Visual Processes in Reading and Reading Disabilities

Edited by

Dale M. Willows
Ontario Institute for Studies in Education
Richard S. Kruk
University of Wollongong, Australia
Evelyne Corcos
Glendon College—York University, Ontario

LEA
1993
LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS
Hillsdale, New Jersey Hove and London

7

Visual Information Processing in Reading

Dominic W. Massaro
University of California, Santa Cruz
Thomas Sanocki
University of South Florida, Tampa

INTRODUCTION

The first half of the title of this book, *Visual Processes in Reading*, might seem redundant because reading is necessarily a visual act. However, reading is influenced by nonvisual information. Our concern is with visual processing, but nonvisual information (in many different forms) contributes to our perceptual experience of reading text. The goal of this chapter is to present a theoretical framework for visual information processing in reading and to sample relevant research. After presenting the model, we address the issue of what visual information is used in word identification. The results are consistent with the view that identification of a word depends on processing its component letters. Therefore, it is important to describe the visual information contributing to letter recognition. After considering letter recognition per se, we discuss a way of using visual information to increase the amount of orthographic and phonological information conveyed by letters. Then we return to the observed perceptual advantage for words and consider competing theoretical explanations. In particular, we focus on an account that emphasizes knowledge of spelling regularities (orthographic structure) and an account that utilizes specific words. We evaluate these accounts in terms of how they handle effects of processing time and other variables in word perception tasks. Finally, we close with consideration of effects of the discourse context on word recognition.

FUZZY LOGICAL MODEL OF PERCEPTION (FLMP)

Within the framework of the Fuzzy Logical Model of Perception (FLMP), visual processing in reading can be conceptualized as a pattern recognition process. Well-learned patterns are recognized in accordance with a general algorithm, regardless of the modality or particular nature of the patterns (Massaro, 1987; Oden & Massaro, 1978). The model consists of three operations: feature evaluation, feature integration, and decision. *Continuously valued features* are evaluated, integrated, and matched against prototype descriptions in memory, and an identification decision is made on the basis of the relative goodness-of-match of the stimulus information with the relevant prototype descriptions. Prototypes are generated for the task at hand. The *sensory* systems transduce the physical event and make available various sources of information called *features*. During the first operation in the model, the features are evaluated in terms of the *prototypes* in memory. For each feature and for each prototype, feature evaluation provides information about the degree to which the feature in the signal matches the corresponding feature value of the prototype.

Feature evaluation provides the degree to which each feature in the stimulus matches the corresponding feature in each prototype in memory. The goal, of course, is to determine the overall goodness-of-match of each prototype with the stimulus. All of the features are capable of contributing to this process and the second operation of the model is called *feature integration*. That is, the features (actually the degrees of matches) corresponding to each prototype are combined (or conjoined in logical terms). The outcome of feature integration consists of the degree to which each prototype matches the stimulus. In the model, all features contribute to the final value, but with the property that the least ambiguous features have the most impact on the outcome.

The third operation is *decision*. During this stage, the merit of each relevant prototype is evaluated relative to the sum of the merits of the other relevant prototypes. This relative goodness-of-match gives the proportion of times the stimulus is identified as an instance of the prototype. The relative goodness-of-match could also be determined from a rating judgment indicating the degree to which the stimulus matches the category. The three operations between presentation of a pattern and its categorization can be formalized mathematically, but this is beyond the purview of this chapter (but see Massaro & Hary, 1986; Massaro, 1987). We begin with the study of the visual features that are functional in reading words.

WORD IDENTIFICATION

If the first psychologists from a century ago were to return to life and read our journals, they would be impressed by how much the fundamental questions they proposed remain at the center of research. Some of the first studies gave a surprising result. Subjects could read words under conditions that did not permit the identification of the component letters when they were presented in random letter strings. A great deal of excitement in psychological and educational circles was generated by this work. The research results convinced many people that word recognition was not dependent on the recognition of individual letters. Researchers and educators concluded that words are learned as patterns of unique shapes rather than as unique sequences of letters. If words have unique shapes, readers would learn words in terms of relatively gross properties that define their shape. We call these properties *supraletter features* because they supposedly are composed of multiletter patterns and even whole word patterns. This belief was responsible for the whole-word method of teaching reading.

As appealing as the concept of supraletter features might be, there is no evidence for this idea (Anderson & Dearborn, 1952; Gibson & Levin, 1975; Huey, 1968). One of the strongest arguments against the idea of supraletter features is their small potential contributions to reading. Overall word shape, for example, does not sufficiently differentiate among the words of a language (Groff, 1975). There is also experimental evidence against the idea of word recognition based on supraletter features. It has been shown that words with rare shapes are not more perceptible than those with more common shapes (Paap, Newsome, & Noel, 1984). A rare shape means that this shape specifies only one or a few words, whereas a common shape is consistent with many possible words. If word shape is functional in reading, words with rare shapes should have been easier to read.

The role of supraletter features has also been evaluated in a number of studies by determining whether mixing the case (McClelland, 1976) or type fonts (Adams, 1979) of letters eliminates the tachistoscopic identification advantage of word over nonword letter strings. For example, Adams (1979) studied the tachistoscopic recognition of words, pseudowords very high in orthographic structure, and nonwords very low in structure. The items were presented in a single type font or the items were constructed from a wide variety of fonts. Performance was more accurate for words than pseudowords and poorest for nonwords. Most importantly, the size of the differences among the three types of items did not change when the letters of the items were presented in a variety of type fonts. If supraletter features or whole-word cues contribute to the perceptual advantage of well-structured strings, the advantage of the word and pseudoword strings should have been drastically attenuated in the mixed-font presentation.

LETTER RECOGNITION

Although word recognition depends on identification of component letters, several sources of nonvisual information (i.e., orthographic, lexical, and syntactic and semantic) may contribute to letter identification. Therefore, in order to study letter recognition per se, it is necessary to strip away these nonvisual sources of information and study either single letters or strings of unrelated letters. A starting point in the study of letter recognition is the fact that instances of the same letter can vary markedly, depending on the font. In general, researchers have responded to differences between instances in either of two ways.

General Features

One approach is to assume that the features used in recognition are abstract, general features that apply to varying instances of the same letter. That is, prototype descriptions for letters would be in terms of general features. For example, Gibson's (1969) well-known feature list includes general binary contrasts such as the presence or absence of "straight-vertical" and "intersection." Intersection is present in instances of A and P, and absent in C and U. Such contrasts may be important in letter recognition because they are generally applicable and easy for the visual system to detect. However, there has not been much research in the last 20 years that directly tests these features, perhaps because of the very general level at which such features are (necessarily) specified. One exception is research examining fuzzy contrasts, in which features are present or absent to varying degrees (e.g., Massaro & Hary, 1986; Oden, 1979).

Recently, the importance of binary contrasts has taken on added meaning in models of object recognition (e.g., Biederman, 1987; Lowe, 1985). For example, Biederman (1987) proposed that objects are represented in terms of their parts, termed "geons," and geons are identified by detecting a limited number of easy-to-detect binary or ternary contrasts. Biederman referred to the features as "nonaccidental properties" because they have two important characteristics: They are unlikely to occur by accident, and they remain invariant across changes in viewpoint (see Lowe, 1985; Witkin & Tenenbaum, 1983). These characteristics are important because it is necessary in object recognition to distinguish between image properties that are valid indicators of object properties and image properties that may occur because of an accidental alignment of stimulus features. For example, parallel edges (over small visual angles) are nonaccidental properties because they are much more likely to be caused by parallel edges in the object rather than be caused by accident (e.g., by accident alignment of nonparallel edges when seen from a particular viewpoint).

Font-Specific Features

A second response to differences between instances of the category is to simply focus on a single font and to assume that features correspond to local parts or global properties within that font (e.g., Keren & Baggen, 1981; Townsend & Ashby, 1982). This approach has been used more often, and models have been fit with more precision, but it is not clear that a model fit to data from one font will fit data from another font (see Gilmore, 1985). Several different types of feature sets have been considered in this research. In one study, Keren and Baggen (1981) used features derived from a multidimensional scaling method. The features ranged from global ones (e.g., "facing right") to local ones (e.g., "horizontal line in center"). In a second study, Keren and Baggen used segments (of numerals, as on a calculator) as features, supplemented by two global features ("open-left" and "open-right"). Keren and Baggen showed that, when incorporated into Tversky's (1977) contrast model, such features predicted confusion data quite well. It is not clear, however, that Keren and Baggen's features would be easily detected by the visual system. For example, the feature "line in center" cannot be resolved until spatial relations within the pattern have been established. "Facing right" cannot be established until the parts of the figure that do the facing have been detected.

Font-Specific Tuning

One issue relevant to the two approaches already mentioned is the degree to which the processes underlying recognition become "tuned" to the details that characterize particular fonts. If features are only abstract and general, then varying the font should not matter. However, if tuning occurs, then letter prototypes could become modified for the current font. As a result, letter recognition should be more accurate with target sets of *n* letters of a single font than with *n* letters of two or more fonts because tuning can be more precise with a single font. Sanocki (1987, 1988) examined the recognition of strings of (typically) four unrelated letters in consistent and mixed-font conditions. Results from both a letter–nonletter reaction time task and a backward masking task indicated that tuning occurs. Advantages for consistent font conditions have been quite large in the letter–nonletter task, which requires explicit attention to visual structure (effects typically exceed 120 ms), but smaller in the masking task, which requires only identification (effects averaged about 3%). These results indicate that recognition can use features that are specific to particular fonts. However, the rather small effects in the masking task qualified the extent to which such features contribute to recognition; more general features may be more important.

Global Features

The assumption that recognition involves both specific and general features contrasts with a simpler assumption that features correspond to mutually exclusive segments of letters. In this simpler model, each segment is detected by an independent feature mechanism. Townsend and Ashby (1982) presented letters composed from line segments and obtained reports of letter identity and of the segments perceived. They found reasonably good fits for models that assumed segment features. However, a further prediction is that, if segment-features are detected independently, then the probabilities of detecting different segments should not be correlated. Townsend and Ashby found that the probability of detecting a given segment increased with the probability of detecting other segments in the letter (see also, Townsend, Hu, & Evans, 1984). These results appear to be inconsistent with the simple-segment model. One way to explain this result would be to assume that, in the initial stages of the letter recognition process, only global features (e.g., overall letter shape) have been extracted. Segments might then be perceived more accurately when they are contained within accurately perceived global features.

Townsend, Hu, and Kadlec (1988) replicated the intersegment correlations while finding moderately strong support for more detailed predictions of a global-to-local model of letter recognition. In that study, they also found evidence against explanations of the correlations in terms of a periodic waxing and waning of attention (i.e., more or fewer segments are perceived as alertness or attention gradually increases or decreases). However, their data do not rule out an explanation in terms of random trial-to-trial fluctuations in attention. That is, a subject's overall performance might fluctuate from trial to trial because of a variety of factors and this variability might be sufficient to produce a positive correlation among feature-detection probabilities. Further evidence against simple segment-feature models stems from the fact that when Keren and Baggen (1981) used segment features, they found it necessary to supplement the segment features with two global features. Unfortunately, they did not report results for a segment-features-only model. The importance of global features was demonstrated some time ago in Bouma's (1971) study of confusions between lowercase letters. Bouma found that letters with similar envelopes (enclosing outlines) were more likely to be confused.

Some global features may correspond to lower spatial frequency components of the stimulus. Gervais, Harvey, and Roberts (1984) analyzed letters into their spatial frequency components and then filtered the results to attenuate higher spatial frequency components (the filter was based on contrast sensitivity functions). The resultant "internal representations" were then used to predict confusions and did so with moderately high success (and with more success than a template or feature-based model). However, it is important to note that although models such as Gervais et al.'s have few free parameters, the internal representations assumed are quite complex. For example, the frequency analysis for each letter used as input a 50 × 45 cell array (2,250 cells in total) and the results of the analysis were represented in arrays of (apparently) similar complexity. In contrast, feature-based representations were simpler, consisting of trinary values on a small set of features.

Although global features in the low pass stimuli of Gervais et al. are not detailed, other global features could involve more detailed relations between parts of letters. Sanocki (1991a) studied the importance of relations between letter parts by measuring the perceptibility of backward-masked letters from four fonts. Sanocki used two different size "normal" fonts (one twice as large as the other), in which the relations between ascender or descender size and body size were typical of that found in most common text fonts (as established in a normative study). Two "abnormal" fonts were then created by interchanging ascenders or descenders and bodies between the two normal fonts. If letter perceptibility was determined by the discriminability of only segment features, then the average perceptibilities of the normal and abnormal pairs of fonts should be equal because they are composed of the same segments. On the other hand, if features more complex than segments are used in recognition, then there should be an advantage for the normal fonts, which have relations more typical of normal letters. Overall accuracy was higher with the normal fonts than with the abnormal fonts, consistent with the claim that relations between segments are important in letter recognition. There was an advantage for both the large and small normal fonts relative to the abnormal fonts (taken as a whole), so the results cannot be explained in terms of overall letter size or some other correlational feature.

Sanocki's (1991a) results are contrary to the impression of letter recognition given in many textbooks. Many texts contain a figure showing a wide variety of different font instances of a letter, all (or most) of which can be recognized. This demonstration is then taken as evidence that letter recognition is a general process that can handle a variety of instances. However, the differences obtained by Sanocki between normal and abnormal letters indicate that instances are not equally recognizable. Less typical patterns are more difficult to recognize, perhaps because they provide relatively poor matches for details of internal prototypes developed through years of experience.

The previous data indicate that to some extent recognition may rely on detailed feature information. This might seem contrary to models assuming general contrastive features such as those of Gibson and Biederman. However, further consideration of object recognition also implicates more specific features. Although Biederman (1987) emphasized simple nonaccidental contrasts such as "straight versus curved," such properties do not seem to provide sufficiently strong constraints for object recognition. Many objects have multiple geons and many edges, and any given edge can participate in many relations with neighboring edges. Simple contrasts such as "straight (vs. curved)" would not constrain the assignment of edges to geons, meaning that an edge may be wrongly combined with edges from other geons. On the other hand, somewhat more complex, relational properties such as vertice type (e.g., arrow vertice or fork vertice) put stronger constraints on geon identity because they can participate in only a limited number of relations with neighboring edges. Thus, in more recent computational work, vertice-types have played important roles (Hummel & Biederman, 1992). Consistent with the apparent importance of vertice-types, Enns and Rensink (1990, 1991) studied them in search tasks and obtained evidence that they are detected easily and in parallel ("pre-attentively visible"; Treisman &

Gormican, 1988). Furthermore, consistent with the importance of details of relations, Enns and Rensink found that vertices with less typical details (e.g., 60 rather than 90 degree angles) are not preattentively visible.

Further data on object recognition indicate that, as in reading, the perceptual system uses multiple sources of information, with the importance of a given type of information depending on its strength. In contrast to models in which only edge information is assumed to be used, Price and Humphreys (1989) found that surface details (e.g., color and shading) also contribute to recognition. The importance of such information increases with objects from categories with many visually similar members, because edge or shape information is less distinctive within such categories.

Integrating Features

In sum, letter recognition may involve a mixture of different features: Local features corresponding to specific letter parts and relations between parts, as well as both specific and more general global features. The type of features used may depend in part on visual conditions (e.g., length of eye fixation and quality of type). Also, many theorists believe that the type of features detected might vary during the process of recognition, with more general global features being detected first, followed by increasingly precise and more local features. Several types of evidence are consistent with this claim (e.g., Krueger & Chignell, 1985; Lupker, 1979; Townsend et al., 1988), but research has only begun to directly address the constancy or variance of information processed across the time course of recognition (Navon, 1991; Sanocki, 1991b). These features must be integrated together to achieve accurate letter recognition (see Oden, 1979, and Massaro & Hary, 1986, for an empirical and theoretical analysis of how features are integrated in letter recognition).

LETTERS AND SOUNDS

The process of recognizing letters in words appears to use multiple sources of information, including information about spellings and phonology (Venezky & Massaro, 1987). The extent to which a source of information will be used depends on its informativeness; in general, the least ambiguous source will have the largest effect. For skilled readers and familiar words, the extraction of visual letter information may be so quick that other types of information will not have an influence. However, with more difficult words or with beginning readers, nonvisual sources of information may become increasingly important. For example, phonological information resulting from the encoding of letter-to-sound correspondences may be useful for beginning readers.

Because of complexities of the English language, the spelling-to-sound translation of any given letter is often ambiguous. For example, *c* has a different sound in *cat* and *city*. Such ambiguities can be very troublesome for beginners. One way to reduce the ambiguity is to increase the amount of visual information associated with letters—that is, to modify the alphabet. For example, in the Initial Teaching Alphabet (i.t.a.; Pittman & St. John, 1969), new symbols were added to the alphabet so that each unique letter sound had its own symbol. Also, the spellings of many words were "regularized" by giving them new spellings. Although the i.t.a. was once widely used, it ran into serious problems. Most important was the fact that students who learned with the i.t.a. experienced great difficulty transferring to the traditional orthography (see, e.g., Downing, 1967).

However, it is important to note that alphabet modification follows from underlying assumptions about the reading process, and the assumptions underlying the i.t.a. are now known to be incorrect. In particular, Pittman subscribed to the aforementioned belief that words were recognized as whole units, and thus, that individual letters played little role in reading beyond preserving word shape (Pittman & St. John, 1969). Pittman allowed some sounds to be represented by new patterns that bore no resemblance to the old ones beyond shared envelope.

Quite different assumptions can be used to motivate alphabet modification. One example is the Graphophonic Alphabet (GP; Sanocki & Rose, 1989–90), in which different sounding versions of the same letter are distinguished by minor modifications of the letter's form. Because the more global properties of the letters remain unchanged, a modified letter should first be identified as an instance of the original letter. In addition, the modifications should become clear as the details of the instance are processed. The modifications correspond to how the appropriate sound is produced, and thereby provide cues to the letter's phonological status. As can be seen in Fig. 7.1, for example, the hard *c* sound is signified by a closure at the back of *c*, which corresponds to the closure in the throat used to make the sound. Soft *c* on the other hand is signified by *s*-like curls at the front of the letter, corresponding to the closure at the front of the mouth (as with *s*). Similarly, hard *g* has a closure at its back, whereas soft *g* is condensed and *j*-like.

The difficulty of transition to the traditional orthography should be minimized because the modified letters are similar to the traditional letters and because the

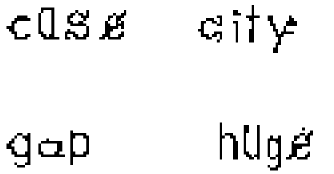

FIG. 7.1. Examples of Graphophonic Letters. In addition to the "c's" and "g's" mentioned in the text, note also the soft "s", the silent "e", the "y" on "city", and the short and long vowels (which are short or tall, respectively).

GP requires no new spellings. In addition, the modifications can be removed gradually in a fading-out process. The effectiveness of the alphabet remains to be seen (research is underway), and no claim is made that the current modifications are the best ones. The important point here is that the amount of visual information in letters can be increased to provide additional phonological and orthographic information. More generally, alphabet modification remains a potentially promising and relatively untested tool for teaching reading.

WORDS

Stimulus information about letters serves as the first source of information used in word recognition. The second is nonvisual information possessed by the sophisticated reader and stored in memory. There are three sources of nonvisual information that can aid the reader in decoding the written message. These sources are the orthographic, syntactic, and semantic structures that exist in English prose. The orthographic constraints define the valid spelling patterns in English. We know that words are separated by blank spaces and must have at least one vowel. Syntactic rules establish the permissible sequences of different parts of speech. For example, "The boy down fell the hill" is grammatically incorrect. Finally, semantic rules allow the reader to predict the word or words that make sense in a given sentence context. "The hill fell down the boy" is syntactically correct but semantically anomalous. All of these rules allow us to agree on the missing word in "Please clean the dirt from your s——s before walking inside."

In terms of the FLMP, the visual stimulus is transformed by the visual system, and letter features are evaluated. Recognition, or the interpretation of this information, depends on the integration of these features with the information possessed by the reader about the occurrences of spelling patterns in English. In Fig. 7.2, although the visual information available about the last letter of the first word is the same as the first letter of the second word, the contribution of what one knows about the valid spelling patterns in English text demands that they be interpreted as different letters. (This knowledge is sometimes referred to as redundancy, because it reduces the number of valid alternatives a particular visual configuration can possess.) In reading, we would expect that this knowledge of English spelling would enable us to extract meaning from a page of text without analyzing all the visual information present or to identify words even when some of the visual information is incomplete or fuzzy.

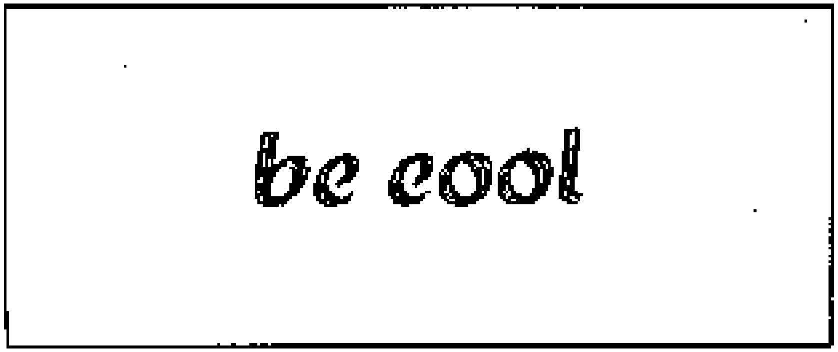

FIG. 7.2. The same visual configuration can be interpreted as two different letters, depending on the meaningful context.

Orthographic structure refers to the fact that a written language, such as English, follows certain rules of spelling. These regularities prohibit certain letter combinations and make some letters and combinations much more likely in certain positions of words than others. It is only natural that readers would use this information in letter and word perception. Concern for orthographic structure in reading has occurred only recently. An important question is the nature of a reader's knowledge about orthographic structure. It is possible to distinguish between two broad categories of orthographic structures: statistical redundancy and rule-governed regularity (Venezky & Massaro, 1987). The first category includes all descriptions derived solely from the frequency of letters and letter sequences in written texts. The second category includes all descriptions derived from the phonological constraints in English and scribal conventions for writing words as sequences of letters. Although these two descriptions are highly correlated in written English, it is possible to create letter strings that allow the descriptions to be orthogonally varied. Given these strings as test items, perceptual recognition tasks have been carried out to decide which general category seems to reflect the manner in which readers store and utilize knowledge of orthographic structure.

Massaro, Taylor, Venezky, Jastrzembski, and Lucas (1980) contrasted specific statistical-redundancy descriptions with specific rule-governed descriptions by comparing letter strings that varied orthogonally with respect to these descriptions. The statistical redundancy measures were summed token single-letter frequency, bigram frequency, and log bigram frequency. The rule-governed regularity measures were various sets of rules based on phonological and scribal constraints. In a typical experiment, six-letter words and anagrams of these words were used as test items. The anagrams were selected to give letter strings that represented the four combinations formed by a factorial arrangement of high or low frequency and regular or irregular orthographic structure. In a series of experiments utilizing a target-search task, subjects were asked to indicate whether or not a target letter was present in these letter strings. Both accuracy and reaction time measures indicated some psychological reality for both frequency and the regularity description of orthographic structure. The results of these studies provided evidence for the utilization of higher order knowledge in the perceptual processing of letter strings. Lexical status, orthographic regularity, and frequency appear to be important components of the higher order knowledge that is used (Massaro et al., 1980). In addition, Venezky and Massaro, in a reanalysis of the results of Waters and Seidenberg (1985), found that word frequency, spelling-to-sound correspondences, and orthographic regularity influence reaction time in naming and lexical decision performance.

Orthographic Knowledge versus Specific Words

There are two fundamentally different accounts of the word advantage. One type of explanation is that the subject can use knowledge of spelling during letter and word perception. In a typical trial, because of the brief duration of the display, there is only partial information about the letters in the display. However, the recognition process can use information about orthography to reduce the number of valid interpretations and arrive at a correct percept based on partial visual information. This information does not have to be consciously known or applied, and does not have to be perfect. That is, this information can still make a positive contribution to word recognition even if it provides only partial information. This explanation, within the context of the FLMP, states that readers have two sources of information given a word and only a single source given a nonword.

Another account of the word advantage dispenses with the idea of orthographic knowledge entirely, and explains the word advantage in terms of the contribution of the specific words in the reader's lexicon (Brooks, 1978; Glushko, 1979). The most complete model within this class is the interactive activation model (IAM). The model was designed to account for context effects in word perception and postulates three levels of units: features, letters, and words. Features activate letters that are consistent with the features and inhibit letters that are inconsistent; letters activate consistent words and inhibit inconsistent words; and most importantly, words activate consistent letters (top-down feedback). In addition, activated words inhibit other words. Interactive activation explains the word advantage over nonwords in terms of interactive activation from the word level to the letter level. The FLMP and IAM make very similar predictions and a more complex experiment is necessary to distinguish between them. We now describe an empirical test between these two accounts of the word advantage.

Context Effects and Backward Masking

It is possible to gradually transform the letter *c* into an *e* by extending the horizontal bar. To the extent the bar is long, there is good visual information for an *e* and poor visual information for a *c*. Now consider the letter presented as the first letter in the context *-oin* and the context *-dit*. Only *c* is orthographically admissible in the first context and only *e* is admissible in the second context. The context *-oin* favors *c*, whereas the context *-dit* favors *e*. The context *-tsa* and *-ast* can be considered to favor neither *e* or *c*. The first remains an inadmissible context whether *c* or *e* is present, and the second is orthographically admissible for both *e* and *c*. Four analogous contexts were used for presentations of the test letter in each of the other three positions of the four-letter string.

The experiment factorially combined six levels of visual information with these four levels of orthographic context, giving a total of 24 experimental conditions (Massaro, 1979). The bar length of the letter took on six values going from a prototypical *c* to a prototypical *e*. The test letter was presented at each of the four letter positions in each of the four contexts. The test string was presented for a short duration followed after some short interval by a masking stimulus composed of random letter features. Subjects were instructed to identify the test letter as *c* or *e* on the basis of what they saw. Figure 7.3 gives the probability of *e* judgments in the task. Both the test letter and the context influenced performance in the expected direction. Further, the effect of context was larger for the more ambiguous test letters along the stimulus continuum.

This study also evaluated context effects as a function of processing time

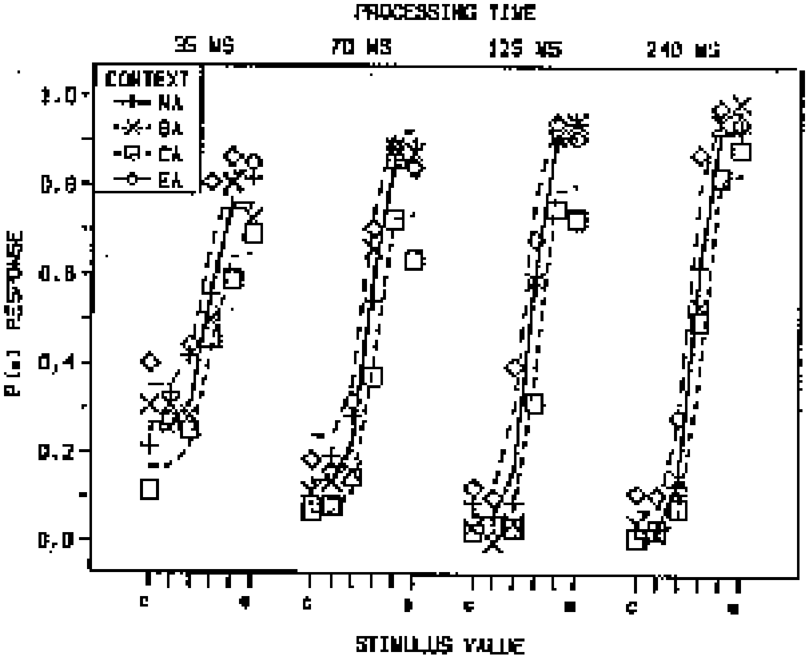

FIG. 7.3. Observed (points) and predicted (lines) probability of *e* identification as a function of the bar length of the test letter, the orthographic context, and the processing interval between the onset of the test stimulus and the onset of the masking stimulus for the dynamic FLMP model (results after Massaro, 1979; predictions from Massaro & Cohen, 1991).

controlled by the time between the test display and a backward masking stimulus. The masking stimulus was composed of nonsense letters created by selecting random feature strokes from the letters of the alphabet. Performance was more chaotic (in the sense of being more random and giving a more restricted range of response probabilities) at the short masking intervals. That is, less processing time leads to less orderly behavior—as expected from research on the time course of perceptual processing. Even for prototypical test letters, subjects did not make consistent identification judgments at short masking intervals. According to the FLMP, there was not sufficient time for feature evaluation and integration to take place before the onset of the masking stimulus.

Both the test letter and the context influenced performance at all masking intervals. The effect of test letter was attenuated at the short relative to the long processing time. That is, the identification functions covered a larger range across the *e–c* continuum with increases in processing time. Context has a significant effect at all masking intervals. In fact, the context effect was larger for the prototypical test letters at the short than at the longer masking intervals. This result follows naturally from the trade-off between stimulus information and context in the FLMP. Context has a larger influence to the extent the stimulus information is ambiguous. We now discuss the tests of the FLMP and IAM against these results.

Tests of the FLMP and IAM

Massaro and Cohen (1991) extended the FLMP to account for the time course of perceptual processing. They assumed that feature evaluation would follow the same negatively accelerating growth function found in backward recognition masking tasks. The backward masking function can be described accurately by a negatively accelerated exponential growth function of processing time,

$$d' = \alpha(1 - e^{-\theta t}) \tag{1}$$

where d' is an index of resolution of the target. The parameter α is the asymptote of the function and θ is the rate of growth to the asymptote. The function putatively describes feature evaluation and can be conceptualized as representing a process that resolves some fixed proportion of the potential information that remains to be resolved per unit of time. The same increment in processing time results in a larger absolute improvement in performance early relative to late in the processing interval.

Early in featural evaluation, the perceiver would have some information about each feature (dimension), but the information would not be sufficient to inform the perceiver about the identity of the stimulus. Integration of the separate features (dimensions) would be continuous and would be based on the outputs of feature evaluation. Similarly, decision (and thus response selection) could occur at any time after the stimulus presentation. For example, a response could be

initiated before sufficient information is accumulated—as might occur in speed accuracy experiments.

Following the theoretical analysis of backward masking, a masking stimulus would terminate any additional processing of the test stimulus. The dynamic model given by Equation 1 can be combined with the FLMP to describe how multiple sources of information are evaluated and integrated over time. The output from evaluation would be fed continuously to the integration process—which would operate as assumed in the FLMP. Integration outputs would be fed forward to the decision process, which would compute the relative goodness of match of the alternatives. In the backward masking task with unlimited response time, it seems reasonable to assume that the decision is not made until evaluation of the sources of information is near asymptote. That is, less processing time leads to less orderly behavior—as expected from research on the time course of perceptual processing.

This dynamic FLMP was tested against the results (Massaro & Cohen, 1991). Given the four masking intervals in the task, it is possible to describe performance in terms of the change in featural information and orthographic context across the four masking intervals. Eleven free parameters are required for the fit of the FLMP: the rate of growth of the functions and 10 asymptotic values for the 10 functions from the 6 levels of stimulus information and the four levels of context. The fit of the model was very good; the root mean square deviation (RMSD) between the observed and predicted points was .0501.

Massaro and Cohen (1991) also fit a variety of stochastic IAM models to the data (see McClelland, 1991). To bring the model in line with the empirical results of Massaro (1989), McClelland (1991) modified the original interactive activation model by allowing variability at the input and/or during each processing cycle and changing the decision rule from a relative goodness rule to a best-one wins rule. The topology of the network tested by Massaro and Cohen (1991) is designed to account for the effects of both the target letter and the orthographic context in the Massaro (1979) study. The network assumes three layers of units: CONTEXT, TARGET, and WORD. There were two-way connections between the CONTEXT and WORD units and the TARGET and WORD units to reflect interactive activation. It was also assumed that the mask terminated further processing, as in the fit of the dynamic FLMP. This model also required 11 free parameters—10 for the inputs corresponding to the 6 levels of stimulus and 4 levels of context and 1 parameter that translates processing time in the number of processing cycles. The RMSD obtained for this model was .1135, about twice that found for the dynamic FLMP. Thus, the orthographic-knowledge account of the FLMP gives a much better description of the results than the specific-word account of interactive activation.

These results provide a dramatic falsification of the need for interactive activation in word recognition in reading. In the IAM, context modifies the representation of the target letter. As shown by the good description given by the

FLMP, however, the perceptual advantage for letters in words can be explained in terms of the word context functioning as an additional independent source of information. The IAM also falsely predicts that the influence of context must accumulate later during perceptual processing than the influence of stimulus information. Contrary to IAM, the FLMP accurately predicts that the influence of context does not necessarily lag behind the influence of stimulus information. These experimental results agree with our reading experience. In the IAM, slow reading means long processing times in which the context can overwhelm the stimulus information about the target letter. In contrast, we read slowly to detect spelling errors, which shows that information about the target letter can remain immune to the context. We recognize letters better in words than in nonwords simply because we recognize words better than nonwords, not because the letter is somehow modified differently in word and nonword contexts.

Word Superiority Effect (WSE)

We began this chapter with the early findings that words give an advantage to the letters that make them up. Psychologists were suspect of these results because post hoc guessing and memory are critical contributions to performance in tachistoscopic experiments, and these could be responsible for the results. In order to assess the perceptual effect, the experiments must account for both post hoc guessing and memory contributions to performance. This might be seen as an insurmountable obstacle, but some good solutions have been proposed. The most influential solution has been the Reicher–Wheeler task (Reicher, 1969; Wheeler, 1970). In this task, subjects see a short target display followed by two test alternatives for one letter position. For example, on one trial, the subject might see the target WORD. If the fourth letter position is tested, two alternatives would be presented after the display, D and K. The subject would have to choose one of these alternatives. In this task, then, the subject must make a choice only on the basis of the information processed during the visual display. Knowledge of the rules of English spelling will *not* help after the display because both alternatives D and K form words given the information WOR–. Of course, a different word is presented on each trial, and the subject does not know which letter position would be tested until the cue appears. Performance in this condition is compared with performance when the subject was presented with a single letter at any of the four letter positions defined by the word. For example, the subject could be presented with D alone and asked whether it was D or K. A third condition is a nonword that does not conform to the spelling rules of English: for example, OWRD or OWRK.

Performance in the task gives a word advantage over nonwords and sometimes over single letters. In Reicher's experiment, subjects picked the correct alternative 12% more often when faced with a word display than when tested with a single-letter display. The nonword display produced performance equivalent to the single-letter condition. The advantage of words over single letters and nonwords has been dubbed the Word Superiority Effect (WSE). Reicher's findings have been replicated repeatedly (e.g., Johnston & McClelland, 1973; Wheeler, 1970) and recent research has focused on the explanation of the findings.

Within the context of the FLMP, visual information and orthographic context are integrated during the perceptual processing of a letter string, and the resulting percept is influenced by both these sources of information. In the Reicher–Wheeler task, simple post hoc guessing is controlled by giving subjects two response alternatives, after the test presentation, but orthographic information can still be used during perception. As an example, suppose we are interested in comparing a word *WORD* with a test nonword *ORWD*. In both cases, the two test alternatives are *D* and *K* for the fourth position. If subjects have no visual information about the test string, they would have no advantage in the word relative to the nonword condition. However, if a curved feature from the fourth letter was derived from the visual information, then the candidates for this position might be *D*, *O*, or *Q*. If the first three letters *WOR* were also recognized in the test word, then orthographic context would eliminate the candidates *O* and *Q*, leaving *D* as the only perceptual alternative. Recognizing *ORW* in the nonword condition would not constrain the alternatives for the fourth position, thus making perception of any of the three alternatives equally likely. In the FLMP, the advantage of words over nonwords in the Reicher–Wheeler task results from this contextual difference. The Reicher–Wheeler control does not eliminate a possible influence of orthographic context during perception; the control only precludes a postperceptual guessing advantage for words.

In the IAM, a WSE occurs because top down connections from the word level to the letter level allow context to modify the representation at the letter level. Although the model can account for many of the existing results on the WSE, it is important to stress that interactive activation is not *necessary* to account for these results. The FLMP, for example, does so without interactive activation. In the FLMP, context operates independently of featural analyses, simply by providing an additional source of information (Massaro, 1984). We now describe a strong test of the FLMP's description of the WSE.

Backward Masking, Lateral Masking, and the WSE

Johnston and McClelland (1973) found a WSE over letters when the test display was followed by a mask, but not when no mask was presented, and offered three possible explanations. Massaro (1975) explained this effect in terms of the tradeoff between the positive contributions of orthographic context and the negative effect of lateral masking. To test this explanation, Massaro and Klitzke (1979) employed four types of display in the Reicher–Wheeler task: words, nonwords, letters, and letters flanked by dollar signs. On each trial, one of these test

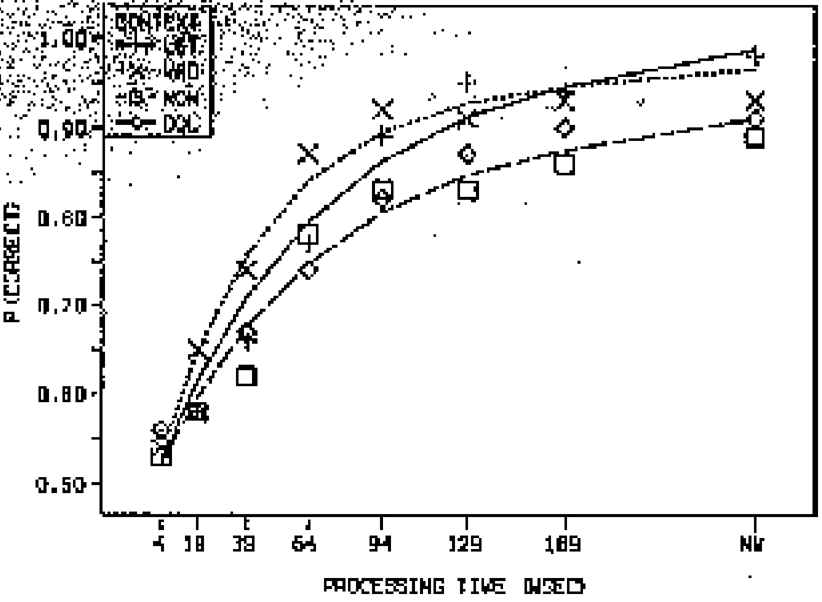

FIG. 7.4. Observed probability (points) of correct identification of the test letter as a function processing time for the letter alone (LET), word (WRD), nonword (NON), and letter-in-dollar-signs context (DOL) conditions. The lines are the predictions of the FLMP. The left-most line gives the predictions for the word context, the middle line for the letter-alone condition, and the right-most line for the nonword and letter-in-dollar-signs contexts. Note that both the observations and the predictions give a crossover between the word and letter-alone conditions (results after Massaro & Klitzke, 1979; predictions from Massaro & Cohen, 1991).

displays was presented alone or followed by a masking display after one of seven stimulus onset asynchronies (SOAs). Two choice-alternatives were presented ½ s after the test display as in the standard Reicher–Wheeler control. The masking stimulus varied from trial to trial and was composed of nonsense letters created by selecting random feature strokes from the letters of the alphabet.

The results shown in Fig. 7.4 indicated that performance improved with increases in processing time. Performance was better for words relative to letters at short processing times, but not at long processing times. Performance for nonwords and letter-in-dollar-signs was poorer than for single letters at all processing times. Because of the lateral interference of adjacent letters on each other, the sensory information in the word, nonword, and letter-in-dollar-signs conditions must necessarily be less than the single-letter condition. Of course, the word condition still has the advantage of orthographic context, whereas the other three types of display do not. Lateral masking and orthographic context must necessarily counteract each other and only a quantitative model can be reasonably tested against the results.

The dynamic FLMP not only predicts a WSE, it predicts a subtle interaction between the WSE, backward masking, and lateral masking. In the FLMP, given a test letter alone, only a single source of information is evaluated. Given a test word, two sources of information are evaluated and integrated. Thus, the test word will tend to accumulate more information over time than the test letter, and a WSE should be observed. This instantiation of the FLMP can predict that two sources of information can lead to better performance than just one. If the visual information about the test letter is presented in a word or nonword context, an advantage for words over nonwords is predicted. This prediction is consistent with the observed results. At asymptote, the letter presented alone gave better performance than a letter presented in a word. The model also predicts a word advantage over single letters with a masking stimulus at short SOAs, but not at long SOAs and when no mask is presented.

Performance in the nonword and letter-in-dollar-signs conditions was poorer than in word and letter alone condition. Performance in the former displays suffers because of lateral masking and the lack of a benefit from orthographic context. These results were predicted using the same free parameters values that were used to describe the letter and word conditions.

The interaction of the WSE with SOA is an important result because it reflects the interaction of a contextual (cognitive) influence with two sensory influences (perceptual processing time and lateral masking). Massaro and Cohen (1991) showed that the dynamic FLMP captures the observed results in a direct and parsimonious manner by accounting for the influences at the appropriate levels of processing. The components of the FLMP reflect the contributions of lateral masking, backward masking, and orthographic context.

The research appears to support the reader's use of spelling regularities in word identification. A potentially problematical result was obtained by Johnston (1978), who found no effect of lexical constraint in test words in the Reicher–Wheeler task. Lexical constraint refers to extent three of the four letters reduces uncertainty about the fourth letter (e.g., the extent to which -hip constrains s--- versus the extent to which -ink contains s---). This result appears to question the assumption that the word context provides an additional source of information. Analogous to Paap, Newsome, McDonald, and Schvaneveldt (1982), we do not find Johnston's results troublesome. He calculated lexical constraint based on the assumption of complete knowledge of the three-letter context and no knowledge of the fourth letter. This analysis might not correlate with the actual constraints when only partial information about the test and context letters is available. Paap et al. provided evidence that the lexical constraint between Johnston's high and low constraint words does not differ when partial information is assumed to be available at each letter position. Paap et al. reanalyzed the results from the individual words in Johnston's experiment and found that lexical con-

[illegible] had an effect when estimated under the assumption of partial information. In addition, we believe that sublexical constraints must also be included for a complete account of performance.

WORDS IN CONTEXT

Just as orthographic and lexical information can influence identification of letters, the identification of words can be influenced by syntactic and semantic information in the discourse context. However, the literature on context effects is conflicting; some experimenters have obtained large advantages for words in an appropriate context (e.g., Sanocki & Oden, 1984; Tulving & Gold, 1963) but others have not (e.g., Fischler & Bloom, 1979; Stanovich & West, 1983). However, these negative findings can be attributed to limitations in the methods used for measuring context effects. Often, the effects of information in a sentence context is gauged by presenting an incomplete sentence followed by a target word that completes the sentence appropriately or inappropriately. However, the occurrence of incongruous context-target relations invalidates the contextual information, with the result that subjects may use it less than otherwise (Sanocki & Oden, 1984). If only congruous context-target relations are used, advantages of an appropriate context increase, a result that obtains when word processing is measured with a lexical decision task (Sanocki & Oden, 1984) and with a naming task (Norris, 1987).

A more subtle problem with many context-effect studies is that the target word is presented in isolation and must be processed immediately. In contrast, during normal reading the processing of a word may be cascaded, beginning as parafoveal information is picked up and sometimes ending when the word's meaning is resolved, after fixation has moved beyond the word (see Sanocki et al., 1985). Therefore, it may be important to study context effects occurring when subjects can read whole sentences and phrases. In one such study, Sanocki et al. (1985, Experiment 3) compared the efficiency of processing words when either lexical, semantic, or both lexical and semantic information could be used. Subjects scanned word strings for violations, defined as either (a) a nonword within a scrambled sentence (in this case, only lexical information can be used), (b) an incongruous word in an otherwise meaningful sentence (semantic information only can be used), or (c) a nonword in a sentence (both lexical and semantic information can be used). Subjects were faster in the third condition when they could use both lexical and semantic information. More importantly, comparisons of the reaction time frequency distributions for the three conditions indicated that the advantage in this condition was greater than would be expected from a horse race in which the fastest of autonomous lexical and semantic processes was chosen (e.g., Forster, 1979). The results are consistent with the idea that lexical and semantic information were conjoined during the process, and with the general claim that two sources of information are better than one.

CONCLUSION

We have seen that visual processes are fundamental to reading letters and words. In addition, the perceptual processing of text is greatly facilitated by nonvisual information. The nonvisual information supplements the visual information and improves performance relative to the situation with just visual information. The IAM incorrectly predicts that nonvisual sources modify the representation of the visual sources rather than adding just an independent source. The results are nicely described by the FLMP, which predicts that two sources of information are more informative than just one. The FLMP is also able to account for the temporal course of perceptual processing of visual and nonvisual sources of information.

ACKNOWLEDGMENT

The research reported in this paper and the writing of the paper were supported, in part, by grants from the Public Health Service (PHS R01 NS 20314) and the National Science Foundation (BNS 8812728) to DWM.

REFERENCES

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

Anderson, I. H., & Dearborn, W. F. (1952). *The psychology of teaching reading*. New York: Ronald Press.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*, 115–147.

Bouma, H. (1971). Visual recognition of isolated lower-case letters. *Vision Research*, *11*, 459–474.

Brooks, L. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 169–211). Hillsdale, NJ: Lawrence Erlbaum Associates.

Downing, J. (1967). What's wrong with the i.t.a.? *Phi Delta Kappa*, *February*, 262–266.

Enns, J. T., & Rensink, R. A. (1990). Sensitivity to three-dimensional orientation in visual search. *Psychological Science*, *1*, 323–326.

Enns, J. T., & Rensink, R. A. (1991). Preattentive recovery of three-dimensional orientation from line drawings. *Psychological Review*, *98*, 335–351.

Fischler, I., & Bloom, P. A. (1979). Automatic and attentional processes in the effects of sentence contexts on word recognition. *Journal of Verbal Learning and Verbal Behavior*, *18*, 1–20.

Forster, K. I. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper & E. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 27–85). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gervais, M. J., Harvey, L. O., & Roberts, J. O. (1984). Identification confusions among letters of the alphabet. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 655–666.

Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts.

Gibson, E. J., & Levin, H. (1975). *The psychology of reading*. Cambridge, MA: MIT Press.

Gilmore, G. C. (1985). Letters are visual stimuli: A comment on the use of confusion matrices. *Perception and Psychophysics, 37*, 389–390.

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 5*, 674–691.

Groff, P. (1975). Research in brief: Shapes as cues to word recognition. *Visible Language, 9*, 67–71.

Huey, E. B. (1968). *The psychology and pedagogy of reading*. Cambridge, MA: MIT Press. (Original work published 1908)

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review, 99*, 480–517.

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

Johnston, J. C., & McClelland, J. L. (1973). Visual factors in word perception. *Perception & Psychophysics, 14*, 365–370.

Keren, G., & Baggen, S. (1981). Recognition models of alphanumeric characters. *Perception and Psychophysics, 29*, 234–246.

Krueger, L. E., & Chignell, M. H. (1985). Same-different judgments under high speed stress: Missing feature principle predominates in early processing. *Perception and Psychophysics, 38*, 188–193.

Lowe, D. G. (1985). *Perceptual organization and visual recognition*. Boston: Kluwer.

Lupker, S. J. (1979). On the nature of perceptual information during letter perception. *Perception and Psychophysics, 25*, 303–312.

Massaro, D. W. (1975). Primary and secondary recognition in reading. In D. W. Massaro (Ed.), *Understanding language: An information processing analysis of speech perception, reading, and psycholinguistics* (pp. 241–289). New York: Academic Press.

Massaro, D. W. (1979). Letter information and orthographic context in word perception. *Journal of Experimental Psychology: Human Perception and Performance, 5*, 595–609.

Massaro, D. W. (1984). Building and testing models of reading processes. In P. D. Pearson (Ed.), *Handbook of reading research* (pp. 111–146). New York: Longman.

Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Massaro, D. W., & Cohen, M. M. (1991). Integration versus interactive activation: The joint influence of stimulus and context in perception. *Cognitive Psychology, 23*, 558–614.

Massaro, D. W., & Hary, J. M. (1986). Addressing issues in letter recognition. *Psychological Research, 48*, 123–132.

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

Massaro, D. W., Taylor, G. A., Venezky, R. L., Jastrzembski, J. E., & Lucas, P.A. (1980). *Letter and word perception: Orthographic structure and visual processing in reading*. Amsterdam: North-Holland.

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

McClelland, J. L. (1991). Stochastic interactive processes and the effect of context on perception. *Cognitive Psychology, 23*, 1–44.

Navon, D. (1991). Testing a queue hypothesis for the processing of global and local information. *Journal of Experimental Psychology: General, 120*, 151–172.

Norris, D. (1987). Strategic control of sentence context effects in a naming task. *Quarterly Journal of Experimental Psychology, 39A*, 253–275.

Oden, G. C. (1979). A fuzzy logical model of letter identification. *Journal of Experimental Psychology: Human Perception and Performance, 5*, 336–352.

Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological Review, 85*, 172–191.

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

Paap, K. R., Newsome, S. L., & Noel, R. W. (1984). Word shape's in poor shape for the race to the lexicon. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 413–428.

Pitman, J., & St. John, J. (1969). *Alphabets and reading*. New York: Pitman.

Price, C. J., & Humphreys, G. W. (1989). The effects of surface detail on object categorization and naming. *Quarterly Journal of Experimental Psychology, 41A*, 797–828.

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology, 81*, 275–280.

Sanocki, T. (1987). Visual knowledge underlying letter perception: Font-specific, schematic tuning. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 267–278.

Sanocki, T. (1988). Font regularity constraints on the process of letter recognition. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 472–480.

Sanocki, T. (1991a). Intra- and interpattern relations in letter recognition. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 924–941.

Sanocki, T. (1991b). Effects of early common features on form recognition. *Perception and Psychophysics, 50*, 490–497.

Sanocki, T., Goldman, K., Waltz, J., Cook, C., Epstein, W., & Oden, G. C. (1985). Interaction of stimulus and contextual information during reading: Identifying words within sentences. *Memory and Cognition, 13*, 145–157.

Sanocki, T., & Oden, G. C. (1984). Contextual validity and the effects of low-constraint sentence contexts on lexical decisions. *Quarterly Journal of Experimental Psychology, 36A*, 145–156.

Sanocki, T., & Rose V. (1989–90). Modifiable letterforms for teaching reading: The graphophonic alphabet. *Journal of Educational Technology Systems, 18*, 173–183.

Stanovich, K. E., & West, R. F. (1983). On priming by a sentence context. *Journal of Experimental Psychology: General, 112*, 1–36.

Townsend, J. T., & Ashby, F. G. (1982). Experimental test of contemporary mathematical models of visual letter recognition. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 834–864.

Townsend, J. T., Hu, G. G., & Evans, R. J. (1984). Modeling feature perception in brief displays with evidence for positive interdependencies. *Perception and Psychophysics, 36*, 35–49.

Townsend, J. T., Hu, G. G., & Kadlec, H. (1988). Feature sensitivity, bias, and interdependencies as a function of energy and payoffs. *Perception and Psychophysics, 43*, 575–591.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review, 95*, 15–48.

Tulving, E., & Gold, C. (1963). Stimulus information and contextual information as determinants of tachistoscopic recognition of words. *Journal of Experimental Psychology, 66*, 319–327.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327–352.

Waters, G. S., & Seidenberg, M. S. (1985). Spelling-sound effects in reading: Time-course and decision criteria. *Memory & Cognition, 13*, 557–572.

Venezky, R. L., & Massaro, D. W. (1987). Orthographic structure and spelling-sound regularity in reading English words. In D. A. Allport, D. G. MacKay, W. Prinz, & E. Scheerer (Eds.), *Language perception and production: Shared mechanisms in listening, speaking, reading and writing* (pp. 111–129). London: Academic Press.

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

Witkin, A. P., & Tenenbaum, J. M. (1983). On the role of structure in vision. In J. Beck, B. Hope, & A. Rosenfeld (Eds.), *Human and machine vision* (pp. 481–543). New York: Academic Press.