Reading and Listening (Tutorial Paper)

Dominic W. Massaro

Reading text and listening to speech involve a sequence of processing stages that take the language user from the spoken or written message to meaning. This paper describes each stage in an information processing model. At each stage of the model there are storage and process components: the storage component defines the type of information available at the stage of processing; the process component specifies the procedures that operate on the information held in the corresponding storage component. The model is used heuristically to incorporate data and theory from a variety of approaches to reading and speech perception. Some relevant issues are the properties of the various storage structures, the dynamics of the functional processes, and the active utilization of the various sources of knowledge available to the language receiver. One central assumption is that analogous structures and processes occur in speech perception and reading. It is valuable, therefore, to ask similar experimental questions in both areas, and to attempt to develop a unitary model of both reading and speech perception.

A session on Reading and Listening must be motivated by the belief that these two skills are not unrelated. And, in fact, it has been commonly believed by many that reading is somehow a skill parasitic to listening: that is, all that is necessary in order to read is to translate the written word into its spoken counterpart, and then to listen to the message. Since most literate people learn to listen long before they learn to read, it seems natural that reading instruction should exploit what the child knows about listening. Faced with a new word in written text, the child is encouraged to decode it (sound it out) on the assumption that the spoken rendering will more likely be

recognized. The primacy of speech has been one motivating influence on theories of reading that view reading as a process of going from print to some speech code enroute to the meaning (Gough, 1972). Since it is a common belief that man spoke and listened long before he wrote and read, one could also argue that language learning recapitulates the development of spoken and written language. However Wrolstad (1976), in a stimulating and fascinating manifesto, argues the reverse.

This author rejects the idea that speech is primary and that reading is subservient to or dependent on listening. Reading and listening can be viewed as independent but analogous processes, the goal of which is to derive the meaning of a message. One argument for these being analogous processes can be formulated in terms of convergent evolution: Two independent processes may develop similar solutions to a problem of survival. As an example, the eye of the octopus (a cephalopod) and the eye of man (a mammal) evolved completely independently of one another. But, as Blakemore (1977) observed, the eyes evolved to be very similar to one another. Both organisms developed functionally similar solutions to the problem of seeing. Following this logic, the assumption that reading and listening can be viewed as similar processes does not necessitate an assumption of a common phylogenetic or ontogenetic evolution. Given that reading and listening solve the same problem, it is not unreasonable to assume that they are analogous rather than hierarchical solutions to the problem.

Analogous solutions to language understanding would not be limited to reading and listening. Consider the recent proliferation of research on the processing of manual-visual languages such as American Sign Language (ASL). Remarkable parallels have been found between understanding signs and understanding speech. Lane, Boyes-Braem, and Bellugi (1976) found that perceptual confusions among signs can be described utilizing a distinctive feature system, analogous to systems developed for perceptual confusions in speech. Tweney, Heiman, and Hoemann (1977) demonstrated that grammatical structure in sign language plays the same functional role that it does in spoken language. These results support the claim that grammatical processes of language are relatively general and abstract—not tied uniquely to the input modality. Although this paper is limited to reading and listening, the work on ASL encourages the belief that there are similar and analogous processes in all forms of language understanding.

A Stage Model of Language Processing

Reading can be defined as the abstraction of meaning from printed text; listening as the abstraction of meaning from speech. Deriving meaning from a spoken or written message requires a series of transformations of the energy signal after it arrives at the appropriate receptors. Language processing can be studied as a sequence of processing stages or operations that occur between the energy stimulus and the meaning. In this framework, language processing can be understood only to the extent that each of these processing stages is described. In a previous effort, a general information-processing model was utilized for a theoretical analysis of speech perception, reading, and psycholinguistics (Massaro, 1975a). The model was used heuristically to incorporate data and theory from a wide variety of approaches in the study

of language processing. The model should be seen as an organizational structure for the state of the art in language processing. In this paper, I will present a general overview of the information-processing model, and use the model to describe and incorporate recent research.

Figure 1 presents a flow chart of the temporal course of reading and listening. At each stage the system contains storage and process components. The storage component defines the information available to a particular stage of processing. The process component specifies the procedures and processes that operate on the information held in the corresponding storage component. The model distinguishes four process components: feature detection, primary recognition, secondary recognition, and rehearsal-recoding. The model will be described in more detail below, and contrasted with other models of language processing.

The primary assumption of a stage model is that, at some level, the stages are necessarily successive. The feature-detection process transforms the energy pattern created by the language stimulus and transduced by the appropriate receptor system into a set of features held in preperceptual storage (PS). Primary recognition which evaluates and integrates the features in PS into a synthesized percept must necessarily follow feature detection. However, not all of the features must be detected before primary recognition begins. Given that different features require different detection times, not

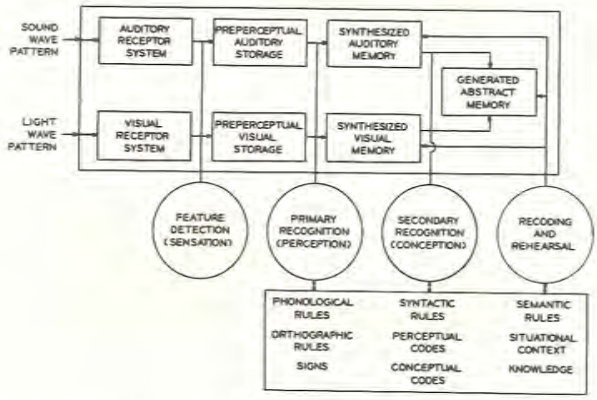

Figure 1. A flow diagram of the temporal course of reading and listening.

all of the features will enter PS simultaneously. Primary recognition could, therefore, evaluate, integrate, and continually update its analysis of the contents of PS while feature detection goes to completion. As can be seen in this example, a basic assumption is that the successive processes necessarily overlap in time. Although this assumption complicates the traditional serial nonoverlapping-stages model, the model continues to have predictive and descriptive power. The serial model is necessary, in the present view, because later stages of processing do not have access to early, low levels of information at early stages. As an example, rehearsal-recoding does not appear to have direct access to the features that are detected and held in preperceptual storage.

Criticisms of Stage Models

The model under discussion here can be classified in a successive-stage information-processing framework; it therefore has some similarities to the stage models developed by Estes (1975), Gough (1972), and LaBerge and Samuels (1974). Rumelhart (1977) justly criticizes Gough's and LaBerge and Samuels' models because of the assumption that "no higher level can in any way modify or change the analysis at a lower level". As evidence against this assumption, he mentions the well-known finding that the perception of a letter can be influenced by the letters that surround it. A letter that is part of a word or a pseudoword is more accurately perceived than the same letter presented in a nonword context. In this case, then, it appears that higher-order knowledge of letter groups or words influences the perception of individual letters. In Rumelhart's opinion and also that of Allport (1977), this result cannot be easily handled by serial models. I would like to stress, however, that not only is this result easily described by the present serial model, but the model was also instrumental in generating hypotheses and experiments on exactly this phenomenon (Massaro, 1973, 1975a; Massaro, Venezky, & Taylor, in press; Thompson & Massaro, 1973). In our model two sources of information are available at the primary recognition stage in reading: (1) the featural information in perceptual visual storage (PVS) and 2) knowledge about the orthographic structure of English spelling stored in long-term memory (LTM). Both of these sources of information contribute to the perception of individual letters in a letter string. (See Massaro, 1975a; and Venezky & Massaro, in press, for a more detailed description).

Although it is clear that the present model allows orthographic structure to influence letter perception, the influence of syntactic/semantic constraints may not be immediately obvious. It seems obvious that the same visual information can be interpreted differently in different syntactic/semantic contexts (Nash-Weber, 1975). Figure 2 shows that, faced with the same visual configuration, the preceding context supports the recognition of cast in one case and east in the other. In our model, knowledge obtained at the level of generated abstract memory (GAM) would be combined with the featural information in PVS. Given the same featural information, different contexts can lead to different outcomes of primary recognition. Although this possibility was acknowledged in our earlier work, very few data were available and, therefore, the influence of semantic/syntactic constraints at primary recognition was not made explicit (however, see Massaro, 1975a, pp. 258-259). More recently, Marslen-Wilson and Welsh (1978) have found that the message

It was an old film with a east of thousands.

The tourist took a train going east.

Figure 2. The same visual configuration can be interpreted differently in different syntactic/semantic contexts.

being understood at the level of GAM can influence the recognition of the message. I will extend the model by developing a quantitative formulation of how syntactic/semantic constraints and acoustic information are combined and integrated at the recognition stage of processing (see the section on Secondary Recognition).

Parallel Interactive Models

In two recent schemata of reading systems, Allport (1977) and Rumelhart (1977) have drawn heavily from contemporary work in artificial intelligence (AI). The distinctive feature of Rumelhart's system is a set of independent knowledge sources, each of which applies specialized knowledge about some aspect of reading. More specifically, each knowledge source generates hypotheses about the input and relays these to a message center which maintains a running list of all incoming hypotheses. The knowledge sources evaluate their current hypotheses and generate new ones until some decision can be reached. In this model, all levels of knowledge interact, and each contributes independent evidence for the final interpretation.

Although this view has some appeal, there are a number of deficiencies when it is evaluated in terms of a psychologically real description of reading. The primary problem arises because the model is a direct application of a nonhuman system. In this case, the system HEARSAY II was developed for automatic speech understanding (Lesser, Fennell, Erman, & Reddy, 1975). Although the system proved effective in understanding speech, it had to perform in a manner that was clearly different from the way humans understand speech.

The specific advantage that HEARSAY II has over its human counterpart is that it has a high resolution memory. In HEARSAY II, the parametric level holds the most basic representation of the utterance in terms of fine-grained acoustic parameters. These low-level decisions are not final since the original parameters can be processed and evaluated continuously throughout the understanding process. By failing to make final, low-level decisions, the number of hypotheses and amount of data that have to be maintained would far exceed the capacity of human memory structures. We know that preperceptual auditory storage holds featural information for roughly a quarter of a second, less than the time required to hold a complete sentence until it is finally interpreted (Massaro, 1975a).

This observation is not meant as a criticism of automatic speech recognition, but rather as a caveat for those who seek psychological models in that arena. Although HEARSAY II is a very successful system, it does not make any final decisions about the first part of a sentence until all of the sentence has been processed. Humans recognize speech while it is being spoken, if not before. My criterion for the application of AI theories to humans is not that machines perform as well as humans, but that machines perform like humans.

Allowing low-level data to be continuously available may be more compatible with reading than with speech perception, since the reader can always make a regressive eye movement to refresh the image. In this case, the low-level information in reading can be maintained indefinitely as it is in automatic speech recognition. Considering the number of hypotheses that might have to be checked at this level, however, the number of regressive eye movements would be enormous and the reader would never finish the passage. The small number of regressions in normal reading infirms the idea that low level information in reading is continually rechecked until a decision is made. Although some concepts developed in AI understanding systems have obvious relevance to psychological theory, current AI schemes are not good models of what is known about how humans process language.

Summary

The model presented here was developed in the genre of a serial stage and information processing framework. A number of recent statements have pointed out limitations to this approach, and have argued for a system based on parallel and interactive channels (Allport, 1977; Rumelhart, 1977). In the serial stage model, each stage has contact with only the information available to that stage and some neighboring stages, and with the appropriate information in long-term memory. For example, the secondary recognition operation in our model has contact with synthesized memory but not with preperceptual storage. Accordingly, the information that was lost during the primary recognition transformation cannot influence secondary recognition. In an interactive model, the featural information at preperceptual storage would influence very late and high-level processing stages. This assumption is unreasonable because the early low-level information does not remain intact long enough to contribute to later stages of processing. For example, an eye movement in reading erases and replaces the featural information derived in the earlier fixation. It is only the transformed information, then, that can contribute to later processing. The temporal limitations on retention of the various information sources is the primary motivation for utilizing a serial model, and in this regard the parallel interactive models are not good psychological models. The stage model will now be developed in more detail by tracing the flow of information processed in reading and listening. By reviewing relevant issues and research, it is hoped that the heuristic value of the model will become apparent.

Feature Detection

The feature detection process transforms the energy pattern into features in preperceptual storage. The features for reading and listening are described as visual and acoustic, respectively, since it is assumed that there is a direct relationship between the nature of the energy signal and the information in preperceptual storage. There are a number of prominent characteristics in speech and reading stimuli that are potential candidates for features. A characteristic is called a feature only when it is psychologically functional, that is, when it is used to distinguish between different speech sounds or written symbols (Massaro, 1975a).

Audible Features

The study of features in speech perception has had a long history. A good portion of the work was influenced by linguistic descriptions of binary all-or-none distinctive features (Jakobson, Fant, & Halle, 1961). One of the goals of distinctive feature theory was to minimize the number of distinctive features of the language. Distinctive features were designed to be general rather than specific: if a distinctive-feature difference distinguished two phonemes in the language, that same distinction was assumed to distinguish several other phoneme pairs. Given the distinctive feature of voicing, for example, the distinction of voiced versus voiceless could account for the differences between /z/ and /s/, /v/ and /f/, and so on.

In contrast to the linguistic description of binary all-or-none features, the features held in preperceptual auditory storage (PAS) are assumed to be continuous, so that a feature indicates the degree to which the quality is present in the speech sound. This assumption is similar to the more recent treatment of distinctive features provided by Chomsky and Halle (1968) and Ladefoged (1975). Chomsky and Halle distinguish between the classificatory and phonetic function of distinctive features. The features are assumed to be binary in their classificatory function, but multivalued in their phonetic or descriptive function. Ladefoged distinguishes between the phonetic and phonemic level of feature description. A feature describing the phonetic quality of a sound has a value along a continuous scale whereas a feature classifying the phonemic composition is given a discrete value.

In the present model an acoustic feature in PAS is expressed as a continuous value. That is to say, the listener will be able to hear the degree of presence or absence of a particular feature, even though his judgment in a forced choice task will be discrete. Thus, Oden and Massaro (1978) have described acoustic features as fuzzy predicates which may be more or less true rather than only absolutely true or false (Zadeh, 1971). In addition to being concerned with the acoustic features in preperceptual storage, this analysis of the feature evaluation process makes apparent that an important question in speech perception research is how the various continuous features are integrated into a synthesized percept.

The integration of features has not been extensively studied in speech perception research for two main reasons. If distinctive features are binary, the integration of information from two or more features would be a trivial problem. Given binary features representing voicing and place of articulation, for example, the integration would be a simple logical conjunction. If the

consonant /b/ is represented as voiced and labial, a voiced labial sound would be identified as /b/ whereas a voiceless labial sound or a voiced non-labial sound would not be identified as /b/.

A second reason for the neglect of the integration problem is methodological. The primary method of study involved experiments in which the speech sound was varied along a single relevant dimension. For example, in a study of voicing, all voicing cues were made neutral except one, such as voice onset time, and then this dimension was varied through the relevant values. Very few experiments independently varied two cues. Therefore, no information was available about how two or more cues were integrated into a synthesized percept.

More recently, Massaro and Cohen (1976) and Oden and Massaro (1978) have utilized factorial designs and functional measurement techniques (Anderson, 1974) to study the integration of acoustic features in speech perception. The work has illuminated how fuzzy features are integrated into a synthesized percept. The methods allow the investigator to determine which acoustic characteristics in the signal are perceptually functional, the relative weight that each feature carries in the integration process, and the exact combinatorial algorithm that is used.

Visible Features

One of the oldest areas of reading-related research is the study of the functional cues in recognizing printed characters. Much of this work was performed by typographers and artists concerned with good design and legibility of type faces. Although many of the early conclusions appear to be still valid (see Spencer, 1969, for a summary), they are almost totally ignored in the contemporary study of letter recognition. One influence in current studies has been the well-known neurophysiological findings of an amazing stimulus-selectivity in the responses of cells in the visual cortex (Hubel & Wiesel, 1962). There appear to be specialized detectors in the visual system for lines of specific size and orientation (for intelligible reviews of this work see Blakemore, 1973; Lindsay & Norman, 1977).

As in speech, psychological descriptions have centered around binary all-or-none features. Feature sets usually consist of the presence or absence of horizontal, vertical, or oblique lines, curves, intersections, angles, and so on. The feature sets are typically derived from and tested against the recognition confusions of capital letters (see Massaro, 1975a, Chapter 6, for a review). In contrast to the idea of binary all-or-none features, however, visible features, like audible features, may be fuzzy. Rather than a feature being present or absent, the information in preperceptual visual storage (PVS) could represent the degree to which a given feature is present in the signal. Given this conceptualization, it is necessary to develop a methodological and theoretical study of visible features that parallels the work on audible features.

Recently, Blesser and his colleagues have developed and studied ambiguous characters (Blesser, Shillman, Kuklinski, Cox, Eden, & Ventura, 1974; Shillman, Cox, Kuklinski, Ventura, Blesser, & Eden, 1974). Figure 3 presents a matrix of ambiguous characters having one or more functional attributes in transition between two letters. A completely ambiguous character is one that would

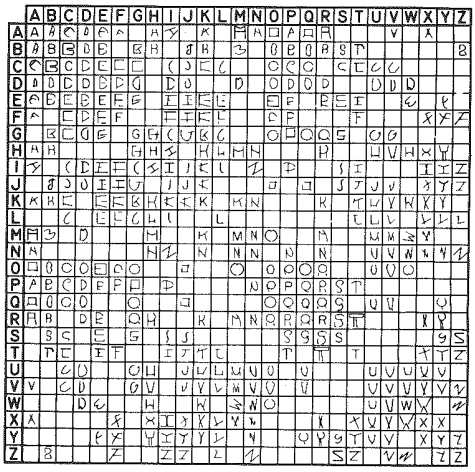

Figure 3. A matrix of ambiguous characters having one or more functional attributes in transition between two letters. (After Naus & Shillman, 1976).

be assigned to either letter class with equal probability. As an example, a v can be gradually transformed into a y by continuously increasing the right oblique line below the intersection (Naus & Shillman, 1976). Analogous to the recent work in speech perception, the theoretical notion of fuzzy information and the experimental methods of factorial designs and functional measurement techniques should advance the study of visible features in reading.

Given that most of the contemporary research on features is carried out with uppercase letters, it is important to know to what extent these features may generalize to perception of lowercase. Bouma (1971) showed that the visual confusions in the recognition of lowercase letters were better described in terms of the overall shape or envelope of the letter rather than the commonly assumed component features. The envelope of the letter is defined as the smallest enclosing polygon without indentations. The envelope for e would be circular, whereas v would be enclosed by an inverted triangle.

That the overall envelope of a letter may be resolved before details of its component features is consistent with Fourier-analysis studies. There is good evidence that low spatial frequency information is processed faster

than high frequencies (Breitmeyer & Ganz, 1976; Broadbent, 1977). Either limited processing time or a reduced figure-ground contrast could allow envelope resolution without corresponding resolution of letter detail. In terms of the present analysis, the reader will have to decide which letter best fits the perceived but fuzzy envelope. For example, recognition of a circular envelope would limit the alternatives to the three letters e, c, and o.

Primary Recognition

In the framework of the information processing model, the primary recognition process integrates the featural integration held in preperceptual storage into a percept in synthesized memory. One question relevant to the model concerns the functional units at this stage of processing. The functional units are called perceptual units; they correspond to units that are described in long-term memory. The primary recognition process finds the best match between the featural information and the descriptions of perceptual units in long-term memory.

Perceptual Units in Speech

Sound patterns of V, CV, or VC size are the best candidates for perceptual units since these patterns can be described by relatively invariant acoustic features (Fujimura, 1975). Smaller patterns such as phonemes lack invariance, and, in fact, this lack of invariance has been a central focus of speech perception theory (cf. Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Although perceptual units of larger size have also been proposed, research has demonstrated that preperceptual auditory storage cannot maintain featural information for a greater period than roughly one quarter of a second (Massaro, 1975a, Chapter 4). It follows that clauses, phrases, or even words are inappropriate perceptual units at the primary recognition stage of processing. (For further discussion see Massaro, 1975a, Chapter 4).

Target Search in Listening

If syllables are perceptual units in speech processing, they should be accessed before the phonemes that make them up. One test of this idea has been to study the reaction times to syllable and phoneme targets in speech processing. Savin and Bever (1970) asked listeners to release a telegraph key as soon as they heard a specified target in a sequence of nonsense consonant-vowel-consonant syllables (CVCs). Listeners were slower in responding to initial consonant phoneme targets such as /b/ or /s/ than in responding to a CVC target such as /bæb/ or /sæb/. Savin and Bever interpreted the differences in reaction time (RTs) as directly reflecting differences in the time required to perceive the phoneme and syllable targets and concluded that phonemes must be perceived only by an analysis of already perceived syllables. Warren (1971) reached similar conclusions from a target-search task through spoken sentences.

Foss and Swinney (1973) objected to the conclusion that phoneme perception must follow syllable perception. They used search lists of two-syllable words and asked subjects to search for an initial phoneme, initial syllable, or a complete word. Reaction times were actually somewhat shorter for word than for syllable targets, although phoneme targets were responded

to slowest of all. In terms of the logic of the target search task, their results would argue that words are accessed before the syllables that make them up. Burying the logic of the target search task even deeper, McNeill and Lindig (1973) included sentence targets and lists in a complete factorial design of phoneme, syllable, word, and sentence targets crossed with the same four kinds of lists. In general, RTs were slowed when the linguistic level of the target differed from that of the search list.

The latter studies have been used to support the thesis that the target-search task is inappropriate for a study of perceptual units in speech processing. This conclusion may be premature, however, given the inappropriate design of the experiments. The critical confounding was that appropriate test foils were not included in the experiments. For example, Foss and Swinney's (1973) results indicated that subjects were not faster in responding to syllable versus word targets. However, the foils were chosen in such a way that subjects could have based their decision on the first syllable of the word in both the target syllable and target word conditions. Given either the syllable target "can" or the word target "candy," subjects could have selected a reliable response after processing just the first syllable, since the test foils never began with "can." Apparently, this is the strategy that was used since the RT averaged about 350 msec, much less than the 500 msec duration of the test words. Similarly, McNeill and Lindig's subjects responded in less than 1/3 sec, showing that a response was initiated long before a word or sentence could have been processed. Therefore the target search task should be reconsidered as an appropriate paradigm, but it should be structured to account for both perceptual and comparison processes (see Ball, Wood, & Smith, 1975; Massaro & Klitzke, 1977; Sloboda, 1977).

Perceptual Units in Reading

In contrast to the acoustic structure of phonemes, printed letters of a given face are invariantly defined with respect to their visual characteristics. It follows that units as small as letters are reasonable candidates for perceptual units at the primary recognition stage in reading. Units of larger size have been proposed, however. Spelling patterns, pronunciation units, words, and even phrases have been proposed as perceptual units at the stage of processing responsible for the integration of visual features into a synthesized percept (Massaro, 1975a). If letters are basic perceptual units in primary recognition, words must be perceived by the letters that make them up. This hypothesis addresses the old and familiar question of whether word recognition can be described in terms of component letter recognition, or whether a word is recognized on the basis of supraletter features without reference to the letters that make it up.

Target Search in Reading

The target search task has also been used to address the issue of whether word recognition is mediated by letter perception. In an experiment carried out by Johnson (1975), subjects indicated whether or not a test item was the same as a target item. Response latencies did not depend on whether the target and test items were both single letters or both whole words. These results led Johnson to reject the idea that word recognition is mediated by letter recognition. However, Johnson's results are not necessarily incompatible

with preliminary letter recognition if the letters in a word are processed in parallel.

Massaro and Klitzke (1977) observed that the population of words of a given length may be more dissimilar from one another than are a set of single letters. If this is the case, a subject could perform with a less stringent criterion for accepting a match in the word condition than in the letter condition. Johnson (1975) made no attempt to control for the similarity of the test foils to the target items, thereby producing test foils that were naturally more dissimilar on word than on letter trials. Increasing the similarity between the target and test foils should decrease the advantages of parallel processing on word trials and, therefore, produce a letter advantage because of reduced figure-ground contrast (lateral masking) in processing word strings.

Massaro and Klitzke (1977) replicated Johnson's experiment while simultaneously investigating the role of similarity in the identification of words and letters. Half of the test foils were highly similar to the target and half were dissimilar. The manipulation of similarity was effective in that responses to similar test foils required 40 msec more than responses to dissimilar nontargets. In contrast to Johnson's finding of no difference, reaction times were 8 % (43 msec) faster for letter stimuli than word stimuli. The letter advantage was significant even when the analysis was restricted to target and dissimilar test foil trials. By including similar test foils, subjects could not maintain a lax criterion of sameness because a test foil word could be very similar to the target word. In this case, subjects could not select and execute a positive response as quickly as they might have with a more lax criterion of sameness. Massaro and Klitzke interpret the results as suggesting that perceptual processing and comparison are continuous and overlapping processes in the target search task. A subject does not first perceive the test item and then compare it to the target item in memory but rather the reader is able to make comparisons as partial information about the test item is resolved. Notice that this interpretation is consistent with the general idea of successive but overlapping stages in language processing. Therefore, the normal dissimilarity advantage of word strings over single letters can work to neutralize the word's disadvantage of lateral masking. Another important advantage for words is orthographic structure. But in both cases, the word advantage is not incompatible with the idea of letter units mediating word recognition (Sloboda, 1976, 1977).

Secondary Recognition

Secondary recognition transforms synthesized percepts into meaningful forms in generated abstract memory. In speech perception, it is assumed that the input is analyzed syllable by syllable for meaning. In reading, letter sequences are closed off in word units. In both cases, the secondary recognition process makes the transformation from percept to meaning by finding the best match between the perceptual information and the lexicon in long-term memory. Each word in the lexicon contains both perceptual and conceptual information. The concept recognized is a function of two independent sources of information: the perceptual information in synthesized memory and the syntactic/semantic context in the message.

Perceptual and Contextual Contributions to Listening

Abstracting meaning is a joint function of the independent contributions of the perceptual and contextual information available. In one experiment, Cole (1973) asked subjects to push a button every time they heard a mispronunciation in a spoken rendering of Lewis Carroll's "Through-the-Looking-Glass." A mispronunciation involved changing a phoneme by 1, 2, or 4 distinctive features (for example, "confusion" mispronounced as "gunfusion," "bunfusion," and "sunfusion," respectively). The probability of recognizing a mispronunciation increased from 30 to 75% with increases in the number of feature changes, making apparent the contribution of the perceptual information passed on by the primary recognition process. The contribution of contextual information should work against the recognition of a mispronunciation since context would support a correct rendering of the mispronounced word. In support of this idea, all mispronunciations were correctly recognized when the syllables were isolated and removed from the passage.

Cole and Jakimik (1977) extended Cole's (1973) mispronunciation task to evaluate how higher-order contextual information can influence sentence processing. To the extent that a word is predicted by its preceding context, the listener should be faster at detecting a mispronunciation. This follows from the idea that the quickest way to detect a mispronunciation is first to determine what the intended word is and then notice a mismatch with what was said. Given the sentences, "He sat reading a (book/bill) until it was time to go home for his tea," mispronouncing the /b/ in "book" as /v/ should be detected faster than the same mispronunciation of "bill." In fact, listeners were 150 msec faster detecting mispronunciations in highly predictable words than in unpredictable words.

Marslen-Wilson (1973) asked subjects to shadow (repeat back) prose as quickly as they heard it. Some individuals were able to shadow the speech at extremely close delays with lags of 250 msec, about the duration of a syllable or so. One might argue that the shadowing response was simply a sound-to-sound mapping without any higher order semantic-syntactic analyses. When subjects make errors in shadowing, however, the errors are syntactically and semantically appropriate given the preceding context. For example, given the sentence "He had heard at the Brigade," some subjects repeated "He had heard that the Brigade." The nature of the errors did not vary with their latency; the shadowing errors were always well-formed given the preceding context.

Marslen-Wilson and Welsh (1978) asked observers to shadow spoken passages from a popular novel. At random throughout the passage, common three-syllable words were mispronounced. When the words were mispronounced, only a single consonant phoneme was changed to a new consonant phoneme. The new phoneme differed from the original by one or three distinctive features, based on Keyser and Halle's (1968) classification system. Independently of the degree of feature change, the changes could occur in the first or third syllable of the three-syllable word. Finally, the mispronounced words were either highly predictable or unpredictable given the preceding portion of the passage. Subjects were not told that words could be mispronounced although they probably became aware of this early in the experiment. All subjects shadowed at relatively long delays greater than 600 msec. The primary dependent measure in the task was the percentage of fluent restorations, that is, the proportion of times the shadowers repeated what should have been said rather than what was said. About half of the mispronounced words were restored, the

restorations were made on-line with an average latency, and the shadowing was not disrupted. (When the mispronunciation was not restored, shadowing was disrupted and response times increased).

The change in the percentage of restorations as a function of the three independent variables in Marslen-Wilson and Welsh's study can illuminate how the listener integrates acoustic information and higher-order context. Figure 4 presents the observed results in terms of the percentage of fluent restorations. All three variables influenced the likelihood of a restoration. Shadowers were more likely to restore a one-feature than a three-feature change, a change in the third than in the first syllable, and a change in a highly predictable than in an unpredictable word.

The serial processing model can be quantified to predict the results of this experiment. The central assumption of the model is that the information passed on by feature detection does not change with changes in semantic/syntactic constraints. In terms of the model, the listener should be more likely to restore a single-feature change than a three-feature change. Substituting /f/ for /v/ should give the subject more acoustic information for /v/ than substituting /n/ for /v/. Letting A_i represent the acoustic information available for the original sound, then A_1 for a one-feature change should be larger than A_3 for a three-feature change. Mispronouncing the word in the first or third syllable should also influence the likelihood of a restoration. I have argued previously that speech is processed syllable by syllable (Massaro, 1975a), and the listener attempts to make lexical access at each syllable boundary. We might expect that lexical access of an upcoming word is usually not possible before some of the word is heard. This means that the lexical information available for the mispronounced word should be much higher at the time of the third than the first syllable. Accordingly, a subject should be much more likely to restore a mispronunciation in the third than the first syllable. Letting L_i equal the amount of lexical information that is available, L_3 should be much larger than L_1, where 1 and 3 refer to a mispronunciation in the first and third syllables, respectively. Finally, highly constrained context should make a restoration much more likely than unconstrained context. Therefore, letting S_c and S_u represent the amount of preceding sentential information under constrained and unconstrained contexts, we would expect S_c to be much larger than S_u.

The critical feature of our analysis is that the three sources of information—acoustic, lexical, and sentential—are assumed to have independent influences on a restoration. That is to say, in contrast to some top-down processing models, highly constrained context does not change low-level perceptual analyses. In this model, the values of A_1 and A_3 do not change as a function of the amount of preceding sentential context. This assumption is similar to Morton's (1969) idea that the stimulus information available to the Logogen System does not vary with the amount of context information. Similarly, low-level perceptual processes do not override higher-order constraints: the values of S_c and S_u do not change as a function of the acoustic information that is available.

Figure 4 also presents the predicted results based on a quantification of this model (Massaro, 1977). In addition to the good fit of the model, the parameter estimates confirm our analysis. The value of A for a single-feature change was .76 whereas it was only .24 for a three-feature change. The value of L representing the degree of lexical constraint was .44 when the first syllable was mispronounced and .56 when the third syllable was mispronounced.

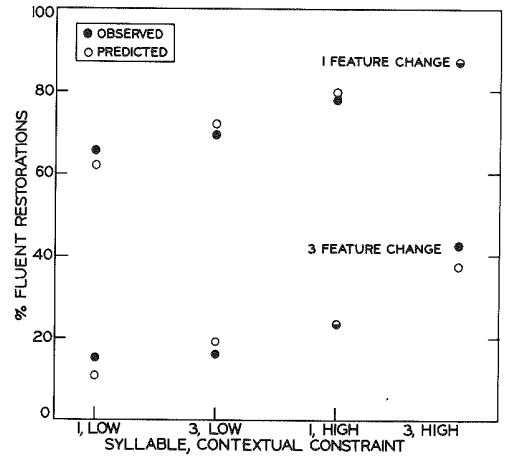

Figure 4. The percentage of fluent restorations as a function of the amount of acoustic change, the syllable changed, and the predictability of the changed word.

Finally, the degree of sentential constraint was .60 with highly predictable words and only .38 with a word nonpredictable from the preceding context.

These results support the common belief that both low-level perceptual information and high-level contextual information contribute to speech processing (and reading). What remains to be discovered is exactly how these two sources of information are combined in the abstraction of the meaning of the message. Empirically, this calls for the independent variation of the two sources of information. On the theoretical level, it is necessary to develop and test quantitative models of how the sources of information are integrated into meaning. It seems likely that many of the upcoming generation of language experiments will address this issue.

Phonetic Mediation in Reading

One persistent question in reading-related research is the extent to which the speech code is necessary for the derivation of meaning. (Phonetic, phonological, and speech codes are used synonymously to mean a code based on sound and/or articulation.) Figure 5 presents two extreme answers to the phonetic mediation question. In the first model, letters are identified and mapped into a speech code using spelling-to-sound rules, and meaning is

determined on the basis of the derived speech code. In the second model, meaning is determined from letter resolution and a speech code is not made available until after meaning has been accessed. In an earlier review, I concluded that, "although phonological mediation can be rejected in word recognition, it may play a significant role in later processing stages such as rehearsal and recoding in generated abstract memory" (Massaro, 1975a). Although a relatively large amount of research has been added to the data base, the story does not require any modification.

Gough and Cosky (1976) believe that they have some new data in support of the phonological-mediation view of Gough (1972). Subjects were asked to read aloud as rapidly as possible words that violated or obeyed spelling-to-sound rules. If phonological mediation occurs, regular words which conform to spelling-to-sound rules should be converted to a speech code faster than exception words which violate the rules. Accordingly, the time required to comprehend the word and name it aloud should take longer for words that violate spelling-to-sound rules. In support of their hypothesis, the pronunciation times for exception words averaged 27 msec longer than the pronunciation times for regular words. However, there is no assurance that differences in pronunciation time result from differences in word recognition time. The differences in reaction time could also have resulted from differences in the time for response selection and programming after the word had already been identified (see Massaro, 1975a; Massaro, 1977).

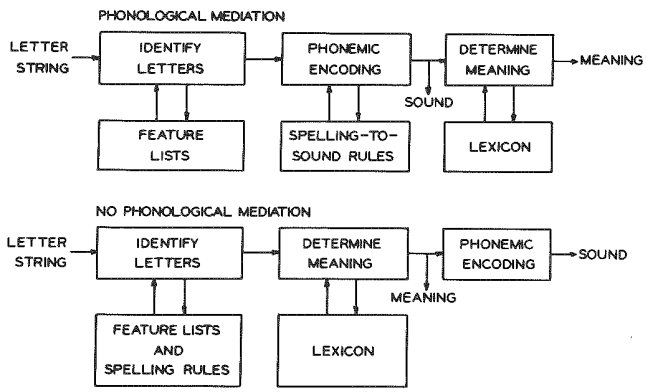

Figure 5. Two stage models of the locus of phonological mediation in reading.

Green and Shallice (1976) asked subjects to judge whether two words rhymed or whether they belonged to the same broad semantic category (animate or inanimate). On half of the trials, the words were misspelled to give homophonic pseudowords ("wise" would be spelled "wize"). Misspelling the words produced a much larger decrement in the semantic than the rhyming task. If lexical access occurs via phonological coding, there is no reason that the semantic task should have been slowed more by misspelling than the rhyming task was. Spelling-to-sound rules would have been sufficient to perform the rhyming task and misspelling should have very little effect on this process. In support of this, misspelling the words increased reaction times by only 11% in the rhyming task. Lexical access should be drastically influenced by misspelling, however, if it occurs via a visual code. Reaction times were slowed by 58 percent in the semantic task, arguing against the idea of phonological or speech recoding in lexical access and derivation of meaning. The results support other findings that demonstrate that phonemic encoding is not necessary in processing written language for meaning (see Massaro, 1975a; Massaro, 1977).

Rehearsal and Recoding

In this model, the same abstract structure stores the meaning of both listening and reading. Generated abstract memory (GAM) in the model corresponds to the working memory of contemporary information processing theory. Rehearsal and recoding processes operate at this stage to maintain and build semantic/syntactic structures. There is good evidence that this memory has a limited capacity, holding about 5 ± 2 chunks of information. For a more detailed discussion of processing at this stage, see Massaro (1975b, Chapter 27).

Generated Abstract Memory

Although generated abstract memory (GAM) is assumed to hold abstract information relative to the earlier auditory and visual memories, the nature of the information appears to be tied to the surface structure of the language. Some relevant research comes from experiments carried out with bilingual subjects (Dornic, 1975). Recall from immediate memory (supposedly tapping GAM) does not differ for unilingual and bilingual lists, whereas recall of items assumed to be no longer in GAM was poorer in bilingual than unilingual lists (Tulving & Colotla, 1970). Similarly, Kintsch and Kintsch (1969) showed that the semantic relationship between the words in different languages did not influence immediate memory, but did affect recall of items no longer active in GAM. Saegert, Hamayan and Ahmar (1975) showed that multilingual subjects remembered the specific language of words in a mixed-language list of unrelated words, but this information was forgotten when the words were presented in sentence contexts. Dornic (1975) points out that surface structure and item information are integrally related in immediate memory; subjects seldom

report translations for the words. If the items are remembered, so are the appropriate surface structure forms.

Not only is GAM tied to the surface structure of the language, it appears to be closely tied to a speech code. Given an alphabetic writing system, it may not be surprising that memory would contain a speech-code dimension since the written language has a fairly direct mapping to the spoken language. However, there is now good evidence that memory for non-alphabetic material also exploits a speech code. Chinese logograms and Japanese Kanji are non-phonetic symbols and map into speech at the level of words rather than at the level of letters or spelling patterns. It is more difficult to remember an auditory presentation of a group of letters that are phonemically similar than a group of letters that are phonemically dissimilar (Wickelgren, 1965). If a person remembers the letters in terms of a speech code, it is not surprising that similar sounding letters will be more difficult to remember. Given this result, it becomes important to ask whether similarity effects will also occur in memory for visual presentations of nonalphabetic characters.

Erickson, Mattingly, and Turvey (1972) present some evidence that GAM for Japanese Kanji is based on a speech code. Their Japanese subjects had more difficulty in immediate memory for phonetically similar or homophonic words than for semantically similar or unrelated words. Given the short-term memory requirements, it appears that the phonetic coding of the Kanji was a better medium for retention than the corresponding meaning from which the phonetic reading may have been derived.

Tzeng, Hung, and Wang (1977) asked Chinese subjects to remember lists of phonetically similar or phonetically dissimilar Chinese characters. Phonetic similarity was defined in terms of the number of characters that had the same consonant or vowel sounds in their spoken rendering. Subjects were given a list of four characters followed by 15 sec of shadowing before a serial recall of the test list. Recall was significantly poorer for the similar than for the dissimilar lists. In a second task, subjects judged whether or not a sentence was grammatical and meaningful. The judgment times were longer for the sentences composed of phonetically similar characters. The results of Erickson et al. and Tzeng et al. show that a speech code is a critical dimension in GAM for nonalphabetic characters. Given that meaning was probably accessed before or simultaneously with the speech code, it is important to note that a speech code in GAM does not imply that the meaning of written characters is derived by their sound (see Phonetic Mediation in Reading).

Memory Codes in Reading

Although phonological encoding is not necessary to derive meaning from print, there is growing evidence that visual information in reading must be transformed into another code very quickly after it is viewed. Observers in tachistoscopic tasks appear to remember very little about the visual properties of a short display presentation. Thompson and Massaro (1973) and Juola, Choe, and Leavitt (1974) varied the similarity of the two forced-choice alternatives in the Reicher (1969) task. Subjects saw a word or letter followed by two letter alternatives and reported which letter was present in the display. Performance did not differ for confusable and distinctive letter alternatives. This result indicates that observers did not maintain partial visual featural

information to aid in their choice between the letter alternatives. The subjects must have recoded the information into whole letters or words or an abstract code, preventing the utilization of partial visual information in this task (Massaro, 1975a).

McClelland (1976) presented words and letter strings in upper, lower, or mixed cases. In addition to reporting the letters themselves, McClelland asked the subjects to indicate whether the letter strings were upper, lower, or mixed case. Subjects averaged only 39% correct where chance would be 33%. Even when subjects were able to report all of the four letters in the string in a whole report, they were able to identify the appropriate case only half of the time.

More recently, Hawkins, Reicher, Rogers, and Peterson (1976) evaluated the degree to which subjects are dependent on a phonetic code in the Reicher paradigm. On word trials, the two alternatives were either homophones (cent, sent) or phonetically different (sold, cold). Observers were given either a large or small proportion of word trials that had homophone alternatives. Single-letter trials were also interspersed throughout the session. In the condition with a high proportion of homophones, Hawkins et al. found the same 15% word advantage over single letters for trials with homophone alternatives as for trials with nonhomophone alternatives. Accordingly, a phonetic code did not improve performance in this task either because subjects did not use it or because it did not contribute more information than that given by other (semantic or visual) codes. In the condition with a low proportion of trials with homophone alternatives, however, the word advantage was only 6% with homophone alternatives and 20% with nonhomophone alternatives. The phonetic code in this condition improved performance 14% relative to information given by other codes.

Why might a phonetic code have been more successful than the visual or semantic code in the Hawkins et al. study? One reason is that subjects may have had both the phonetic and semantic codes available but they were more likely to discard or forget the semantic code, especially when the phonetic code was usually successful in the low-homophone condition. A second reason might be that subjects were unlikely to recognize the whole word in the task. If correct recognition of a single letter was .7, then recognition of all four letters would be $.7^4$ or about one time in four. Given partial information about the letter string and the fragility of a visual representation, subjects might be likely to utilize a speech code. Given a phonetic code in the low-homophone condition, they would be more likely to be correct when the word alternatives were phonetically different. The fact that homophone alternatives do not disrupt performance in the high-homophone condition shows that other codes besides a phonetic code are available.

Kleiman (1975) has recently presented a nice series of experiments showing that although speech recoding is not necessary in deriving meaning from print, it is a critical determinant of storage in working memory. The work centers on the idea that the disruptive effect of concurrent shadowing in different tasks reveals the degree to which speech recoding occurs in each of the tasks. Subjects performed each of three tasks in the first experiment. Subjects saw two words side by side and had to indicate whether the words were graphemically similar, phonemically similar, or semantically similar. For example, the subject in the phonemic task would respond true to tickle-pickle but false to lemon-demon since the former rhyme but the latter do not. A "false"

response in the graphemic task was made if the words differed by more than a single letter. Each of the three tasks was performed with and without the shadowing task. It is unlikely that the graphemic task requires speech decoding, and the interference caused by shadowing should provide an index of the general disruptive effects of shadowing. The additional interference caused by shadowing in the phonemic task, which requires speech decoding, indexes the direct interference of shadowing on speech decoding. Given these measures, the degree of interference caused by shadowing on the semantic task should reveal the degree that speech decoding is present in this task. The results showed that the interference caused by shadowing was equivalent in the graphemic and semantic tasks and significantly greater in the phonemic task. This means that the amount of phonemic encoding was equivalent in the graphemic and semantic tasks and signficantly greater in the phonemic task. In a second experiment Kleiman showed that phonemic similarity did not disrupt the graphemic task. This result shows that no phonemic encoding occurred in the graphemic task. Therefore, phonemic encoding did not occur in the semantic task either, since shadowing had equivalent effects in the graphemic and semantic tasks.

In a third experiment, observers were given five-word sentences in two different tasks with and without shadowing. In the category task, subjects were given a category name and asked to indicate whether or not the following sentence contained a member of that category. In the acceptability task, the subject indicated whether or not the word string was a semantically acceptable sentence. Shadowing increased reaction time 78 msec in the category task and 394 msec in the acceptability task. Given that the category task could be performed without much working memory load whereas the acceptability task could not, the author concludes that speech recoding is important to storage in working memory but not to initial processing for meaning.

Conclusion

It seems valuable to attack reading and listening with similar methodological and theoretical forces in the framework of an information-processing model. Our concern is with how the reader and listener perform, and with the dynamics of this performance. Although the surface structures of written text and speech present unique questions to each skill, the apparent similarities in deep structure offer the hope of a single framework for understanding both reading and listening.

The author's research is supported by grants from the Public Health Service, the National Institute of Education, and the Wisconsin Alumni Research Foundation. The substantive contributions of Michael M. Cohen, David Klitzke, Gregg C. Oden, Glen A. Taylor, and Richard L. Venezky are gratefully acknowledged. Herman Bouma and Glen Taylor provided very helpful reviews of an earlier version of the paper.

REFERENCES

Allport, D.A. On knowing the meaning of words we are unable to report: The effects of visual masking. In S. Dornic (Ed.), Attention and Performance VI. Hillsdale, NJ: Erlbaum, 1977.

Anderson, N.H. Information integration theory: A brief survey. In D.H. Krantz, R.C. Atkinson, R.D. Luce, & P. Suppes (Eds.), Contemporary developments in mathematical psychology (Vol. 2). San Francisco: W.H. Freeman, 1974.

Ball, F., Wood, C., & Smith, E.E. When are semantic targets detected faster than visual or acoustic ones? Perception & Psychophysics, 1975, 7, 1-8.

Blakemore, C. The baffled brain. In R.L. Gregory & E.H. Gombrich (Eds.), Illusion in nature and art. London: Duckworth, 1973.

Blakemore, C. Mechanics of the mind. New York: Cambridge University Press, 1977.

Blesser, B., Shillman, R., Kuklinski, T., Cox, C., Eden, M., & Ventura, J. A theoretical approach for character recognition based on phenomenological attributes. International Journal of Man-Machine Studies, 1974, 6, 701-714.

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

Breitmeyer, B.G., & Ganz, L. Implications of sustained and transient channels for theories of visual pattern masking, saccadic suppression, and information processing. Psychological Review, 1976, 83, 1-36.

Broadbent, D.A. The hidden preattentive processes. American Psychologist, 1977, 32, 109-118.

Chomsky, N., & Halle, M. The sound pattern of English. New York: Harper & Row, 1968.

Cole, R.A. Listening for mispronunciations: A measure of what we hear during speech. Perception & Psychophysics, 1973, 13, 153-156.

Cole, R.A., & Jakimik, J. Understanding speech: How words are heard. Technical report, Department of Psychology, Carnegie-Mellon Univeristy, 1977.

Dornic, S. Human information processing and bilingualism. Report from the Institute of Applied Psychology. The University of Stockholm, No. 67, 1975.

Erickson, D., Mattingly, I.G., & Turvey, M.T. Phonetic coding of Kanji. Paper presented at the 83rd meeting of the Acoustical Society of America, Buffalo, New York, 1972.

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

Foss, D.J., & Swinney, D.A. On the psychological reality of the phoneme: Perception, identification, and consciousness. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 246-257.

Fujimura, O. Syllable as a unit of speech perception. IEEE Transactions on acoustics, speech, and signal processing, 1975, ASSP-23, 82-86.

Gough, P.B. One second of reading. In J.F. Kavanagh & I. G. Mattingly (Eds.), Language by ear and by eye. Cambridge, MA: MIT Press, 1972.

Gough, P.B., & Cosky, M.J. One second of reading again. In N.J. Castellan, Jr., D.B. Pisoni, & G.R. Potts (Eds.), Cognitive theory (Vol. 2). Hillsdale, NJ: Erlbaum, 1976.

Green, D.W., & Shallice, T. Direct visual access in reading for meaning. Memory & Cognition, 1976, 4, 753-758.
Hawkins, H .L., Reicher, G.M., Rogers, M., & Peterson, L. Flexible coding in word recognition. Journal of Experimental Psychology: Human Perception and Performance, 1976, 2, 380-385.
Hubel, D.N., & Wiesel, T.N. Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. Journal of Physiology, 1962, 160, 106-154.
Jakobson, R., Fant, C.G.M., & Halle, M. Preliminaries to speech analysis: The distinctive features and their correlates. Cambridge, MA: MIT Press, 1961.
Johnson, N.F. On the function of letters in word identification: Some data and a preliminary model. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 17-29.
Juola, J.F., Choe, C.S., & Leavitt, D.D. A reanalysis of the word-superiority effect. Paper presented at the Psychonomic Society, Boston, MA, November, 1974.
Keyser, S.J., & Halle, M. What we do when we speak. In P.A. Kolers & M. Eden (Eds.), Recognizing patterns. Cambridge, MA: MIT Press, 1968.
Kintsch, W., & Kintsch, E. Interlingual interference and memory processes. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 16-19.
Kleiman, G.M. Speech recoding in reading. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 323-340.
Ladefoged, P. A course in phonetics. New York: Harcourt, Brace, & Jovanovich, 1975.
LaBerge, D., & Samuels, S.J. Toward a theory of automatic information processing in reading. Cognitive Psychology, 1974, 6, 293-323.
Lane, H., Boyes-Braem, P., & Bellugi, U. Preliminaries to a distinctive feature analysis of handshapes in American sign language. Cognitive Psychology, 1976, 8, 263-289.
Lesser, V.R., Fennell, R.D., Erman, L.D., & Reddy, D.J. Organization of the Hearsay II Speech Understanding System. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1975, ASSP-23, 11-23.
Liberman, A.M., Cooper, F.S., Shankweiler, D.P., & Studdert-Kennedy, M. Perception of the speech code. Psychological Review, 1967, 74, 431-461.
Lindsay, P.H., & Norman, D.A. Human information processing. New York: Academic Press, 1977.
Marslen-Wilson, W.D. Linguistic structure and speech shadowing at very short latencies. Nature, 1973, 244, 522-523.
Marslen-Wilson, W.D., & Welsh, A. Processing interactions and lexical access during word recognition in continuous speech. Cognitive Psychology, 1978, 10, 29-63.
Massaro, D.W. Perception of letters, words, and nonwords. Journal of Experimental Psychology, 1973, 100, 349-353.
Massaro, D.W. Understanidng language: An information processing model of speech perception, reading, and psycholinguistics. New York: Academic Press, 1975. (a)
Massaro, D.W. Experimental psychology and information processing. Chicago: Rand-McNally, 1975. (b)

Massaro, D.W. Reading and listening. Technical Report, Wisconsin Research and Development Center for Cognitive Learning. University of Wisconsin, Madison, Wisconsin, 1977.
Massaro, D.W., & Cohen, M.M. The contribution of fundamental frequency and voice onset time to the /zi/-/si/ distinction. Journal of the Acoustical Society of America, 1976, 60, 704-717.
Massaro, D.W., & Klitzke, D. Letters are functional in word identification. Memory & Cognition, 1977, 5, 292-298.
Massaro, D.W., Venezky, R.L., & Taylor, G.A. Orthographic regularity, positional frequency, and visual processing of letter strings. Journal of Experimental Psychology: General, in press.
McClelland, J.L. Preliminary letter identification in the perception of words and nonwords. Journal of Experimental Psychology: Human Perception and Performance, 1976, 2, 80-91.
McNeill, D., & Lindig, K. The perceptual reality of phonemes, syllables, words, and sentences. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 419-430.
Morton, J. Interaction of information in word recognition. Psychological Review, 1969, 76, 165-178.
Nash-Weber, B. The role of semantics in automatic speech understanding. In D.G. Bobrow & A. Collins (Eds.), Representation and understanding. New York: Academic Press, 1975.
Naus, M.J., & Shillman, R.J. Why a Y is not a V: A new look at the distinctive features of letters. Journal of Experimental Psychology: Human Perception and Performance, 1976, 2, 394-400.
Oden, G.C., & Massaro, D.W. Integration of featural information in speech perception. Psychological Review, 1978, 85, 172-191.
Reicher, G.M. Perceptual recognition as a function of meaningfulness of stimulus material. Journal of Experimental Psychology, 1969, 81, 275-280.
Rumelhart, D.E. Toward an interactive model of reading. In S. Dornic (Ed.), Attention and performance VI. Hillsdale, NJ: Erlbaum, 1977.
Saegert, J., Hamayan, E., &Ahmar, H. Memory for language input in polyglots. Journal of Experimental Psychology: Human Learning and Memory, 1975, 1, 607-613.
Savin, H.B., & Bever, T.G. The nonperceptual reality of the phoneme. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 295-302.
Shillman, R.J., Cox, C., Kuklinski, T., Ventura, J., Blesser, B., & Eden, M. A bibliography in character recognition. Techniques for describing characters. Visible Language, 1974, 8, 151-166.
Sloboda, J.A. Decision times for word and letter search: A holistic word identification model examined. Journal of Verbal Learning and Verbal Behavior, 1976, 15, 93-101.
Sloboda, J.A. The locus of the word-priority effect in a target-detection task. Memory & Cognition, 1977, 5, 371-376.
Spencer, H. The visible word. New York: Hastings, 1968.
Thompson, M.C., & Massaro, D.W. The role of visual information and redundancy in reading. Journal of Experimental Psychology, 1973, 98, 49-54.
Tulving, E., & Colotla, V.A. Free recall of trilingual lists. Cognitive Psychology, 1970, 1, 86-98.

Tweney, R.D., Heiman, G.W., & Hoemann, H.W. Psychological processing of sign language: Effects of visual description on sign intelligibility. Journal of Experimental Psychology: General, 1977, 106, 255-268.
Tzeng, O.J.L., Hung,D.L., & Wang, W.S-Y. Speech recoding in reading Chinese characters. Journal of Experimental Psychology: Human Learning and Memory, 1977, 3, 621-630.
Venezky, R.L., & Massaro, D.W. The role of orthographic regularity in word recognition. In L. Resnick & P. Weaver (Eds.), Theory and practice of early reading. Hillsdale, NJ: Erlbaum, in press.
Warren, R.M. Identification times for phonemic components of graded complexity and for spelling of speech. Perception & Psychophysics, 1971, 9, 345-349.
Wickelgren, W.A. Short-term memory for phonemically similar lists. American Journal of Psychology, 1965, 78, 567-574.
Wrolstad, M.E. A manifesto for visible language. Visible Language, 1976, 10, 4-40.
Zadeh, L.A. Quantitative fuzzy semantics. Information Sciences, 1971, 3, 159-176.

Working Memory and Reading

A. D. Baddeley

The possible role of short-term or working memory in reading is discussed. A distinction is drawn between two hypothetical components of working memory -- a central executive that is responsible for information processing and decision taking, and an articulatory loop which acts as a slave system, enabling verbal material to be maintained sub-vocally. On the basis of existing evidence, it is argued that the articulatory loop plays an important role in learning to read, but is less essential for fluent reading. Experiments are presented which show that subjects may read and comprehend statements without utilisation of the articulatory loop. Conditions under which the loop will be utilised are discussed. It is suggested that these include situations where 1) judgments of phonological similarity are required; 2) retention of the surface structure of the passage is required; 3) strict word order is crucial to comprehension, and possibly 4) the rate of input of material exceeds the rate of semantic processing.

My primary research interest is in human memory, and my interest in reading stems from this. A few years ago Graham Hitch and I began a series of experiments aimed at exploring the role of short-term memory in a number of information-processing tasks. This resulted in a substantial change in our concept of the nature of short-term memory, and the resulting hypothetical system, which we term working memory, seemed to suggest some interesting hypotheses about the nature of reading. In the present chapter, I want to discuss these hypotheses and describe some experiments stemming from them.

We carried out a series of experiments which explored the role of short-term memory in tasks such as verbal reasoning, verbal memory, and comprehension (Baddeley & Hitch, 1974). On the basis of the results, we produced