Featural Priming: Data Consistent With the Interactive Activation Model of Visual Word Recognition

Robert Kachelski and John Theios

Department of Psychology
University of Wisconsin
Madison, WI 53706

Abstract

The effect of featural priming on word identification was investigated as a test of the interactive activation model of word perception put forth by McClelland and Rumelhart (1981). Observers were presented with a 250 msec featural prime, which was either consistent with or inconsistent with the letters in the target word that immediately followed. Reading latencies were recorded for 96 trials per subject. A neutral prime condition consisting of a random dot pattern was used as a control in order to obtain baseline identification times. The prediction of the interactive activation model that mean reading latency would be significantly longer for words that were primed with inconsistent features than for those that were primed with consistent features was confirmed, adding to the empirical support for the model.

Introduction

The interactive activation model of visual word perception was proposed by McClelland and Rumelhart in 1981. Essentially, it was a combination of the basic assumptions of the Rumelhart (1977) interactive model with the spreading activation notions of the McClelland (1979) cascade model. The model was designed to represent the processes involved in word recognition. In the discussion of this model that follows, the initial emphasis will be on the assumptions made in its construction, and these assumptions will then be examined in light of the predictions they suggest for the present experiment.

The first main assumption of the interactive activation model is that perceptual processing takes place in a system of several levels of processing. In their original discussion, McClelland and Rumelhart (1981) described three such levels in depth: the feature level, the letter level, and the word level. Each level is made up of a set of parallel nodes, with one node representing each possible element at that level. For instance, at the feature level, there is a node for each possible feature at each letter position in a word. Similarly, there is a letter node for each letter in each position in a word, and there is a word node for each word at the word level. Higher levels of processing, which supply so-called "top-down" information to the word level, were also assumed to exist.

A second major assumption of the model is that visual perception involves parallel processing. This means not only parallelism in terms of processing information about several letters of a word simultaneously, but also in terms of processing information at all of the levels simultaneously.

The third assumption is closely related to the second. It is the idea that since processing in the visual system is massively parallel, there is a great deal of interaction both between and within the levels of processing, and this interaction ultimately determines what is perceived. In the model, there are top-down, conceptually driven aspects of processing, such as knowledge of words and orthographic structure, which occur simultaneously with bottom-up, data-driven aspects, such as featural information. The multiple constraints created by these two types of processing result in perception of a word or letter that is consistent with both (cf. Norman & Bobrow, 1975).

The interaction between the nodes discussed above is achieved through simple excitatory and inhibitory activations of a neural type (Anderson & Hinton, 1981; McClelland, Rumelhart & Hinton, 1986). Each node is connected to a number of other nodes in a bi-directional fashion. These connections may be within a given level and between adjacent levels. A connection between nodes will be excitatory if the nodes are consistent with each other; if not, the connection will be inhibitory. Every word node inhibits all other word nodes, and every letter node inhibits all other letter nodes at a particular position in a word. Connections between the word level and the letter level may be either excitatory or inhibitory, depending on whether the letter is a part of the word in that particular letter position. In the same way, the type of connections between the feature level and the letter level depend on whether or not the features present at a given letter position are consistent with the letter in question.

At any given time, each node has a specific activation level that varies continuously between a minimum and maximum level. When there is no input from neighboring nodes, a node is said to be inactive, and its activation level decays to a resting level. However, when the neighbors of a node are active, the result is either excitation or inhibition of the node, depending on the relationships among the nodes. The net input to the node is a weighted sum of the activation levels of the other nodes in the system and the connection strengths from those nodes. The net input, current activity level, resting level, and decay rate are then used to update the activity level of the node (cf. Golden, 1986; McClelland & Rumelhart, 1981).

In this model, then, when a word is presented, the parallel layer of feature nodes is activated first, presumably by feature detectors in the visual system. Activated feature nodes then send excitatory messages to letter nodes that contain those features and inhibitory messages to those that do not. Activated letter nodes can then activate consistent word nodes, and once a word node becomes activated, it begins to inhibit all other word nodes. Top-down, conceptually driven processing also occurs, since partially activated word nodes excite their component letter nodes and inhibit all other letter nodes. These interactions among the nodes at the different levels continue until a single word node gains superior activation and drives the activation of the other nodes down through massive inhibition. When this occurs, the system has essentially "recognized" the stimulus word as the word represented by the node with the highest activation value at the time of output.

When the interactive activation model was experimentally tested (Rumelhart & McClelland, 1982), the results provided a very close representation of some of the major phenomena in letter and word perception, including the way contextual inputs influence perceptual processing. In particular, Rumelhart and McClelland tested the perceptibility of a single target letter in four-letter displays. They manipulated the onset and offset times of the context letters in order to determine the extent to which they could facilitate perception of the target. Consistent with the predictions of the model, the duration and timing of the presentation of contextual information had significant effects on the perceptibility of a target letter embedded in a word.

Specifically, greater accuracy on a forced-choice task of target-letter identification was associated with longer context durations and greater numbers of letters enhanced within the context. In other words, the longer the context letters (or "primes") were present before the target letter appeared, the easier it was to identify the target letter. Similarly, the more letters that were used as a prime, the greater the percentage of correct target identifications. The finding that these context-enhancement effects were more profound when the extra contextual information came before rather than after the target letter points to the conclusion that the context must affect perception of the target letter itself as it is being processed. Presumably, this occurs because the context spreads activation to the appropriate letter node, and this activation is reinforced when the target letter appears, allowing the node to reach an activation level sufficient for letter identification more quickly than would be possible without a contextual prime.

The spreading activation notions central to the interactive activation model of McClelland and Rumelhart (1981) are by no means new, and they are certainly not exclusive to this model. Ever since Quillian introduced his theory of spreading activation (Quillian, 1962, 1967), studies have been conducted to assess whether this concept is a useful one in building a model of human perceptual processing (e.g. Collins & Quillian, 1969; Meyer & Schvaneveldt, 1971; Ratcliff & McKoon, 1981). Based on the results of such experimentation, the idea of spreading activation has become widely popular, forming the central mechanism in a number of models of memory and information processing (e.g. Anderson, 1976; Collins & Loftus, 1975; Dell & Reich, 1977; Hinton, 1981).

A large portion of the research on spreading activation has centered on semantic priming. Significant semantic priming effects have been reported by a number of researchers (e.g. Collins & Quillian, 1969; Meyer, 1970; Meyer & Schvaneveldt, 1971). This research supports the notion that memory storage is structured so that associated words and concepts can activate each other. Letter recognition tasks have also been used to study spreading activation. In addition to the experiments mentioned earlier that were conducted by Rumelhart & McClelland (1982), other research has shown effects of orthographic structure on letter perception that are consistent with the top-down processing ideas in the interactive activation model (Massaro & Klitzke, 1979; Spoehr & Smith, 1975).

As one can see from the discussion above, research evaluating the notion of spreading activation, and in particular the model put forth by McClelland and Rumelhart (1981), has concentrated almost exclusively on the letter and word levels of processing. To date, little evaluation of the model has been done in terms of how well it can predict and account for the results of experimentation at the level of basic features. Since the processing that takes place at the feature level is vital to the functioning of the model, such evaluation is necessary before an assessment of the model as a whole can be made.

The present experiment is designed to test the interactive activation model by attempting to produce priming effects at the feature level. If the processing of words proceeds according to the interactive activation model, priming an observer with features that are consistent with the letters in a target word should facilitate recognition of that word. In this case, the appropriate letter and word nodes presumably would be activated sooner than if there were no prime. In the words of McClelland and Rumelhart (1981, p.382), "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."

Similarly, a prime consisting of features that are inconsistent with the letters in a target word should interfere with the recognition of the word. The model would predict that inappropriate letter and word nodes would be activated, and the appropriate letter and word nodes would be initially inhibited. The letters and words would compete with each other and, "perhaps no single set of letters or single word will get enough activation to dominate the others. In this case the various active units might strangle each other through mutual inhibition" (McClelland & Rumelhart, 1981, p. 382).

In terms of reading latencies, then, the model would predict that consistent featural primes would result in shorter reading latencies than neutral primes, and that inconsistent featural primes would result in longer reading latencies than neutral primes.

Method

Observers

The observers were 18 graduate and undergraduate students at the University of Wisconsin-Madison. All observers had normal or corrected-to-normal vision.

Apparatus

The stimuli were presented using a Scientific Prototype model 800F two-channel tachistoscope with a Psionix, Inc. timer (model 1248B) and a LaFayette Instrument voice-activated relay. The sensitivity of the voice-activated relay was set at a level that was effectively activated by human speech, but was sufficiently insensitive to extraneous noise so that there were no inappropriate activations. Observers spoke into a Wollensak 3M microphone that was placed in a stand about 30 cm from their mouths.

Materials

Ninety-six words were chosen for this experiment. The words used were all four to six letters long. All of the words chosen were relatively common (25 appearances or more in print per million words). The words were chosen such that most of their letters had predominantly vertical, horizontal, slanted, or curved visual features, for example: HELP, THERE, MANY, or DROOP, respectively.

Each of the target words was printed on a card measuring 13.5 cm wide by 24 cm high in order to be presented in the tachistoscope. These stimulus items were generated by a Macintosh Plus computer with an Apple laser printer, using Chicago font. The letters in each word were in upper case and were 1.1 cm wide by 1.2 cm high. When the words were presented in the tachistoscope, each letter subtended .84 degrees of visual angle in width and .96 degrees of visual angle in height.

Stimulus cards to be used as primes in the experiment were constructed for each word. A prime card for a particular word consisted of features consistent with the letters of that word in their appropriate spatial position. For example, the prime for the word "HELP" would consist of the five vertical features of the four letters, each in their correct position. There were four general types of primes: vertical, horizontal, slanted, and curved, corresponding to the words with these predominant features.

A "neutral" prime card was also constructed. It consisted of a random dot pattern of the same dimensions as the featural primes on the other cards.

Procedure

Observers were first given a vision test in order to ensure that only those with normal vision were included in the experiment. They were then instructed to read each word that was presented to them out loud as quickly as possible without making errors. Observers were given six practice trials in order to eliminate any confusion and to orient them to the task. The 96 experimental trials were then presented, with a five-minute break at the halfway point.

For each trial, the experimenter gave the observer a verbal ready signal and then presented the stimuli. The prime was presented for 250 msec and was immediately followed by the target word, which remained on until the observer read the word, thus activating the voice relay. There were three types of primes: (a) consistent = the prime created for that particular target word using its predominant features (all features being either vertical, horizontal, slanted, or curved), (b) inconsistent = a prime whose features did not match the predominant features of the target word, and (c) neutral = the random dot pattern. Reading latencies were recorded in milliseconds for each trial.

The 96 trials were presented in a different random sequence to each observer. Both word order and priming condition (consistent prime, inconsistent prime, or neutral prime) were randomized, with the constraint that there was an equal number of trials with each type of prime. Observers saw each target word only once.

Results

When all of the data were collected, a single-factor, within-subjects ANOVA was conducted on the reading latencies. The result was a significant effect for type of prime [$F(2,34) = 3.18$, $p<.05$]. The mean reading latencies for the words in each priming condition were: consistent prime = 556 msec, neutral prime = 561 msec, and inconsistent prime = 567 msec. A planned comparison between the means of the inconsistent and consistent priming conditions revealed that mean reading latency for words that were primed inconsistently were significantly longer than the mean latency for words that were primed consistently [$F(1,17) = 5.45$, $p<.05$]. The mean latency for the neutral prime condition was not significantly different from either the mean latency for the inconsistent prime condition [$F(1,17) = 1.17$, $p>.05$] or the mean latency for the consistent prime condition [$F(1,17) = 1.12$, $p>.05$].

To examine the data more closely, separate sign tests were conducted between the neutral prime condition and each of the other two priming conditions. It was found that 13 of the 18 observers had shorter mean reading latencies in the consistent priming condition than in the neutral priming condition. This result was significant ($p<.05$). For the sign test comparing the neutral prime condition to the inconsistent prime condition, 11 of the 18 observers had shorter mean reading latencies in the neutral prime condition than in the inconsistent prime condition. This result was not significant ($p=.24$).

Discussion

The finding that the mean reading latency for consistently primed words was significantly shorter than the mean latency for inconsistently primed words is supportive of the interactive activation model's representation of the role of the feature level of processing in visual word recognition. It appears that having accurate information about the features of the letters in a word before the word is actually presented gives the appropriate letter nodes and the target-word node a "head start" in terms of activation as compared to a situation in which the featural information is inconsistent with the letters in the target word. On the other hand, when the featural information is inconsistent with the target word, the relevant letter nodes and the target-word node itself tend to be initially inhibited, resulting in longer word recognition times. This result fits in nicely with the research findings on priming at the letter and word levels of processing. It lends support to the idea of spreading activation in human information processing, and in particular to the parallel-distributed-processing, connectionist formulation of the interactive activation model of visual word recognition.

It should be stated, however, that the present experiment ws not designed to provide a critical test of the interactive activation model *in comparison to* other models of word recognition. The purpose here was to test a specific prediction about the role of featural input in this model. The present results would also have been predicted by a number of other models of word recognition, such as Morton's (1969) logogen model, Smith's (1971) feature model, and Paap et al.'s (1982) activation-verification model, to name a few. Although these models are very different, they all share the assumption that analysis of the features in words is critical for successful word perception. The featural priming effects reported here should be viewed as supportive of this assumption in general. On the other hand, the present data could not be accounted for explanations of word recogntion that rely on "holistic" or global features of words, such as word "envelope" or shape (e.g. Haber, Haber & Furlin, 1983; Monk & Hulme, 1983) or on the contours of adjacent letters (e.g. Wheeler, 1970). Features of only *some* of the letters were presented as primes in this experiment, so the overall word shape could not have been determined from the prime. In addition, only upper-case letters were used for the stimuli, so these words did not have the distinctive shapes, defined by ascending, descending, and neutral letters, that supposedly underlie word recognition by word shape.

Although the results of the planned comparisons revealed that the mean reading latency for the neutral prime condition was not significantly different from the mean of either the consistent or the inconsistent prime conditions, the sign test showed that the number of observers whose mean for the consistent prime condition was lower than their mean for the neutral prime condition was significantly greater than would be expected if the two types of primes had the same effect. The 5 msec difference between the overall means for the consistent prime condition and the neutral prime condition was quantitatively small, but it appeared fairly consistently across our sample of observers. Therefore, consistent featural priming seems to have a slight facilitative effect on word recognition speed as compared to even a neutral prime.

There are at least three reasons why the reading latencies in the neutral prime condition were not significantly different from those in the inconsistent prime condition. First, observers may recover rather quickly from misleading featural information in a word recognition task. The initial inhibition of appropriate feature, letter, and word nodes due to the inconsistent prime may be easily and quickly overcome once the target word appears. This may be a reflection of more emphasis being given to excitatory connections than inhibitory connections in the lower featural levels of the visual word recognition process (in contrast to the emphasis on inhibition in McClelland and Rumelhart's (1981) simulations). Second, the correct choice of a true "neutral" prime or baseline control in these types of experiments is problematic. The random dot prime itself may cause some degree of inhibition. After all, it is a type of visual noise. Thus, we may have been optimistic to assume that the random dot prime would not have an interfering effect on the perception of the target. Finally, the 250 msec exposure duration we used for the primes may have been too long to achieve maximal excitation and inhibition effects. Given Di Lollo's (1980) work on visual persistence and visual integration, a prime exposure duration of under 100 msec may be a better choice for future research on featural priming. Although further research is necessary to identify the underlying mechanisms are that are responsible for the results of the present study, it is clear that they are generally consistent with the predictions that follow from Rumelhart and McClelland's (1981) parallel-distributed-processing, interactive activation model of visual word recognition.

References

Anderson, J. A. (1976). *Language, memory, and thought*. Hillsdale, NJ: Erlbaum.

Anderson, J. A., & Hinton, G. (1981). Models of information processing in the brain. In G. Hinton & J.A. Anderson (Eds.), *Parallel models of associative memory* (pp. 9-48). Hillsdale, NJ: Erlbaum.

Collins, A., & Loftus, E. (1975). A spreading activation theory of semantic processing. *Psychological Review*, *82*, 407-428.

Collins, A., & Quillian, M. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*, 240-248.

Dell, G. S., & Reich, P. A. (1977). A model of slips of the tongue. In R. Dipietro & E. Blansitt (Eds.), *The third LACUS forum*. Columbia, SC: Hornbeam Press.

Di Lollo, V. (1980). Temporal integration in visual memory. *Journal of Experimental Psychology: General*, *109*, 75-97.

Golden, R. (1986). A developmental neural model of visual word perception. *Cognitive Science*, *10*, 241-276.

Haber, L. R., Haber, R. N., & Furlin, K. R. (1983). Word length and word shape as sources of information in reading. *Reading Research Quarterly*, *18*, 165-189.

Hinton, G. (1981). Implementing semantic networks in parallel hardware. In G. Hinton & J. A. Anderson (Eds.), *Parallel models of associative memory*. Hillsdale, JN: Erlbaum.

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

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

McClelland, J. L., & Rumelhart D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, *88*, 375-407.

McClelland, J. L., Rumelhart, D. E., & Hinton, G. E. (1986). The appeal of parallel distributed processing. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations* (pp. 3-44). Cambridge, MA: MIT Press/Bradford.

Meyer, D. E. (1970). On the representation and retrieval of stored semantic information. *Cognitive Psychology*, *1*, 242-300.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90*, 227-234.

Monk, A. F., & Hulme, C. (1983). Errors in proofreading: Evidence for the use of word shape in word recognition. *Memory & Cognition, 11*, 16-23.

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

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology, 7*, 44-64.

Paap, K. R., Newsome, S. L., McDonald, J. E., & Schvaneveldt, R. (1982). An activation-verification model for letter and word recognition: The word superiority effect. *Psychological Review, 89*, 573-594.

Quillian, M. R. (1962). A revised design for an understanding machine. *Mechanical Translation, 7*, 17-29.

Quillian, M. R. (1967). Word concepts: A theory and simulation of some basic semantic capabilities. *Behavioral Science, 12*, 410-430.

Ratcliff, R., & McKoon, G. (1981). Does activation really spread? *Psychological Review, 88*, 454-462.

Rumelhart, D. E. (1977). Toward an interactive model of reading. In S. Dornic (Ed.), *Attention and performance IV*. Hillsdale, NJ: Erlbaum.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review, 89*, 60-94.

Smith, F. (1971). *Understanding reading*. New York: Holt, Rinehart and Winston.

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

Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology, 1*, 59-85.