office of Naval Research
800 N. Quincy St.
Arlington, VA 22217

1 DR. A.L. SLAFKOSKY
SCIENTIFIC ADVISOR (CODE RD-1)
HQ, U.S. MARINE CORPS
WASHINGTON, DC 20380

CoastGuard

1 Chief, Psychological Reserch Branch
U. S. Coast Guard (G-P-1/2/TP42)
Washington, DC 20593

Other DoD

12 Defense Documentation Center
Cameron Station, Bldg. 5
Alexandria, VA 22314
Attn: TC

1 Dr. Craig I. Fields
Advanced Research Projects Agency
1400 Wilson Blvd.
Arlington, VA 22209

1 Dr. Dexter Fletcher
ADVANCED RESEARCH PROJECTS AGENCY
1400 WILSON BLVD.
ARLINGTON, VA 22209

1 Military Assistant for Training and Personnel Technology
Office of the Under Secretary of Defense for Research & Engineering
Room 3D129, The Pentagon
Washington, DC 20301

Civil Govt

1 Dr. Joseph L. Young, Director
Memory & Cognitive Processes
National Science Foundation
Washington, DC 20550

1 Dr. Susan Chipman
Learning and Development
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Mr. James M. Ferstl
Bureau of Training
U.S. Civil Service Commission
Washington, D.C. 20415

1 Dr. Joseph I. Lipson
SEDR W-638
National Science Foundation
Washington, DC 20550

1 Dr. John Mays
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 William J. McLaurin
Rm. 301, Internal Revenue Service
2221 Jefferson Davis Highway
Arlington, VA 22202

1 Dr. Arthur Melmed
National Intitute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. Andrew R. Molnar
Science Education Dev. and Research
National Science Foundation
Washington, DC 20550

1 Dr. H. Wallace Sinaiko
Program Director
Manpower Research and Advisory Services
Smithsonian Institution
801 North Pitt Street
Alexandria, VA 22314

1 Dr. Frank Withrow
U. S. Office of Education
400 Maryland Ave. SW
Washington, DC 20202

Non Govt

1 Ms. Carole A. Bagley
Minnesota Educational Computing Consortium
2354 Hidden Valley Lane
Stillwater, MN 55082

1 Mr Avron Barr
Department of Computer Science
Stanford University
Stanford, CA 94305

1 Dr. Jackson Beatty
Department of Psychology
University of California
Los Angeles, CA 90024

1 Dr. John Bergan
School of Education
University of Arizona
Tuscon AZ 85721

1 Dr. Nicholas A. Bond
Dept. of Psychology
Sacramento State College
600 Jay Street
Sacramento, CA 95819

1 Dr. Lyle Bourne
Department of Psychology
University of Colorado
Boulder, CO 80309

1 Dr. Kenneth Bowles
Institute for Information Sciences
University of California at San Diego
La Jolla, CA 92037

1 Dr. John S. Brown
XEROX Palo Alto Research Center
3333 Coyote Road
Palo Alto, CA 94304

1 Dr. Bruce Buchanan
Department of Computer Science
Stanford University
Stanford, CA 94305

1 Dr. Lynn A. Cooper
Department of psychology
Uris Hall
Cornell University
Ithaca, NY 14850

1 Thomas L. Crandell
35 Leslie Avenue
Conklin, NY 13748

1 Dr. Meredith P. Crawford
American Psychological Association
1200 17th Street, N.W.
Washington, DC 20036

1 Dr. Kenneth B. Cross
Anacapa Sciences, Inc.
P.O. Drawer Q
Santa Barbara, CA 93102

1 Dr. Emmanuel Donchin
Department of Psychology
University of Illinois
Champaign, IL 61820

1 Dr. Hubert Dreyfus
Department of Philosophy
University of California
Berkely, CA 94720

1 LCOL J. C. Eggenberger
DIRECTORATE OF PERSONNEL APPLIED RESEARC
NATIONAL DEFENCE HQ
101 COLONEL BY DRIVE
OTTAWA, CANADA K1A 0K2

1 ERIC Facility-Acquisitions
4833 Rugby Avenue
Bethesda, MD 20014

1 Dr. A. J. Eschenbrenner
Dept. E422, Bldg. 81
McDonnell Douglas Astronautics Co.
P.O.Box 516
St. Louis, MO 63166

1 Dr. Marvin D. Glock
217 Stone Hall
Cornell University
Ithaca, NY 14853

1 Dr. Frank E. Gomer
McDonnell Douglas Astronautics Co.
P. O. Box 516
St. Louis, MO 63166

1 Dr. Daniel Gopher
Industrial & Management Engineering
Technion-Israel Institute of Technology
Haifa
ISRAEL

1 DR. JAMES G. GREENO
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 Dr. Harold Hawkins
Department of Psychology
University of Oregon
Eugene OR 97403

1 Dr. Barbara Hayes-Roth
The Rand Corporation
1700 Main Street
Santa Monica, CA 90406

1 Dr. Frederick Hayes-Roth
The Rand Corporation
1700 Main Street
Santa Monica, CA 90406

1 Dr. Dustin H. Heuston
Wicat, Inc.
Box 986
Orem, UT 84057

1 Dr. James R. Hoffman
Department of Psychology
University of Delaware
Newark, DE 19711

1 Dr. Michael Levine
210 Education Building
University of Illinois
Champaign, IL 61820

1 Dr. Mark Miller
Computer Science Laboratory
Texas Instruments, Inc.
Mail Station 371, P.O. Box 225936
Dallas, TX 75265

1 Dr. Allen Munro
Behavioral Technology Laboratories
1845 Elena Ave., Fourth Floor
Redondo Beach, CA 90277

1 Dr. Seymour A. Papert
Massachusetts Institute of Technology
Artificial Intelligence Lab
545 Technology Square
Cambridge, MA 02139

1 Dr. James A. Paulson
Portland State University
P.O. Box 751
Portland, OR 97207

1 MR. LUIGI PETRULLO
2431 N. EDGEWOOD STREET
ARLINGTON, VA 22207

1 Dr. Martha Polson
Department of Psychology
University of Colorado
Boulder, CO 80302

1 DR. PETER POLSON
DEPT. OF PSYCHOLOGY
UNIVERSITY OF COLORADO
BOULDER, CO 80309

1 DR. DIANE M. RAMSEY-KLEE
R-K RESEARCH & SYSTEM DESIGN
3947 RIDGEMONT DRIVE
MALIBU, CA 90265

1 Dr. Robert Smith
Department of Computer Science
Rutgers University
New Brunswick, NJ 08903

1 Dr. Richard Snow
School of Education
Stanford University
Stanford, CA 94305

1 Dr. Kathryn T. Spoehr
Department of Psychology
Brown University
Providence, RI 02912

1 Dr. Robert Sternberg
Dept. of Psychology
Yale University
Box 11A, Yale Station
New Haven, CT 06520

1 DR. ALBERT STEVENS
BOLT BERANEK & NEWMAN, INC.
50 MOULTON STREET
CAMBRIDGE, MA 02138

1 Dr. David Stone
ED 236
SUNY, Albany
Albany, NY 12222

1 DR. PATRICK SUPPES
INSTITUTE FOR MATHEMATICAL STUDIES IN
THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CA 94305

1 Dr. Kikumi Tatsuoka
Computer Based Education Research
Laboratory
252 Engineering Research Laboratory
University of Illinois
Urbana, IL 61801

1 Dr. John Thomas
IBM Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

1 Dr. J. Arthur Woodward
Department of Psychology
University of California
Los Angeles, CA 90024

1 Dr. Karl Zinn
Center for research on Learning
and Teaching
University of Michigan
Ann Arbor, MI 48104

1 Dr. John R. Anderson
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

1 Thomas H. Anderson, Ph.D.
Center for the Study of Reading
174 Children's Research Center
51 Gerty Drive
Champiagn, IL 61820

1 Dr. John Annett
Department of Psychology
University of Warwick
Coventry CV4 7AL
ENGLAND

1 DR. MICHAEL ATWOOD
SCIENCE APPLICATIONS INSTITUTE
40 DENVER TECH. CENTER WEST
7935 E. PRENTICE AVENUE
ENGLEWOOD, CO 80110

1 1 psychological research unit
Dept. of Defense (Army Office)
Campbell Park Offices
Canberra ACT 2600, Australia

1 Dr. R. A. Avner
University of Illinois
Computer-Based Educational Research Lab
Urbana, IL 61801

1 Dr. Alan Baddeley
Medical Research Council
Applied Psychology Unit
15 Chaucer Road
Cambridge CB2 2EF
ENGLAND

1 Dr. Patricia Baggett
Department of Psychology
University of Denver
University Park
Denver, CO 80208

1 DR. C. VICTOR BUNDERSON
WICAT INC.
UNIVERSITY PLAZA, SUITE 10
1160 SO. STATE ST.
OREM, UT 84057

1 Dr. Anthony Cancelli
School of Education
University of Arizona
Tuscon, AZ 85721

1 Dr. Pat Carpenter
Department of Psychology
Carnegie-Mellon University
Pittsburgh, PA 15213

1 Dr. John B. Carroll
Psychometric Lab
Univ. of No. Carolina
Davie Hall 013A
Chapel Hill, NC 27514

1 Charles Myers Library
Livingstone House
Livingstone Road
Stratford
London E15 2LJ
ENGLAND

1 Dr. William Chase
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

1 Dr. Micheline Chi
Learning R & D Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15213

1 Dr. William Clancey
Department of Computer Science
Stanford University
Stanford, CA 94305

1 Dr. Allan M. Collins
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, Ma 02138

1 Dr. Ed Feigenbaum
Department of Computer Science
Stanford University
Stanford, CA 94305

1 Mr. Wallace Feurzeig
Bolt Beranek & Newman, Inc.
50 Moulton St.
Cambridge, MA 02138

1 Dr. Victor Fields
Dept. of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. Edwin A. Fleishman
Advanced Research Resources Organ.
Suite 900
4330 East West Highway
Washington, DC 20014

1 DR. JOHN D. FOLLEY JR.
APPLIED SCIENCES ASSOCIATES INC
VALENCIA, PA 16059

1 Dr. John R. Frederiksen
Bolt Beranek & Newman
50 Moulton Street
Cambridge, MA 02138

1 Dr. Alinda Friedman
Department of Psychology
University of Alberta
Edmonton, Alberta
CANADA T6G 2E9

1 Dr. R. Edward Geiselman
Department of Psychology
University of California
Los Angeles, CA 90024

1 DR. ROBERT GLASER
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 Glenda Greenwald, Ed.
"Human Intelligence Newsletter"
P. O. Box 1163
Birmingham, MI 48012

1 Dr. Earl Hunt
Dept. of Psychology
University of Washington
Seattle, WA 98105

1 Dr. Steven W. Keele
Dept. of Psychology
University of Oregon
Eugene, OR 97403

1 Dr. Walter Kintsch
Department of Psychology
University of Colorado
Boulder, CO 80302

1 Dr. David Kieras
Department of Psychology
University of Arizona
Tuscon, AZ 85721

1 Dr. Mazie Knerr
Litton-Mellonics
Box 1286
Springfield, VA 22151

1 Dr. Stephen Kosslyn
Harvard University
Department of Psychology
33 Kirkland Street
Cambridge, MA 02138

1 Mr. Marlin Kroger
1117 Via Goleta
Palos Verdes Estates, CA 90274

1 Dr. Jill Larkin
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

1 Dr. Alan Lesgold
Learning R&D Center
University of Pittsburgh
Pittsburgh, PA 15260

1 Dr. Fred Reif
SESAME
c/o Physics Department
University of California
Berkely, CA 94720

1 Dr. Andrew M. Rose
American Institutes for Research
1055 Thomas Jefferson St. NW
Washington, DC 20007

1 Dr. Ernst Z. Rothkopf
Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974

1 Dr. David Rumelhart
Center for Human Information Processing
Univ. of California, San Diego
La Jolla, CA 92093

1 DR. WALTER SCHNEIDER
DEPT. OF PSYCHOLOGY
UNIVERSITY OF ILLINOIS
CHAMPAIGN, IL 61820

1 Dr. Alan Schoenfeld
Department of Mathematics
Hamilton College
Clinton, NY 13323

1 DR. ROBERT J. SEIDEL
INSTRUCTIONAL TECHNOLOGY GROUP
HUMRRO
300 N. WASHINGTON ST.
ALEXANDRIA, VA 22314

1 Committee on Cognitive Research
% Dr. Lonnie R. Sherrod
Social Science Research Council
605 Third Avenue
New York, NY 10016

1 Robert S. Siegler
Associate Professor
Carnegie-Mellon University
Department of Psychology
Schenley Park
Pittsburgh, PA 15213

1 DR. PERRY THORNDYKE
THE RAND CORPORATION
1700 MAIN STREET
SANTA MONICA, CA 90406

1 Dr. Douglas Towne
Univ. of So. California
Behavioral Technology Labs
1845 S. Elena Ave.
Redondo Beach, CA 90277

1 Dr. Benton J. Underwood
Dept. of Psychology
Northwestern University
Evanston, IL 60201

1 Dr. Phyllis Weaver
Graduate School of Education
Harvard University
200 Larsen Hall, Appian Way
Cambridge, MA 02138

1 Dr. David J. Weiss
N660 Elliott Hall
University of Minnesota
75 E. River Road
Minneapolis, MN 55455

1 DR. GERSHON WELTMAN
PERCEPTRONICS INC.
6271 VARIEL AVE.
WOODLAND HILLS, CA 91367

Dr Keith T. Wescourt
Information Sciences Dept.
The Rand Corporation
1700 Main St.
Santa Monica, CA 90406

1 DR. SUSAN E. WHITELY
PSYCHOLOGY DEPARTMENT
UNIVERSITY OF KANSAS
LAWRENCE, KANSAS 66044

1 Dr. Christopher Wickens
Department of Psychology
University of Illinois
Champaign, Ill 61820