Language processing is the abstraction of meaning from a physical signal such as a printed text or sequence of speech sounds. The goal of an information-processing model is to describe how language is processed, not simply what the reader or listener must know to understand language. Language processing is viewed as a sequence of internal processing stages or operations that occur between the language stimulus and meaning. The operations of a particular stage take time and transform the information in some way, making the transformed information available to the following stage of processing. In the present model the storage component describes the nature of the information at a particular stage of processing whereas the functional component describes the operations of a stage of processing. The information-processing model is used heuristically to incorporate data and theory from a variety of studies of language processing.

1. Introduction

Does it matter that I wrote this contribution rather than spoke it? Or does it make a difference that you are reading rather than listening? Or are you in fact not only reading the article but simultaneously hearing a thing spoken by the live human reader in your head? Regardless of the modality of the input, the special journal issue does not offer convenient solutions to these and other important problems in communication. We will, however, present some recent research and theory on the psychological processes involved in listening to speech and reading printed text. Our goal is to stimulate your interest and involvement in our study.

One of the persistent questions about understanding language is whether the modality of input (what might be called the true surface structure) makes a difference. The first answer that comes to mind is why should it? The purpose of all language is to communicate and understand a message (or at least to camouflage a message). Language production and understanding processes must solve the same problems in both visible and audible language. Although reading and listening may have developed independently, the processes involved may still be analogous in the same.
by that analogous biological processes have developed in convergent evolution. It is commonly
appreciated that two organisms may develop similar solutions to the problem of survival, even though
evolved independently of one another. As an example, the eye of the octopus (Oc. vulgaris) and the
eye of man (a hominid) function similarly, although they evolved independently of one another (Bakken,
1957). Following this logic, the assumption that reading and listening can be considered 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 unreasonable to assume that they are analogous rather than hierarchically
related to each other.

Band (1976) presents a similar argument and supporting evidence. Recent research on the prosody of manual-visual language (American Sign Language) indicates that analogous
processes are involved in language understanding and beyond reading and listening on every dimension that has been explored. Remarkable parallels have been found between the
processes involved in understanding spoken language and visual language. For example, the
processes involved in understanding spoken language and visual language are both influenced by the
modality-specific processes necessary to transform the sound vibrations of speech and the light
waves of print. Several other obvious differences come to mind. Spoken language comes in one ear
and goes out the other, whereas the print remains available at the beck and call of a regressive eye movement or a flip of the page. It is true that some compulsive listeners might record the message and capitalize on the sound and play option for particularly difficult sections of a spoken message. In the information-processing model presented here, however, we will draw inferences between the general modality-specific processes in language processing. Returning to our argument of analogous processes, we begin to see how the different types of visual
language and speech can interact within the same information-processing framework.

II. Information-Processing Models

Reading and listening can be defined as the abstraction of meaning from printed text and from
speech, respectively. In order to derive meaning from speech or text, the speaker or author must encode the information available at the appropriate stage of processing. Language processing can be described as a sequence of processing stages or operations that occur between the input stimulus and 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, an information-processing model was utilized for a theoretical analysis of speech perception, reading, and psycholinguistics (Massaro, 1975). This model was used heuristically to incorporate data and theory from a variety of approaches to the study of language processing. The model should be conceptualized as an organizational structure for the analysis of the art of language processing. In this paper, I will present a general overview of the information-processing model and use the model to describe and incorporate some recent research.

Figure 1 presents a flow diagram of the temporal course of reading and listening. At each stage
the system contains information and functional components. The storage component represents the information available at a particular stage of processing. The functional component specifies the procedures and processes that operate on the
storage component. The model distinguishes four functional components: feature detection, primary recognition, secondary recognition, and rehearsal. Reading the corresponding storage component represents the information available at each of these stages of processing.

III. Feature Detection and
Primary Recognition

The feature detection process transforms the energy patterns in speech into a set of features held in perceptual storage. Primary recognition evaluates and integrates these features into a percept which is then synthesized into a memory. In speech, for example, the changes in sound pressure set the ear's membrane in motion and these mechanical vibrations are transformed into a set of neural impulses. It is assumed that the signal in the form of continuous changes in vibration pattern is transformed into a set of relative discrete features. Features do not have to be relative to the magnitude of energy in a particular frequency band, but they may include information about the direction and rate of frequency change. It would be possible, for example, to have a feature detector that responds to the rising first formant transition that is characteristic of the class of voiced stop consonants.
A. Audible features

One traditional concern in speech research has been to determine the acoustic features that are utilized in perception. In terms of our model, the feature detection process places features in a brief temporary storage called the perceptual auditory storage (PAS), which holds information from the feature detection process for about 250 msec. The primary recognition process integrates these features into a synthesized percept which is placed in synthesized auditory memory. One critical question is what features are utilized and a second important question is how are all of the features integrated together. Does the listener only process the least ambiguous features and ignore all others, or are the features given equal weight, and so on? Despite the overwhelming amount of research on acoustic features, very little is known about how the listener combines the multitude of acoustic features in the signal in order to arrive at a synthesized percept.

The integration of acoustic features has not been extensively studied for two apparent reasons. The first is that research in this area was highly influenced by linguistic descriptions of speech sounds in terms of binary all-or-none distinctive features (Jakobson, Fant, & Halle, 1961). One of the goals of distinctive feature theory was to describe all of the functional differences among speech sounds by a minimal number of distinctive features of the language. Therefore, distinctive features were designated as either present or absent in each phoneme. This led to a reductionist view of speech perception, where listeners were assumed to perceive the sound as a whole, without considering the integration of multiple features. 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. In a typical study of voicing all voicing cues were made neutral except one, such as vocal onset time and then varying the remaining dimension through the relevant values. Similarly, place of articulation was studied by neutralizing all cues but one, and then varying the remaining dimension through the appropriate values. Very few experiments independently varied no voicing cues and place cues within a particular experiment so that little information was available about how these cues were integrated into a synthesized percept.
values. 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 incorrect. Oden and Massaro (1977) have used this description to describe acoustic features as fuzzy; that is, varying continuously from one speech sound to another.

In this representation features are represented as fuzzy predicates which may be more or less true rather than only absolutely true or false (Zadeh, 1971). In terms of the model, fuzzy predicates represent the feature detection and evaluation process; each predicate is applied to the speech sound and specifies the degree to which it is true that the sound is a relevant acoustic feature. For example, rather than assuming that a sound is voiced or unvoiced, the voicing feature of a sound is expressed as a fuzzy predicate.

\[ P(\text{voiced} | S_i) = 0.85 \]  

The model given by Equation 1 represents the fact that it is 0.85 true that speech sound \( S_i \) is perceived to be voiced. In terms of our model, then, the feature detection process makes available a set of fuzzy predicates at the level of PAS. In addition to being concerned with the acoustic features in perceptual 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.

As an example of the study of acoustic features, consider the dimension of voicing of speech sounds. In English the stops, fricatives, and affricates can be grouped into cognate pairs that have the same place and manner of articulation but contrast in voicing. The question of interest is what acoustic features are responsible for this distinction and how the various features are integrated together in order to provide the perceptual distinction. The integration question has not been extensively studied, however, since the common procedure in these experiments is to study just a single acoustic feature at a time. Consider two possible cues to the voicing distinction in stop consonant syllables: voice onset time (VOT), the time between the onset of the syllable and the onset of vocal cord vibration, and the fundamental frequency of vocal cord vibration at its onset. Each of these cues has been shown to be functional in psychophysical experiments when all other cues have been held constant at neutral values. However, it is difficult to generalize these results to the perception of real speech, since no information is provided about the weight that these features will carry when other features are also present in the signal. To overcome this problem it is necessary to independently vary two or more acoustic features in the signal. The results of this type of experiment not only provide information about the cue value of one feature when other features are present in the signal, but also allow the investigator to evaluate how the various cues are weighted.

B. Audible features in fluent speech

The success of finding acoustic features in perception of isolated speech sounds might lead one to expect that perception of fluent speech is a straightforward process. Sound segments could be recognized on the basis of their features and the successive segments could be combined into higher-order units of words, phrases, and sentences. However, the acoustic structure of words in fluent speech differ significantly from the same words spoken in isolation. Two solines contribute to the large variation in words in fluent speech: enunciation and psychological parsimony (Cole & Jakimik, 1977; Ross, 1975).

In fluent speech the speech articulators must assume an ordered series of postures corresponding to the intended sounds, and the articulators cannot always reach their intended targets because of the influence of adjacent movements. Coarticulation refers to altering the articulation of one sound because of neighboring sounds. The words "did it" and "said it" are pronounced as [dild] and [sid] in isolation but as [dild] and [sid] in combination because of palatalization. The alveolar stop followed by a front glide when combined produce the front-palatal affricate [t∫]. Even though a word boundary is present, the consonant "t" is not affected.
Visible features.

One of the oldest areas of reading-related research is the study of the functional cues used in recognizing printed characters. Much of this work was performed by typographers and artists concerned with the relative merits of good design and legibility in type fonts (Spencer, 1968). Although many of the early conclusions remain valid today, they are almost totally ignored in the contemporary study of letter recognition. The primary influence in extant studies has been the well-known neurophysiological finding that the responses of cells in the visual cortex demonstrate an amazing stimulus selectivity (Hubel & Wiesel, 1962). For example, there appear to be specialized detectors in the system for lines of specific size and orientation (see Blakemore, 1973, and Lindsay & Normun, 1977, for reviews of this work).

Consistent with an all-or-none response of specialized cells, psychological descriptions have centered around binary all-or-none features. Feature sets usually consist of the presence or absence of horizontal, vertical, oblique lines, curves, intersections, angles, and so on. The feature sets are typically derived from and tested against the recognition confusions of upper-case letters (see Massar, 1976, chapter 6). In contrast to the idea of binary all-or-none features, however, visible features, like other all-or-none features, may be fuzzy. Rather than have a feature 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 the idea of fuzzy information, it is important to carry out research that manipulates the degree to which a feature is present in a letter. Recently, Blesser and his colleagues have developed and studied ambiguous characters (Blesser, Shillman, Kukinski, Cox, Eden, and Ventura, 1974; Shillman, Cox, Kukinski, Ventura, Blesser, and Eden, 1974). A completely ambiguous character is one that would be assigned to either letter class with equal probability. As an example, a k can be gradually transformed into a y by continuously increasing the right oblique line below the intersection (Naja & Shillman, 1978). This work with ambiguous characters is consistent with the idea of fuzzy visible features, since each feature must be represented in terms of the degree to which it is present in the character. 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.

D. Visible features in printed text

Javan (cited in Huynh, 1968) showed that reading is much easier when the top half rather than the bottom half of a line of print is exposed. There are seven ascending letters and five descending letters and, in addition, ascending letters are about two times more frequent in text than are descending letters (Mayner and Tresselt, 1968). Therefore, it is not surprising the top half is more important than the bottom half of a line of print. It would be interesting to repeat Javan's experiment with a type font that equates the vertical extent for all of the letters. This font was actually developed by Andrew Tauer in the 1860's (cited in Spencer, 1968) and is only now being used on some computer printouts of lowercase type.

Are words perceived by way of the letters that make them up? This old and familiar question addresses whether word recognition can be described in terms of component letter recognition or whether a word is recognized on the basis of suprasegmental features without reference to the letters that make it up. If words are recognized via the letters that make them up, Caskey (1976) reasoned that the case of word recognition would be a direct function of the case of recognition of the component letters. He measured the time to name letters presented aloud and in the time to discriminate between letters in creating an index of letter legibility. Then he composed words of the 13 most legible letters and the 23 least legible letters. Empirizing these words in a naming task, he found no effect of letter legibility and concluded that letter perception does not mediate word perception. However, in addition to interpreting differences in naming times as a direct index of differences in recognition time there are a number of limitations to Caskey's study. Most importantly, Caskey (1976) did not
The advantage of words over single letters and nonwords is not incompatible with the idea that a letter is a basic perceptual unit, however (Massaro, 1975b). In the present model, the primary recognition process operates on a number of letters in parallel. The visual features read out at each letter position define a candidate set of possible letters for that position. The recognition process is not limited to feature information, but can also utilize knowledge about the orthographic structure of English spelling. The letter that is synthesized at each position, therefore, will not only correspond to the visual information that is available from feature detection and evaluation, but will also correspond to the orthographic constraints in the language. For example, consider the case in which the subject is given the lowercase string coi and has resolved just the circular envelope of the first letter and all of the last three letters. Given that $c$, $o$, and $i$ are the only letters that are consistent with the circular envelope, those are the only possible letters at this position. If the reader further assumes that the string must conform to English orthography, only one is possible since the strings coi and coi are illegal English spellings. In this case, the reader can synthesize the string coi since it is the only valid alternative. When the single letter $c$ is presented, on the other hand, the perception of the envelope does not allow an unambiguous choice among $c$, $o$, and $i$. Accordingly, the reader is less likely to synthesize the correct alternative and will be attracted only to the visual feature of the letter as the potential letter.

A. Perceptual and contextual contributions to listening

Our conceptualization of speech processing is one that is perceptually, and therefore, acoustically driven. We assume that the secondary recognition process operates by syllable on the output of primary recognition.

However, contextual constraints also play a strong influence on this stage of processing, so that both contributions must be accounted for. In describing how meaning is imposed on the spoken message, a series of recent studies has shown that abstracting meaning is a joint function of the perceptual and contextual information. 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 pheme by 1, 2, or 4 distinctive features (for example, confusion mispronounced as gunfusion, bunfusion, and sunfusion, respectively). The probability of recognizing a one-feature mispronunciation was 3 whereas a four-feature change was recognized with probability 75. This result makes apparent the contribution of the perceptual information passed on by the primary recognition process in our view some of the mispronunciations went unnoticed because the contribution of contextual information worked against the recognition of a mispronunciation. The syntactic/semantic context of the story would support a correct rendering of the mispronounced word; in other words, the perceptual information is not the crucial factor in support of the idea that mispronunciations were correctly recognized when the syllables were isolated and removed from the passage.

Co et al. (1971) reasoned that the listener should be faster at detecting a mispronunciation to the extent that a word is predicted by its preceding context. This follows from the idea that the evident way to detect a mispronunciation is to check the context against which the word is predicted.
text were the exclusive and overriding factor, we might expect subjects to replace the syntactically semantically anomalous word with the appropriate word. This did not occur, however, showing that both context and acoustic information influenced speech processing.

Marlesen-Wilson and Walsh (1978) asked observers to shadow (repeat back) spoken passages from a popular novel. The words of the passage were read at a rate of 150 words per minute. The subjects were told to repeat back exactly what they heard. 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 of three phonemic distinctive features, based on Kayser 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 restorations 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 repeated exactly, i.e., 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 Marlesen-Wilson and Walsh's study can illuminate how acoustic information and high-order context are integrated by the listener in language processing. Figure 2 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.

Marlesen-Wilson and his colleagues interpret this series of experiments as evidence against aural theories of language processing, which assume that 'vast degrees of delay before information at any one level of analysis can interact with information at a higher level' (Marlesen-Wilson and Tyler, 1975, p. 784). However, the results do show exactly such a delay. Restorations seldom occur when the first syllable is mispronounced by three features even though the word is relatively predictable given the preceding context. This means that some low-level perceptual analyses of the word, occurred regardless of the higher-order one being available and then

A second paradigm that has been used to study speech processing is the shadowing task, in which the listener repeats back the message as it is heard. It is well-known that shadowing performance improves with increases in the syntactic-semantic constraints in the message (Rosenberg & Lambert, 1974; Treisman, 1965). Recent research has been directed at how these higher-order constraints are integrated with the ongoing perceptual analyses in order to arrive at the meaning of the message. Marlesen-Wilson (1973) asked subjects to shadow prose as quickly as they heard it. Some individuals were able to shadow the speech at extremely close delays with lags of 200 msec, about the duration of a syllable or so. When subjects made errors in shadowing, the errors were 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.' In this example, that shares acoustic information with at and is also syntactically semantically appropriate in the same position in the sentence.

In another experiment (Marlesen-Wilson, 1975) subjects shadowed sentences that had one of the syllables mispronounced in a three syllable word. Subjects never restored the word, that is, repeated back what should have been said when the mispronunciations occurred in the first syllable. With mispronunciations in the second syllable and third syllable, a significant proportion of restorations occurred. If the mispronounced word was syntactically and semantically anomalous, however, restorations did not occur for any mispronounced syllable. These results indicate that restorations will not occur if the shadower does not have sufficient acoustic information and syntactic/semantic context to make the restorations appropriate. If necessary, the shadower could determine whether the first syllable was anomalous and then repeat the entire word, but the second and third syllables would not be restored.
the outcomes of these analyses were combined with the higher-order constraints. The fact that higher-order constraints in the passage influence shadowing does not mean that some analyses do not begin before others. Moreover, their view might be interpreted to mean that higher-order analyses modify the output of lower-level analyses. However, a quantitative model that assumes that both levels of analyses are functionally independent can account for these results. Figure 2 also presents the predictions of a quantitative formulation of the independence model (see Massaro, 1977, for the exact form of the model). The model assumes that the information passed on by the feature detection and evaluation process is equivalent regardless of the higher-order constraints in the passage. Therefore, it is not necessary to assume that higher-order constraints allow the subject to selectively attend to or preselectively process certain acoustic properties of the speech input. In this model, higher-order constraints do not modify the nature of low-level perceptual analyses performed on the input data.

B. Phonological mediation in reading

A persistent question in reading-related research is the extent to which the reader translates print into some kind of speech code before meaning is accessed. A similar but not identical question is the extent to which the speech code is necessary for the derivation of meaning. Figure 3 presents two extreme answers to the phonological 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 the letter sequence, and a speech code is not made available until after meaning has been accessed.

Gough and Cosky (1972) believe that they have accumulated 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 for which spelling-to-sound rules should be converted to a speech code faster than exception words which violate the rules. Accordingly, the time to comprehend the word and name it aloud should take longer for the 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, 1975), p. 263).
In order to provide evidence that differences in naming times are closely tied to processing feedforward times in word recognition time, it is necessary to perform a stage analysis of the naming task and to include a number of other independent variables known to influence specific stages on the task (Massaro, 1975a; Sternberg, 1969). Consider this task in terms of the model depicted in Figure 1. Naming a written pattern requires feature detection, priming, recognition, and secondary recognition processes and also response selection and response programming. Operations in terms of the analysis the reaction times (RT) between the onset of the written pattern and the onset of the spoken response is a composite of 5 component times:

$$\text{RT} = \text{FD} + \text{PR} + \text{SR} + \text{RS} + \text{RP}$$

where FD, PR, SR, RS, and RP represent the times for the five respective processes.

The critical components in the naming task include the time to detect the visual features and shape the visual pattern that is to be named (the pattern), attach a name (speech code) to the seen pattern, to select the appropriate articulatory program: for the speech code, and finally to program the articulators to execute the response. The response execution time would not contribute to the actual RT since the RT is measured at the onset of the naming response. One simplification of the analysis would be to divide up the RT into input and output components. In this case, the five

three operations would entail stimulus processing, whereas the last two would represent response operations.

$$\text{RT} = \text{SP} + \text{RO}$$

where SP and RO equal the times for stimulus processing and response operations, respectively.

The present concern for localizing naming time differences is at a particular stage of processing is not new; in fact, James McKean Catell (1888) provided exactly the analysis almost 100 years ago. Many contemporary investigators have cited Catell's finding that a short word can be named in less time than a single letter, with the implication that words are perceived in less time than are single letters. Catell realized, however, that the naming task included both perception time and “will-time,” as he called the time to choose a response. Relative differences in perception time were determined by using a Donders Type C reaction, in which observer was required to make a simple response such as lifting a finger off a key in response to one of many possible alternatives. In different tests the subject was asked to respond to just one of many possible letters, words, colors, and so on. In contrast to the naming task the results showed shorter reaction times for single letters than for word alternatives. Catell's interpretation appears to be still valid today. To quote from his Popular Science Monthly article in 1888:

"The time it takes to think,” “A letter can be seen more quickly than a word..."
but not pseudowords. Frederiksen-may have found much larger differences than those because backing the words in a sentence would encourage the subjects to pronounce the words via lexical access. Randomizing words and pseudowords in the Thes and Meaco study might have encouraged announcing some of the words by way of spelling-to-sound rules rather than by way of lexical access. In agreement with this interpretation, Frederiksen and Kroll (1975b) found a larger effect of word frequency on naming RTs when only words were presented in a block of trials relative to a random mixture of words and pseudoword trials.

Green and Shallice (1976) asked subjects to judge whether two words rhymed or whether they belonged to the same broad semantic category. Mispronouncing the words as homophones 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 slower here by mispronouncing the rhyming task. The fact that the rhyming task was performed about twice as fast as the semantic task shows that lexical access was not necessary in the former task, although it was in the latter. Spelling-to-sound rules would have been sufficient to perform the rhyming task, and mispronouncing should have very little effect on this procedure. In support of this, mispronouncing the words increased reaction times by only 11 percent. Lexical access should be drastically influenced by mispronouncing, however, if it occurs via a visual code. Reaction times were slowed by 16 percent in the semantic task, arguing against the idea of phonological or speech reading in lexical access and derivation of meaning. The results support other negative findings on the necessity of phonemic encoding in processing written language for meaning (see Massaro, 1976a).

V. Rehearsal and Recoding

In the present model, the same abstract structure stores the meaning of both listening and reading. Generated abstract memory (GAM) in our memory 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 5-9 chunks of information. For a more detailed discussion of processing at this stage, see Massaro (1976a, Chapter 27).

Although GAM is presumed to be abstract relative to SAM and SVM, the nature of the information appears to be tied to the surface structure of the language rather than in terms of underlying meaning that is language independent. Some relevant research comes from work experiments carried out with bilingual subjects (Dornic, 1975). This "limited capacity" rule has provided a reasonable description of the acquisition and forgetting of information in GAM (v. Massaro, 1976a, Chapter 27). A crucial question for the recording operation centers around the size of the units that are recoded. It seems unlikely that recording occurs word by word given that many words are ambiguous until later context disambiguates them a meaning from their context.

VI. Conclusion

It seems valuable to attack reading and listening with similar methodology 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 structure of written text and speech presents questions unique to each skill, the apparent similarities in deep structure offer the hope of a single framework for understanding both reading and listening.

VI. Preview of Contributions on Reading and Listening

One reason that speech has been considered primary and reading and writing secondary is the supposed uniqueness of certain speech perception phenomena. At the top of the list is the categorical perception of speech sounds. Categorical perception refers to a basic perceptual limitation in the perception of speech sounds. Certain sounds cannot be discriminated from one another unless they are, in fact, categorized differently. For example, the two sounds /ba/ and /da/ can be synthesized electronically so that they differ only along a single dimension called voicing or voice onset time (VOT, the time between the onset of the stop release and the onset of vocal cord vibration in real speech). If two of the sounds differ by a VOT of 10 msec, they will not be distinguished from each other in speech.
they are normally categorized as the same (for example, /bar/) but will be discriminated if they are categorized differently (for example, /bar/ and /pat/). In this example a 10-msec difference could be perceived (discriminated) when the two sounds are from different phonemic categories but not when the two sounds are from the same phonemic category.

The phenomenon of categorical perception has attracted renewed interest in the last five years and it would take more than a superficial reference to give it adequate coverage. The bulk of the empirical and theoretical work, however, is usually summarized. It is now generally accepted that categorical perception does not necessarily reflect a perceptual limitation in the processing of certain speech sounds, and in addition, analogous phenomena have been demonstrated with non-speech sounds. Working within this framework, Pastore develops the idea of a reference point for an important aspect in the establishment of categorical perception phenomena. Rather than studying the processing of speech sounds, however, he shows that this idea is equally applicable to the processing of alphabetic symbols.

Given an alphabetic writing system, it is only natural to expect the relationship between spelling and sound to be exploited. There is now sufficient evidence against the idea that the lexical access of printed words occurs via their sound. However, there are many other processing stages and tasks in which utilization of spelling-to-sound correspondences is important. One such task is spelling. In many situations the spelling of a word can be at least partially disambiguated by spelling-to-sound correspondences. Consider the noun and verb forms of an opinion, counsel, or recommendation. The noun pronounced with a final voiceless fricative must be advice or advise whereas the verb with a final voiced fricative must be advise or advise. Although sound does not provide a complete disambiguation in this example, at least one incorrect alternative has been eliminated in both cases. The voiceless fricative can not be spelled with a z and the voiced fricative can not be spelled with a s. In a new study of children, Hilt shows that good readers who are also good spellers, have mastered both the spelling-to-meaning and spelling-to-sound correspondences of the language. In contrast, good readers who are poor spellers show a deficit in their knowledge of spelling-to-sound correspondences but not in spelling-to-meaning correspondences. This result shows that reading and writing utilize different processes and that excellence in one does not insure excellence in the other.

Continuing the exploration into spelling-to-sound correspondences, Baron and Hodge attempt to distinguish among four processes underlying the learning of associations between printed and spoken words. The nature of the learning process has implications for both theories of reading and educational practice. In a series of experiments the authors are not only able to clarify the theoretical viewpoints; they also provide new evidence about the effectiveness of various instructional techniques.

Speech suppression did not affect performance in this condition supporting the idea that a speech code is not necessary for imposing meaning on printed text but is critical when exact wording information must be maintained. This result is consistent with our idea that GAM which maintains the surface form of a sentence is at least partially organized along a speech code dimension.

relations between the printed words and the spoken responses are critical. Readers learned spelling-to-sound correspondences only when it was the case that similar stimuli had similar responses.

Marvin, Multer, and Mills explore one of the differences in processing spoken and written language. The spoken sentence is most compelling, and the periodic cues in the passage may contribute significantly to comprehension. In reading, the sentence is presented in static form and there are no external guides to pace reading. Working within the framework of Martin's rhythmic theory of the temporal organization of speech, the authors devise a situation in which a written sentence is presented dynamically in synchrony with a spoken version. In this rhythmic presentation, each written syllable appears on the screen simultaneously with the onset of the spoken syllable. High school students learning Spanish were trained on either rhythmic or nonrhythmic presentations and the results showed that the rhythmic training facilitated the reading fluency of new sentences. These results open up a series of questions on the relationship between reading and listening and how it should be implemented in pedagogical practice.

Levy's contribution clarifies the locus of the speech suppression effect which refers to interfering memory of a written passage by requiring the reader to count rapidly or to shadow spoken material during reading. The interference supposedly occurs because speaking


References


Journal of Experimental Psychology
1968, 79, 390-402

Cognitive Psychology
1972, 4, 225-242

Psychological Review
1970, 77, 311-326

Behavioral Research and Theory
1972, 10, 387-400

Nature
1969, 223, 325-332

Science
1970, 170, 32-38

Journal of Verbal Learning and Verbal Behavior
1969, 8, 184-192

Psychological Bulletin
1968, 68, 334-344

Journal of Memory and Language
1972, 13, 121-122

Cognitive Psychology
1972, 4, 225-242

Behavioral Research and Theory
1972, 10, 387-400

Nature
1969, 223, 325-332
The purpose of this paper is to examine the perception of components of written and spoken codes of human communication from a broad perspective in an attempt to identify possible similarities. We will focus on the perception of phonemes and alphanumeric stimuli as equivalent levels of components or codes of an effective communication system, and intentionally will side-step the important, often discussed but poorly answered question whether these components represent the basic units of language.

A. An Overview

Human language normally involves two distinct sets of physical representations of its information code. One of these physical codes involves patterns of sounds varying in time, frequency, intensity, and complexity. The sound patterns are produced by the articulatory system and are perceived by the auditory system. The other code involves spatial patterns of lines, written or perceived in a temporal and spatial manner through the visual system. One level of units for the spoken language is the phoneme; whereas it is the letter for the written language (at least for Western cultures). There are many well-