An Information-Processing Approach to Speech Perception

JAMES E. CUTTING AND DAVID B. PISONI

For most of us perceiving speech is an effortless and overlooked task. When engaged in conservation one is primarily aware of tracking meaning; the sound pattern of what is heard is "linguistically transparent" (Polanyi 1964), that is, it goes largely unnoticed. The nature of this unnoticed but crucial half of language's dual structure is of particular interest to psychologists, linguists, and engineers—speech perception is the primary means of picking up linguistic information. The process of converting acoustic information into linguistic message, the underlying structure of that process, and its seemingly unusual design fascinate those who study speech perception. The process is also of particular interest to teachers and applied speech scientists; when it goes awry in the young child or adult, it is they who must try to bolster, realign, or circumvent the vocal/auditory system. The plan in this chapter is to take a small step toward mapping some emerging theoretical views of speech perception onto some of the findings in the school and clinic. The view presented, however, is not accepted dogma. It is the authors' own coalition of material from two separate subdisciplines within psychology: information processing and speech perception. Many of our colleagues may disagree with this description of speech perception processes.

In this discussion we will speak of a process in speech perception because speech perception is not instantaneous; it takes time. We will also speak of a series of stages in this process organized roughly in a hierarchical fashion. Information enters a particular stage, is transformed into something new, enters a new stage, is transformed again, and so forth until the linguistic message is understood. Between some of these stages are memory stores, temporary repositories for information that has flowed in. They help break up the dogged linearity of the auditory system. Finally, many of these stages and memory stores have limited capacity. That is, each can hold only so much information before it becomes saturated and information is lost. For those familiar with the *Zeitgeist* of cognitive psychology, these assumptions are easily recognized as hallmarks of the information-processing approach to perception (Broadbent 1965; Neisser 1967; Haber 1969), offspring of information theory and computer modeling within psychology.

First, the validity of these four assumptions will be established: (1) that speech perception is a process, (2) made up of stages, (3) involving memory stores, and (4) that the stages and memories are limited in their capacity. Second, the various stages of speech perception will be assembled into a flow diagram, both at a macrolevel including the entire speech/language system and at a microlevel including only those portions relevent to the lower levels of speech processing. Third, evidence supporting the layout and flow of information to and from each stage will be presented. Fourth, the nature of each phonetic level stage will be considered first in experimental terms and then in terms of ontogenetic development, comparative organization in animals, and possible neurological locus. Fifth, recurring issues within this information-processing approach will be broached, focusing on the phonetic level. The final step in this progression will be devoted to possible clinical applications.

Speech Perception as Information Processing

AS A PROCESS

People occasionally question the notion that pattern recognition in general, and speech perception in particular, is a process. Malcolm (1971, p. 386), for example, states that "when one recognizes a friend on the street there is usually no process of recognition. You see his face in the crowd; you smile at him and say 'Hi, John.' You do not think, 'Now where have I seen that face before?' "

Similarly, one might paraphrase and extend Malcolm's statement with regard to speech perception: You hear his voice in the crowd, speaking to you; you turn and smile and say "Hi, John." You do not think, "Now what did he say and where have I heard that voice before?"

There are three problems with this kind of refutation of cognitive processing as it is conceived here. First, recognition rarely, if ever, involves subvocal speech. Second, one need not be aware of a process for processing to occur. Third, and most important for this discussion, rapid recognition does not imply instantaneous recognition. If speech perception can be shown to take time, this will constitute strong evidence for a process. In fairness to Malcolm (1971) there is a kernel of argument underlying his statement that is the crux of current controversy in speech perception, but discussion of it—invariance—will be deferred until a later section.

Two kinds of evidence suggest that speech perception takes time. First there is an upper limit on how rapidly speech can be understood by the listener. For artificially compressed speech, comprehension can occur at rates of about 400 words per minute (Foulke and Sticht 1969; Orr, Friedman, and Williams 1965). At an average of about four phonemes per word, this rate translates conservatively into 30 to 40 milliseconds (msec) per phoneme. These high transmission rates are achieved only with considerable practice, only for brief periods of time, and even then with considerable errors. At faster rates speech melts into a patterned blur (Liberman et al. 1967). One twenty-fifth of a second per phoneme, albeit a very brief period of time, is not infinitesimally brief.

A second line of evidence, one more congruous with the information-processing approach, demonstrates that this time domain, roughly 30 to 40 msec per phoneme, is compatible with what is known about processing limitations from other research using different kinds of stimuli. If one presents a synthetic consonant-vowel (CV) syllable, say /ba/ as in *bottle*, to one ear and another syllable, /ga/, to the other ear and if the onset of /ba/ precedes the onset of /ga/ by about 50 msec, the listener will have considerable difficulty in identifying the first syllable as /ba/ (Studdert-Kennedy, Shankweiler, and Schulman 1970; Pisoni 1975a). This difficulty has been attributed to backward masking, a phenomenon in which the identification of a first-arriving item is interfered with by a later-arriving item. That is, the effect of a second stimulus masks backward in time the identity of the first by not allowing pattern recognition processes to operate on it. The domain of this masking effect seems to be about 50 to 150 msec.

MADE UP OF STAGES

To demonstrate that speech perception consists of a series of stages, one could note that the auditory system is made up of component parts: cochlea, cochlear nucleus, trapezoidal body, superior olivary complex, inferior colliculus, and certainly many cortical elements. Given our incomplete knowledge of functional neurophysiology, however, physiological and anatomical data will not be relied on for evidence of separate stages. Instead let us consider some logical requirements of the system.

First, those portions of the central nervous system responsible for speech perception can never be directly affected by an acoustic signal; they can only respond to neural events transmitted along the auditory pathway. Thus there must be at least one stage in the system that transforms acoustic signal into neural signal. Second, these portions of the processing system devoted to speech must share the auditory pathway with many other auditorily based systems, much as a single telephone user might share a party line with several other users. Thus at some level the neural signal must be general enough to be useful to systems that process not only speech but other sounds as well, such as music, environmental sounds, and infant cries. Subsequent to this level, it seems likely that the attributes of the neural signal peculiar to speech are transformed a second time, this time into a linguistic description of what has been said, then a third time into meaning (Licklider 1952).

These three transformations—acoustic signal to general neural code, general neural code to phonetic transcription, and phonetic transcription to meaning—may be only a few of those necessary in processing speech. Indeed, as it will be shown later, there are probably many more. Liberman, Mattingly, and Turvey (1972) have estimated that the transformation of a reasonably intelligible acoustic signal into a phonetic representation of speech—involving two of the transformations just mentioned—is equivalent to transforming a fourty-thousand bit-per-second signal into a forty-bit-per-second signal. This rapid thousandfold reduction of information results in a coded form of speech suitable for even further reduction by coding into meaningful linguistic units. Such magnitude of information reduction is requisite in speech perception. It is our opinion that such feats could only be accomplished by having the signal pass through and be transformed by several different processing stages.

The exact number of stages needed for the processing of speech is not known, although their general layout in a flow chart is commonly agreed on. A familiar convention used in information-processing models is to draw boxes and lines to represent stages and the flow of information between them. That enterprise will be repeated here. Drawing boxes and connecting them with lines is, as Roger Brown (1973, p. 4) says, "an odd interest, dependent, [we] suspect, on some rather kinky gene which, fortunately for our species, is not very widely distributed in the population." Nevertheless, fruitful insights and testable hypotheses arise from such ventures. In formulating these pencil-and-paper systems, however, one must be careful. The information-processing scientist must be selective but not overly economical in the number of stages postulated in that system. On the one hand, he or she must try to avoid such atrocities as Whorf's (1940) tongue-in-cheek fifteen-stage process for translating English into French; yet on the other hand, he or she cannot postulate too few stages, without regressing ultimately to Malcolm's (1971) one-stage direct-processing view of perception.

INVOLVING MEMORY STORES

Most of us have experienced the following embarrassing situation. Imagine yourself sitting across the table from a friend. Both of you are absorbed in different activities but intermittently you talk to one another. A period of silence goes by and suddenly your friend asks you a question. About the time she finishes you say, "Sorry, what did you say?" Distressingly, almost before you finish and certainly before she has a chance to repeat the question, you already know what it was. From where did the question reemerge? The answer must be that it had been stored unused in some kind of memory until the general speech processing system and consciousness gained access to it.

This kind of anecdote demonstrates that speech perception can use memosy. We, as well as Robert Crowder (this volume) hope to demonstrate that speech perception requires the use of memory. This discussion will be prefaced with a biref note on why more than one kind of memory is necessary in an information-processing approach to speech perception. Aside from the anticipated inclusion of echoic memory, short-term memory, lexical memory, and semantic memory within the general speech/language system, one must remember that speech perception is very fast yet is made up of several stages of processing. Add to this the fact that speech is a dynamic signal in which interrelations among spectral parts are constantly changing and changing at a constantly varying rate. Some portions of the signal may be very easy to process, while others are more difficult. These variations must be coped with and processed. According to an information-processing approach, the only plausible manner in which this can be done in "real time"—that is, without stepping outside the natural context and artificially slowing down the signal so that the more recalcitrant aspects of speech become amenable to analysis—is through the copious use of a series of memory stores or buffers whose contents are constantly updated and overwritten by subsequent information. Moreover, the contents of these stores must be accessible to a number of stages in the system. Just how these buffers and stages are interrelated will be developed in the next section.

LIMITED IN THEIR CAPACITY

Although the central nervous system is made up of billions of cells and quite possibly many millions of them are devoted almost exclusively to speech and language, the resources of this system are far from unlimited. Consider first the memories. Echoic memory, often thought of as a kind of an audio tapeloop that is constantly rerecorded, lasts about two seconds. Thus there is a temporal limit to the amount of information that can be stored in a fairly unanalyzed form. If the listener wishes to "play back" a speech sample just heard, it must be done quickly or its relatively unencoded form will be gone forever (Crowder 1971; Darwin and Baddeley 1974; Pisoni 1973).

Short-term memory (STM) is the conscious, largely verbal memory so well studied by psychologists. It is the memory that fails when one forgets a telephone number between directory look-up and dialing. It seems not to be limited so much by time as by amount and content; seven (Miller 1956) or, more likely, five (Broadbent 1975) unrelated items are about all this memory can hold and recycle for further analysis. The items themselves can be syllables, words, or even multiword units, but it seems that they cannot be sentences of any length and certainly not paragraphs. Thus if a linguistic message is to be understood, its gist must be quickly abstracted or it runs the risk of never being fully processed.

Echoic memory seems neurologically expensive, but short-term memory seems less so. If in fidelity echoic memory approaches the quality of tape recording, it would require something approximating the forty-thousand bit-per-second storage mentioned earlier (see also Norman 1972). While the early stages of auditory analysis could easily handle this load, one can see why man was not engineered to have an indefinitely long echoic memory: Billions of brain cells would be involved. Short-term memory, on the other hand, is likely to be of the forty bit-per-second variety and better adapted to hold the more highly coded linguistic message. It is limited in its capacity nonetheless, whether by neurological design or by evolutionary caprice. Only lexical memory (the dictionary-in-the-head) and semantic memory (that used for comprehension) are thought to be functionally unlimited in their capacity. Their lack of limits make them neurologically expensive and in a pneumatic fashion may force capacity constraints on other parts of the system.

The most important limitations, however, are not those placed on memories, but those placed on the entire system by attention. It appears that one can pay attention to, roughly, only one thing at a given time (Broadbent 1958, 1971). Thus from all inputs from all perceptual systems only one source of information can ride high within consciousness. That is, in an information-processing analysis of attention, only one source of input can survive attentional selection. The locus of this squeeze on inputs has been the source of controversy for twenty years, but it now appears that attentional selection occurs quite late in the system, after perceptual processing (see, for example, Shiffrin, Pisoni, and Castaneda-Mendez 1974). This fact is important in speech perception because, as noted in the introduction, one often is aware of attending to the meaning of a discourse without direct awareness of sound pattern. Attentional constraints appear to play a role in the "transparency" of sound. Moreover, if attention is subsequent to perception, the representation of speech in the what-did-you-say situation cited previously is likely to be in a more highly coded form than that of echoic memory. The message was processed and simply awaited attentional focus. Full awareness of the sound pattern of speech may be possible only when the listener has disengaged himself/herself from meaning as when one listens to a conversation spoken in a language that one does not understand or to the first babblings of a young child.

An Information-Processing Model of Speech Perception

PRODUCTION AND PERCEPTION TOGETHER

Before putting these assumptions together into a model of speech perception, it is necessary to establish a conceptual framework for it. De Cordemoy (1668) first postulated a connection between the perception and production of speech, but only since Lashley have psychologists taken this notion seriously. Lashley (1951, p. 120) appealed to parsimony: "The processes of comprehension and production of speech have too much in common to depend on wholly different mechanisms."

Some of the processes thought to be held in common between perception and production are shown schematically in figure 1. In producing speech, for example, one starts with some conceptual representation, coded in "mentalese" (Fodor, Bever, and Garrett 1974), and moves through a series of at least four other stages until the acoustic structure of speech is reached. Thus at the two ends of the process are meaning and sound. Between them are stages of deep structure, surface structure, and phonetic structure and a series of transformational processes—semantics, syntax, phonology, and speech. This system can easily be elaborated,

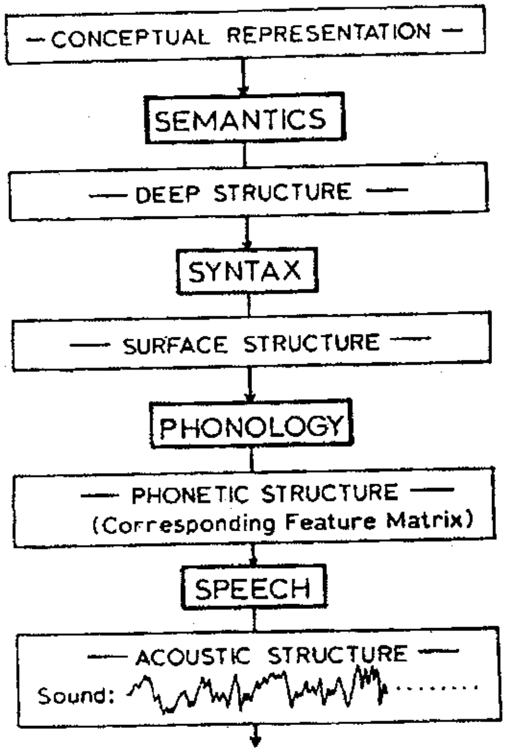

Figure 1. Serial organization of some stages of speech production. Arrows can be reverse for an approximation of the process in speech perception. (Adapted from Liberman 1970.

either at the conceptual end (see Fodor, Bever, and Garrett 1974, p. 391) or at the speech end (see Cooper 1972, p. 34). The importance of this display, however, is that the arrows can be reversed in this representation of the speech production process to achieve a fairly accurate conceptualization of speech perception.

PARALLEL PROCESSES

What is missing in this conceptual display of speech production and perception are the multiple interrelations between and simultaneity of operations within these stages. While the stages are likely to be serial in some respects, they must also be parallel. This is not quite like having one's cake and eating it too. Decisions at one level can be made on the basis of preliminary information sent up or down the system from another level, rather than each stage waiting for ultimate decisions. Again, one needs not be aware that these intermediate decisions are being made; one needs only to be aware of the ultimate outcome. In such a dynamic system there must be careful executive monitoring of parallel processes so that each stage does not act on misinformation. Improper monitoring may result in occasional metatheses, or spoonerisms, in speech production (Fromkin 1971; MacKay 1970) and undetected errors of pronunciation in speech perception (Cole 1973; Marslen-Wilson 1975). In summary, the hierarchical representation of the stages of processing represented in figure 1 is probably not wrong, but it is misleading. Perhaps a better organization, at least from an information-processing point of view, is a more heterarchical one proposed elsewhere (Pisoni 1975b, forthcoming-b) and shown in figure 2.

This is a macrolevel model of speech perception that includes the entire speech/language system. It allows, in real-time operation, the simultaneous functioning of phonetic, phonological, lexical, syntactic, and semantic processes to derive a linguistic representation of a sentence. Its advantage is that it is fundamentally a dynamic approach to speech and language perception rather than a more tem-

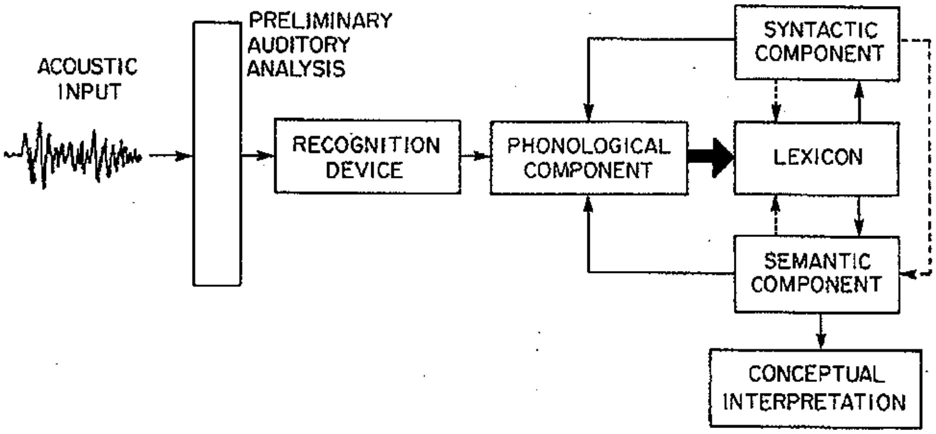

Figure 2. Functional organization of the components of the speech perception system.

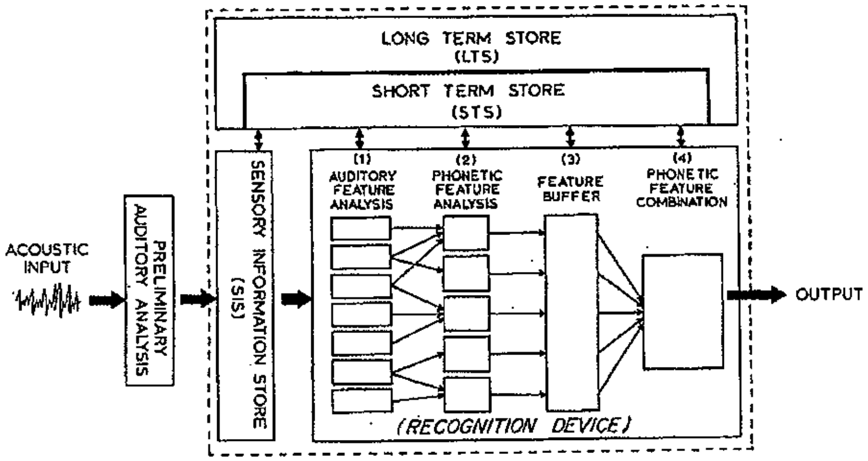

Figure 3. Functional organization of the phonetic recognition component of the speech perception system (from Pisoni and Sawusch 1975).

plate-matching, taxonomic one (Massaro 1975); it allows language to act as a "support system" for speech perception (Bransford and Nitsch, this volume).

Since the brunt of this paper is directed toward speech perception rather than language perception, the recognition device, shown in figure 2, will be further elaborated and described in a microlevel model. This device handles only the phonetic recognition process. It is shown as Pisoni and Sawusch (1975) first conceived of it in figure 3. According to this diagram, the acoustic input enters the system and undergoes a series of transformations. First it enters a stage called *preliminary auditory analysis*, where acoustic energy is transformed into neural energy, preserving to a large degree the time, frequency, and intensity relations among the components of the signal. (The term *acoustic* refers to the signal before it enters the ear; *auditory* refers to the signal after it is transformed by the ear into mechanical and neural impulses.) The information then enters a *sensory information store*. Many information-processing models call this stage preperceptual auditory storage (Massaro 1975; Massaro and Cohen 1975; see also Neisser 1967). This storage is thought to be very brief in nature, on the order of 50 to 250 msec, and is thought to account for, among many other phenomena, the backward masking results discussed earlier. The auditory system does not appear to be able to resolve separate information within a smaller time domain than 20 to 40 msec (Stevens and Klatt 1974; Hirsh 1974). This smearing or integration of the signal occurs in preliminary auditory analysis. Preperceptual auditory storage appears to reflect a "time window" or perceptual moment (Allport 1968) that travels along the acoustic signal. The window is not sharply defined; instead it has graded edges that may extend the integrating field to as much as 100 msec or more in some circumstances.

Information is then transmitted to the recognition device and undergoes a series of at least four stages of analysis. These stages, and the previous registration of the acoustic signal within the sensory information store, are mandatory and not under the conscious control of the listener.

Within the recognition device are stages of *auditory feature analysis* and *phonetic feature analysis*, a *phonetic feature buffer*, and a mixer in which the coded signal undergoes *phonetic feature recombination.* In stage one of the recognition device auditory features of the speech signal are recognized in parallel by a whole system of units whose sole job is to parse the incoming information, looking for prominent auditory features. Following Stevens (1975), some attributes that might be included are the presence or absence of rapid change in the spectrum—information that can aid in the recognition of consonants versus vowels; the direction, extent, and duration of change within a portion of the spectrum—information that can aid in distinguishing consonants from one another; other factors such as the frequency range, duration, and intensity of noise—information that can aid in the distinguishing of fricatives from one another (Hughes and Halle 1956; Gerstman 1957); and the relative onset of periodic and aperiodic portions of the signal (Lisker and Abramson 1964; Lisker 1975), The output of auditory feature analysis is some combination of the (possible) features present in the signal, which is then sent on to the next stage of processing (see Sawusch 1976).

Stage two concerns phonetic feature analysis, where a complex of decision rules maps the many auditory features onto phonetic features. It is within this stage that neural signal becomes language. It is presumed that this stage has knowledge of articulatory constraints of the vocal tract. The output of this many-to-one and one-to-many mapping is a set of abstract phonetic features sent on to the third stage.

Stage three, the feature buffer, is a form of memory not talked about previously. It is simply a holding bin that preserves the phonetic feature composition of the particular syllable being processed. A feature buffer is needed here because it cannot be assumed that all phonetic features are processed at the same rate. Moreover, some memory is needed to preserve and maintain phonetic feature information independently for subsequent linguistic processing, particularly for phonological processing, while prior stages begin to process later-arriving material. This holding bin is intimately related to short-term memory.

Stage four, the final stage within the recognition device, is a mixer for the recombination of phonetic features. It is at this stage that clusters of phonetic features are assembled into a phonetic string. This time-tagged distinctive feature matrix is sent on to higher levels of linguistic processing that modify and extend it through phonological, lexical, syntactic, and semantic analyses.

Before being sent to higher levels of analysis, information from any of these four stages and from the sensory information store can be placed in short-term memory, where the listener first has control over it and can selectively rehearse, encode, or make decisions about it. Long-term memory (LTM) is also assumed to be made accessible during the recognition process; and consistent with recent accounts of their relationship, short-term memory is thought to be simply that portion of long-term memory temporarily activated (Bjork 1975; Shiffrin 1975). Long-term memory includes episodic memory (Tulving 1972), which stores episodes of personal history according to spatial and temporal tags, and more important, semantic and lexical memory (Collins and Quillian 1969; Miller 1972, this volume), which are thought to be those portions of long-term memory "necessary for the use of language. [They are the] mental thesaurus, organized knowledge a person possesses about words and other verbal symbols, their meanings and referents, about relations among them, about rules, formulas, and algorithms for the manipulation of these symbols, concepts, and relations" (Tulving 1972, p. 386). Although not overtly marked in figure 3, STM and LTM must also be able to gain access to the higher levels of the speech/language system (the lexicon and syntactic rules) diagramed in figure 2.

The models in figures 2 and 3 are preliminary. A number of revisions will surely be needed. Nevertheless, their current form provides a convenient framework for many of the results of the past twenty-five years in speech perception. Moreover, these schemata suggest future research questions about speech perception in the normal population and perhaps experiments with listeners belonging to special populations. First, however, let us consider evidence that supports this general model.

Language Contributions to Speech Perception

Consider the macrolevel model presented in figure 2. Evidence from many sources demonstrates contributions of higher order components to speech perception. Perhaps the most direct proof comes from the many studies of the perception of speech under noise. Here "redundancy" of the speech pattern makes otherwise unintelligible sections of speech considerably more intelligible. Within the context of this information-processing approach, this redundancy gain can be attributed to contributions of higher levels of processing in the speech/language system. Bransford and Nitsch (this volume), for example, discuss the contribution of a conceptual component to the perception of speech under noise.

Syntactic and semantic aids to speech perception are shown in the results of Miller, Heise, and Lichten (1951) and Miller (1962). The intelligibility of isolated words at a signal-to-noise ratio of 0 decibels (db)—that is, when the signal and noise have the same amplitude—was only about 40 percent. Those same words, however, when they appeared in sentences under noise, were intelligible nearly 70 percent of the time. Thus when the speech signal is impoverished, the higher level components of the system help to distinguish the important features. How exactly this is done is not known, but many have suspected that linguistic context reduces the number of alternatives. A related effect can be seen in the results of Pollack and Pickett (1963, 1964) where the intelligibility of excerpts from conversation is shown to be directly related to the duration of these excerpts and to the number of words contained in them.

Syntactic and semantic contributions can be isolated, demonstrating separate effects. Again presenting speech in noise, Miller and Isard (1963) found that ungrammatical, largely meaningless strings of words were more difficult to repeat than equally meaningless grammatical strings. These sentences in turn were more difficult to repeat than meaningful grammatical strings. Thus syntax improves speech perception and the addition of an easy to interpret semantic component improves it even further.

Lexical memory, that part of LTM concerned specifically with meanings of words, also plays an important role in perceiving speech. Varying signal-to-noise ratios, Howes (1957) and Savin (1963) have shown that word recognition for common words is very different from that for rare words; common words are recognized and identified when presented in noise 20 db more intense than the corresponding threshold level for rare words. This phenomenon is attributed to the fact that common words are readily accessible in lexical memory. Rare words, on the other hand, subsist in musty, seldom-visited corners of the lexicon. The number of possible alternatives in an ambiguous situation demonstrates a similar effect. Miller, Heise, and Lichten (1951) varied the set size of words to be played under noisy conditions and found that at a signal-to-noise ratio of −12 db words from a set size of 256 items were about 10 percent recognizable, those from a set size of 32 items were about 40 percent recognizable, and those from a set size of 4 items were about 70 percent recognizable. The frequency of a word and the number of possible alternatives in a situation of signal in noise clearly can play a vital role in speech perception.

Phonological cues are also a great aid. The phonological component provides information about the sound structure of a given language, and this information is imposed on the phonetic feature matrix to derive a phonological matrix. Some aspects of this component are universal and some are language specific. General syllabic form and information about intonation contour, or prosody, are also processed at this stage. (But see Crowder, this volume.)

Stress pattern, one aspect of prosody and a speech variable processed by the phonological component, can be a potent cue in word identification. Kozlowski (in press) has used this fact as a tool in the study of tip-of-the-tongue (TOT) phenomenon (Brown and McNeill 1966). The TOT state is a tantalizing mental condition: the individual searches for a word, knows it is there, but cannot obtain it from lexical memory. Certain attributes of the word, however, such as syllabic structure and initial and final letters, can be recalled. Kozlowski explored the nature of this lexical wraith. He presented his subjects with definitions of rare words thought likely to be in a college student's recognition vocabulary but not in their active vocabulary. When his subjects declared themselves to be in the TOT state, he presented them with an auditory cue. This cue was a severely distorted version of the target word in which low-pass filtering essentially removed from the speech signal everything except fundamental frequency, some gross aspects of intensity envelope, and general syllabic structure—hence only the stress pattern of the target word remained. The end result sounded as if an individual were talking through a pillow. Kozlowski found this cue potent enough that his listeners perceived the target word in 35 percent of all TOT states. False cues, filtered words that were not the correct answer, provided only 15 percent "recognition" of the TOT word, a rate that may reflect spontaneous remission of the TOT state. Phonological cues clearly are important to the perception of speech; having available the meaning of a word but not its precise phonological form, one can often obtain the word given only its stress and syllabic patterns.

Before turning to the phonetic level model of figure 3, we should add that one more extraphonetic component aids the perception of speech; it is the speaker's facial display (Erber 1969, 1975; O'Neill 1954; Sumby and Pollack 1954). This component, not shown in figure 2, is often overlooked by the basic speech researcher. Erber (1969), for example, notes that listeners with normal hearing who try to understand words spoken under intense noise (signal-to-noise ratios of −20 db) may not be able to identify any words. If the same words are given to these listeners under the same noise conditions but with the opportunity to observe the speaker, their identification improves to as much as 60 percent. Not surprisingly, much of this gain is attributable to visual access to cues of place of articulation (Binnie, Montgomery, and Jackson 1974), a feature that is often highly masked by a noisy context (Miller and Nicely 1955) but carries a wealth of linguistic impact. Erber (1975) notes that for the hearing-impaired individual most of speech perception is a task very similar to this latter experimental situation. (The fact that conference participants came to Belmont and gathered around a table rather than engaging in one large conference telephone call can be taken as evidence for the importance of visual cues on normal speech perception.)

Note that the particular layout of stages in figure 2 allows nearly maximum interaction between phonological, lexical, syntactic, and semantic components. This arrangement was chosen to stress the heterarchical structure of the speech/language system. The pieces are both semiindependent and intersupportive. This plan also allows for the hedging of bets. The role of the lexicon has been a controversial issue among both linguists (Fodor, Bever, and Garrett 1974) and psychologists (Miller, this volume). A major question concerns where individual isolated words fit into the general scheme of syntax and semantics in the process of generating and perceiving sentences. Since there appears to be no clear answer as yet, the lexicon has been placed near the middle of everything in this model. In linguistics, current views of transformational grammar take this approach as well (Bresnan 1976).[1]

---

[1] At this point the reader may wonder why aspects of language that overtly have nothing to do with speech perception—at least as that subdiscipline has come to be known—have been dealt with so extensively. Specifically, one may ask why the research with speech under noise has been considered. The response to the first point is twofold. First, an information-processing account of a system as complex as speech perception demands thoroughness. The whole system must be considered; without a holistic approach, a proper phonemic description is not possible (Chomsky 1964). Second, language perception is a dynamic, whirling process for which speech perception, in most listeners, is the linchpin. Speech and language are not easily divisible in a working system, just as a whee and its hub are not easily divided in a moving vehicle. To consider speech without regard for the higher processes of language is, if not an empty pursuit, certainly one that wil mislead both the basic and the applied researcher. Our response to the second point i similar in tenor. Speech perception under conditions involving background noise—be i patterned, white, or shaped—is speech perception as it is accomplished every day. Speec and noise are as natural in combination as speech and language.

Phonetic Perception and Information Processing

Now to speech perception, or, as it might properly be called, phonetic perception. Evidence that will be considered in terms of the information-processing model is fairly recent. Results from many experimental paradigms appear to converge on the likelihood of a model such as that shown in figure 3 (see, for example, Studdert-Kennedy, Shankweiler, and Pisoni 1972; Day and Wood 1972; Wood 1974, 1975; Wood, Goff, and Day 1971; Pisoni 1973, 1975a; Cutting 1974, 1976). In general these results point to the facts that speech is treated by the listener as a multidimensional display, some attributes of which are auditory with little bearing on language, and some which are phonetic and integral to the speech code; these auditory and phonetic attributes appear to be coded differently in memory and may be established in different parts of the central nervous system; and auditory processes are logically prior to phonetic processes, but in most situations the two go on in parallel.

EXPERIMENTAL EVIDENCE WITH NORMAL ADULTS

After preliminary auditory analysis, the signal is transmitted to a sensory information store. Some of this information is then shipped to echoic memory (Neisser 1967; Crowder and Morton 1969), which is viewed as a component of short-term memory. Placing this storage early in the system has the advantage that the listener can "postpone classification of some items momentarily, recheck his or her categorization of others, and, generally, transcend the strict pacing" imposed by the temporally linear aspect of audition (Crowder 1972, pp. 254–255). What resides within this memory store is an auditory code of the input particularly well suited to prosodic features and certain speech segments. All speech segments, however, are not equally suited for this code. Stop consonants, for example, appear to be considerably less amenable to such storage than vowels (Crowder 1971, this volume; Pisoni 1973). Moreover, the difference between the two phoneme classes does not appear to be related to phonetic coding but rather to the fact that rapidly moving transients such as those found in stop consonants cannot be laid down in echoic form as easily as the more steady-state attributes of vowels (Darwin and Baddeley 1974). While certain auditory properties of stop consonants can be obtained by short-term memory and consciousness (Barclay 1972), the more typical situation is one in which only the form of the vowels and perhaps the fricatives can be extracted from the sensory information store. Thus much of the unanalyzed signal may bypass the echoic portion of short-term memory and be transmitted directly from sensory information store to auditory feature analysis. This "bypassing" is a signal-dependent process. What can be stored in echoic memory is stored; what cannot be stored is not.

The next stage of processing is auditory feature analysis, which has recently assumed a larger role in accounts of speech perception (Stevens 1972, 1975). This stage may not be vital for vowels and other relatively easy to process segments. An auditory husk may be available for the taking in echoic memory for up to 2 sec before one needs to transform the signal into a more parsimonious phonetic feature code. In fact, steady-state vowels might possibly be directly coded into phonetic form from the echoic portion of STM. As it is conceived here, however, the usual course of events demanded by the speech signal, particularly by those portions that will be perceived as consonants, entails auditory feature analysis.

Perhaps the best experimental evidence supporting the existence of this auditory feature detection stage comes from a paradigm recently imported from vision research (Blakemore and Campbell 1969). It is called selective adaptation. Although its original purpose was to rally support for a direct phonetic feature detection model of speech perception (Eimas and Corbit 1973; Eimas, Cooper, and Corbit 1973), recent evidence using this paradigm supports an auditory feature detection view (Tartter and Eimas 1975; J. L. Miller 1975; J. L. Miller and Eimas 1976; Pisoni and Tash 1975; Cutting, Rosner, and Foard 1976; see Cooper 1975 for a review). In this experimental situation, the listener is presented with dozens, perhaps even hundreds, of tokens of the same utterance and then asked to identify members of an array of stimuli. Results show that previously ambiguous items in the array, those at the boundary between two categories of stimuli, are afterward identified as unambiguous examples of the stimulus category opposite to that which has been adapted. This result is important because it suggests how speech perception might be aided by opponent process binary devices relatively early in the information processing system. The phenomenon is complex, thus it behooves us to discuss first two other phenomena—categorical perception and chromatic afterimages.

Categorical perception is a peculiar, nonlinear mode of perceiving typically associated with stop consonants. For example, if an array of seven stimuli from /ba/to/da/, shown schematically in the top left of figure 4, is randomized and each token is presented many times, the listener usually identifies Stimuli 1 through 3 as /ba/ and Stimuli 5 through 7 as /da/. Stimulus 4 is identified as /ba/ about 40 percent of the time and /da/ about 60 percent of the time. When these items are presented for discrimination listeners/find it very difficult to tell the difference between Stimuli 1 and 3, for example, or between Stimuli 5 and 7, but they have no difficulty discriminating Stimulus 3 from Stimulus 5. This set of results is interesting because the seven stimuli in this /ba/to/da/ array differ from one another in equal acoustic increments in terms of the amount of difference in starting frequency of their second formant transitions. That is, Stimuli 3 and 5 are no more different than Stimuli 1 and 3 or Stimuli 5 and 7 by this acoustic criterion. Typical results of the identification and discrimination tasks are shown in the lower left panel of figure 4 (see Liberman et al 1957; Studdert-Kennedy et al. 1970; Pisoni 1973).

Now let us jump temporarily to the perception of color. A well-known phenomenon is the chromatic afterimage. A viewer who stares at a patch of blue for 15 to 30 sec and then stares at a blank white wall illuminated to the same degree will see a patch of yellow whose contour conforms to the original blue patch. Blue is at some level of analysis the opposite color from yellow; white is chromatically neutral. Staring at the blue patch fatigues the blue receptors in the retina and sets

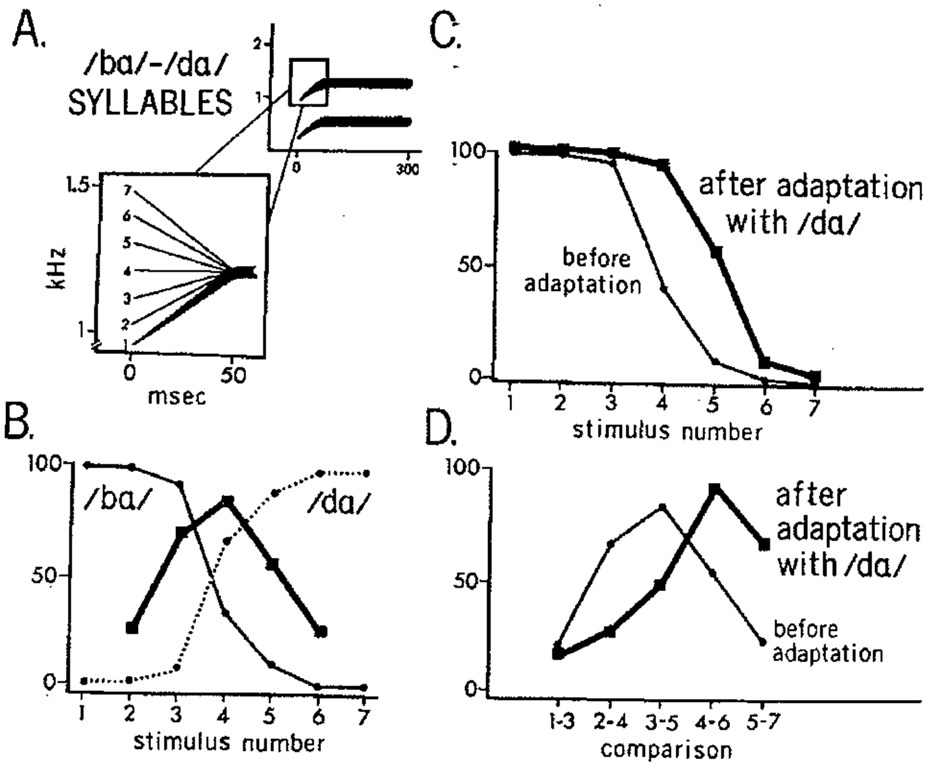

Figure 4 (a). Schematic spectrograms of an array of stimuli from /ba/ to /da/, (b) identification and discrimination functions for that array, (c) and (d) those functions before and after adaptation with Stimulus 7, /da/.

up a temporary imbalance in the opponent processing system for color. Presented with a neutral stimulus, a white wall, the viewer then sees a color that is unambiguously yellow.

In the speech domain, at some level of analysis /ba/ can be thought to be the opposite from /da/ in that it contrasts in place of acticulation. The listener cannot stare at or even listen to a normal speech syllable for any length of time. However, if a syllable such as /da/ is presented over and over, thus replenishing the auditory trace, both the identification and the discrimination functions for the set of stimuli are temporarily shifted toward /da/, as shown in the right panels of figure 4. The analogy between this effect and chromatic afterimage breaks down at this point. The adaptation effect for speech syllables is not in the cochlea (the counterpart to the retina) but farther along in the processing system and at least partly beyond the point in the auditory pathway where the two ears converge (Eimas, Cooper, and Corbit 1973).

Note that these /ba/to/da/ stimuli differ in what Whitfield (1965) has called "auditory edge." Those stimuli identified as /ba/ have a rising second formant transition; those identified as /da/ have a falling second formant transition. It is a bit unusual that the boundary between two stop consonants in CV syllables falls at the level "nonramped" second formant onset, but this case is not unique and probably many other such boundaries can be seen as a variation on this theme. It would appear that in the adaptation situation where /da/ is the adapting stimulus, Stimulus 5 rather than Stimulus 4 is now "perceived" at Stage 1 of our model to have a zero-sloped transition. By this account, it is the auditory feature analysis stage of our model that has been temporarily affected. Arguments for this interpretation instead of a direct, phonetic feature detector interpretation are complex and the reader is referred to Cooper and Blumstein (1974), Pisoni and Tash (1975), Tartter and Eimas (1975), J. L. Miller (1975), Cutting (1977), and Cutting, Rosner, and Foard (1976).[2]

Auditory features from Stage 2 are mapped in a many-to-one and one-to-many fashion onto phonetic features in Stage 3 of the recognition device. This mapping appears to be accomplished partly with reference to invariant information in the signal and partly with regard to knowledge about articulation and its acoustic consequences. How such mapping is accomplished remains one of the recurring unknowns that will be discussed in the next section.

A many-to-one mapping is necessary because a phonetic feature value, such as the voicelessness of /p/ in English, can be cued acoustically in many different ways. In syllable-initial position, for example, this value can be cued by cutback in the first formant transition (Liberman, Delattre, and Cooper 1958) and by delay in voice onset time (VOT) (Lisker and Abramson 1967). This distinction can be cued in intervocalic position by the silent interval between the offset of the previous syllable and the onset of the target syllable (Liberman et al. 1961). In syllable-final position it can be cued by the duration of the previous vowel (Raphael 1972). A one-to-many relationship is also needed since, for example, a single burst can cue different stop consonants (Liberman, Delattre, and Cooper 1952; Schatz 1954). The phonological component further maps phonetic features onto phonemes in a second type of many-to-one and one-to-many fashion not shown in figures 2 and 3.

Support for the existence of a feature buffer (Stage 3 of the recognition device) and a store in which these features are recombined (Stage 4) comes from dichotic listening tasks. If /ba/ is presented to one ear and /ta/ is presented simultaneously to the other ear, the listener often reports hearing a syllable not presented, and most often that syllable is /da/ or /pa/ (Halwes 1969; Cutting 1976). For instance, for the response /da/, it appears that the voicing feature value for /ba/ is perceptually combined with the place of articulation feature value for /ta/, and /da/ results. This is an example of perceptual synthesis of a new syllable from the phonetic feature values of the stop consonants presented to opposite ears. This combination appears to be phonetic because variation in the carrier vowel of the syllables appears to have little effect on the frequency of such "blends" (Studdert-Ken-

---

[2] The assumption that adaptation effects were due to "fatigue" of phonetic feature detectors appears to have arisen from the assumed equivalence of the term *feature* as in phonetic features (regarding distinctive features of Jakobson, Fant, and Halle 1951) and as in feature detectors as the concept was borrowed from the vision literature.

nedy, Shankweiler, and Pisoni 1972); for instance, /bi/-/tu/ pairs appear to yield as many /d/ and /p/ fusion responses as /bi/-/ti/ pairs. These fusions, or phonetic feature value combinations, appear to occur in Stage 4 of figure 3. Stage 3, the feature buffer, would probably contain on a /bi/-/ti/ trial, for example, all phonetic feature values of the two stop consonants—voiced and voiceless manners of production, and labial and alveolar places of articulation. Since a stop consonant cannot simultaneously be voiced and voiceless, or labial and alveolar, only one feature value of each can be combined in a response.

EXPERIMENTAL EVIDENCE WITH YOUNG INFANTS

A one-month-old infant does not possess any lexical, syntactic, or semantic processes or phonological processes that correspond to any degree to those of adults. Thus the infant affords the opportunity to observe the workings of the model presented in figure 3 at some stage of development without the necessity of considering the whole system shown in figure 2. In earlier days many would have referred to the "prelinguistic" child (Kaplan and Kaplan 1971). In some sense, however, this turns out not to be true. The work of Eimas (1974, 1975, et al. 1971), Moffitt (1971), and Morse (1972) has shown that infants are quite sensitive to speech distinctions. They can discriminate phonetic distinctions such as those between /ba/-/pa/ and /ba/-/da/, but they cannot distinguish between members of the same phonemic category. These results are parallel to those for adults in categorical perception (see, for example, Mattingly et al. 1971). Moreover, young infants can distinguish between the initial liquid phonemes in /ra/ and /la/ (Eimas 1975) better than adults in cultures where this distinction is not phonemic (Miyawaki et al. 1975). It would appear then that young infants come equipped with a capacity to discriminate the relevant acoustic attributes that underlie many phonetic features.

The question remains, however, whether these infants are perceiving speech in a linguistic sense (Stevens and Klatt 1974). In terms of our information-processing model, where, for example, does such processing occur? A few years ago it was believed that such discriminations must be accounted for on the basis of phonetic feature analysis (see, for example, Cutting and Eimas 1975). Today that belief is in question. The brunt of the evidence supports the view that auditory feature analyzers determine the results. It was previously claimed that the experiments on selective adaptation appear to work at a stage prior to phonetic analysis. Given that position, it is suggested that the infant studies indicate perception at the same stage, that is, these results indicate not so much that young infants perceive speech as they indicate that they can perceive speech-relevant dimensions of an acoustic signal (Jusczyk et al., 1977) that they will later apply to the process of speech perception. As Roger Brown (1973, p. 37) noted, such perception in infants is "only linguistic by courtesy of its continuity with a system which in fully elaborated form is indeed . . ." speech perception.

An important aspect of infant research on speech perception that is missing at this point is the performance, by the infant, of many-to-one and one-to-many mapping of acoustic features onto phonetic features. Fodor, Garrett, and Brill

(1975) have taken a small step in this direction using older infants. They found that four and five-month-old infants, like adults, perceived the phonemic identity of consonants in the syllables /pi/ and /pu/ as different than in the syllable /ka/, yet in all three syllables the voiceless stop consonant is cued by similar acoustic information (Schatz 1954). Until such evidence for one-cue-to-many-phoneme mappings can be assembled more fully, along with corresponding many-to-one results, the infant data may be regarded as indicating speech-relevant perception rather than speech perception.

EXPERIMENTAL EVIDENCE WITH ANIMALS

Further support for the allocation of sophisticated analyses to auditory feature processing stems from some recent studies with rhesus monkeys and chinchillas. Their discriminations of synthetic speech syllables differing in place of articulation (Morse and Snowden 1975) and identifications along voice onset time (Kuhl and Miller 1975) look suspiciously like categorical perception, although the data are not complete. Although these results may be subject to range effects (see Parducci 1974; Waters and Wilson 1976) and some speech dimensions may not be perceived categorically by animals (Sinnott, et al. 1976), it is clear that the once firm base of empirical data thought to be indicative of phonetic perception may actually be a data base supporting the existence of a sophisticated auditory feature analysis stage of processing (see Cutting and Rosner 1974; Cutting Rosner, and Foard 1976; Cutting 1977; J. D. Miller, et al., 1976; Pisoni 1976). These animals, and perhaps the human infants as well, do not perceive speech as language but as a multidimensional complex of acoustic events.

Embedded in this view is the assumption that underlying categorical perceptions are auditorily based rather than phonetically based decision processes; that is, categorical perception may be accounted for at Stage 1 of the model rather than Stage 2, as assumed in the past. Are the outputs from the separate auditory feature detectors always discrete and categorical? At this point it is simply not known. It may be that more "continuous" perception, such as that found for steady-state vowels (Pisoni 1973), is distinguished from categorical perception solely in the interaction of the roles of echoic memory and of auditory feature analysis.

EXPERIMENTAL EVIDENCE FOR NEUROLOGICAL LOCUS

Mapping the stages of an information-processing model onto neurological structure is not an easy task. While the clinician in particular needs to know about such facts as can be compiled, it must be remembered that the fractionation of the speech/language system with regard to clinical populations can be a hindrance rather than an aid (Jenkins 1975).

In general, the great majority of linguistic processes appear to be associated with the left cerebral hemisphere of the human brain (Geschwind 1970). However, whereas lexical, semantic, and syntactic operations may be best performed by this hemisphere, the right hemisphere also appears to play an important role in the perception of phonological cues such as stress (Blumstein and Goodglass 1972) and intonation (Blumstein and Cooper 1974). Thus the dynamic role of both hemispheres in the language process should not be ruled out, especially in view of the therapeutic value of exercising right hemisphere functions on the recovery of language abilities after stroke (Albert, Sparks, and Helm 1973; Keith and Aronson 1975).

Phonetic-level processing may occur, in part, in the right hemisphere as well as in the left hemisphere. Considering the stages of the recognition device shown in figure 3, there is no reason to assume that preliminary auditory storage is not perfectly bilateral, with equipotential left and right ear components. Most aspects of acoustic feature analysis (Stage 1) may also be bilateral, but certain aspects, such as the analysis of rapid frequency changes (Halperin, Nachshon, and Carmon 1973; Cutting 1974) and the processing of rapid amplitude modulations in the acoustic signal (Blechner, in press). The remaining stages may reside entirely in the left hemisphere, but it is only the stage of phonetic feature combination (Stage 4) for which this seems a logical necessity. Since combination can be seen as a "blending" of phonetic features, be they from two dichotic inputs or one binaural input, a single mixing device is needed. While this mixing could be duplicated in both hemispheres, it seems unlikely given the nature of the auditory pathways to each hemisphere (Milner, Taylor, and Sperry 1968). If this device were in the right hemisphere, it would be removed from many of the other aspects of language. Economy of design, then, would warrant placing it within the left hemisphere. Data supporting this allocation stem from electrophysiological (Wood, Goff, and Day 1971; Wood, 1975) and dichotic listening analyses (Studdert-Kennedy, Shankweiler, and Pisoni 1972).

Short-term memory is certainly bilateral, but different forms of it may be hemispherically specialized. Verbal forms of STM information, for example, appear to occur in the left hemisphere and spatial imagery forms appear to occur in the right hemisphere (Seamon and Gazzaniga 1973). This is a reflection of what the separate hemispheres appear to do best (Kimura 1967).

Some Controversies in Speech Perception

The basic issues in speech perception today are nearly the same as they were twenty-five years ago—the apparent lack of invariance in the acoustic signal, the related problem of segmentation, and the question the appropriate units for analysis. As for the third issue, it can be pointed out that in the approach presented here all units, including auditory features, phonetic features, phonemes, phonological features, morphemes, words, clauses, sentences, and paragraphs are deemed appropriate. All are important in speech perception and all are used, and the great body of literature in experimental psycholinguistics bears out this view. Awareness, or reversing Polanyi's metaphor "linguistic opacity," of some of the speech units, however, appears to develop in the child only after language acquisition is well advanced (Liberman et al 1974; Mattingly 1972) and can delay onset of reading readiness. This topic is discussed further in the next section.

Lack of invariance and the problem of segmentation are less easily dealt with in this model. The acoustic representation of speech is a tour de force in parallel

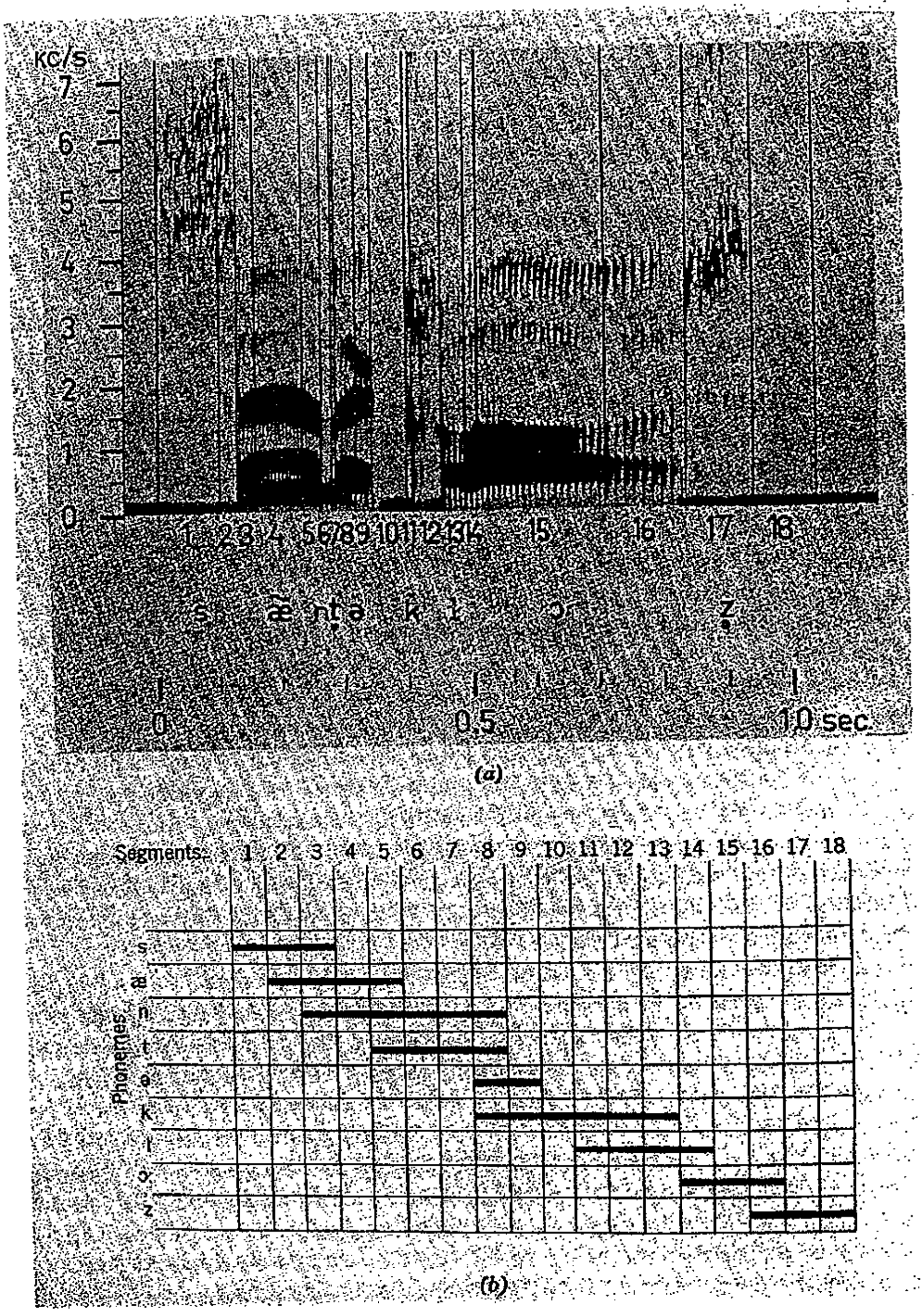

Figure 5. Perceptually, speech sounds seem to follow one another like a train of independent speech segments. Acoustically, however, there is considerable overlap. (a) A spectrogram of the words SANTA CLAUS, where vertical lines mark acoustically different segments, and (b) the assignment of phonemes to those segments. (Adapted with permission from Fant and Lindblom 1961.)

transmission of information, as shown in figure 5. Here the words SANTA CLAUS are spectrographically displayed and parsed roughly according to acoustic segments (a). These acoustic segments are then mapped onto phonemic segments (b), and the amount of parallel transmission can be seen. Segment 8, for example appears to carry information about four phonemes, /ntək/, in only 50 msec; the phone /n/ can be seen to be smeared across acoustic segments 3 through 8, a duration of roughly 200 msec. Thus the acoustic shapes of each phoneme are folded into one another to a very great degree, confounding the problem of invariance in the acoustic display.

Not only is the acoustic shape of one phoneme contingent on its immediate neighbors, but also on phonemes that may be three more phonemes away. For example, the /s/ in STREAK is different from the /s/ in STROKE because in anticipation of the different vowels /i/ and /o/ the lips are unrounded in the first example, creating a more high frequency noise, and rounded in the second, creating noise of slightly lower frequency. Such coarticulation can be so extreme that, as mentioned earlier, identical acoustic events can cue different phonemes in certain contexts and quite different events can cue the same phoneme in others (Liberman et al. 1967).

Some have viewed this position as overstated; and they have a point, especially in view of the time window perceptual moment hypothesis discussed previously. Burst cues of stop consonants, for example, are often only 5 to 10 msec in duration, well below the threshold of auditory resolution. These brief segments surely integrate into following acoustic segments to form a single auditory feature or set of features. Cole and Scott (1974a, 1974b) have claimed that such features are invariant cues that aid the process of speech perception.[3] We remain unconvinced, however, that such cues can account for the perception of all conditioned variation in speech (Liberman et al. 1967).

Without great amounts of invariance speech perception might seem to be an implausible endeavor. Although some aspects of the signal seem invariant—certain aspects of some stressed vowels, of fricatives, and of nasals—many simply are not. How then is speech perceived? Two allied views that suggest a way to cut through (if not unravel) this Gordian knot are the motor theory (Liberman et al. 1967; Studdert-Kennedy et al. 1970; Cooper 1972) and analysis by synthesis (Stevens 1960, 1972; Stevens and House 1972). Both are dynamic views of speech perception and both can be placed easily within the information-processing scheme of figures 2 and 3. In the motor theory the invariance problem is thought to be resolved at the neuromotor level; in analysis by synthesis it is resolved at the neuroacoustic level. Both accounts, when applied to our model, would add the assumption that many-to-one and one-to-many mappings of auditory features onto phonetic features are done with tacit knowledge of articulation (the motor theory) or of its acoustic consequences (analysis by synthesis). Kuhn (1975) has

---

[3] This is a position that Malcolm (1971), for one, would like. If true, it makes possible a simple, *direct* approach to speech perception without circuitous interconnections of stages and without the need for overt contributions of the human perceiver. In our opinion, however, direct perception of speech is not possible despite the attractiveness of this view.

suggested a way in which the latter might be done. A portion of the supralaryngeal vocal tract, the front cavity, may provide invariant information of a kind different from that suggested by Cole and Scott (1974a). The front cavity may allow straightforward computation of place of articulation through selective perception of the second formant or a weighted combination of the second and third formants. The only knowledge necessary for perception of place of articulation, then, is the general size of the vocal tract. The importance of the front cavity resonance to perception has yet to be explored fully, but it may bring us closer to understanding the transformation from auditory feature to phonetic feature and it is knowledge of this transformation that is crucial to understanding speech perception.

Segmentation, according to motor theory and analysis by synthesis views, is also accomplished with regard to articulation. One parses the incoming speech stream according to the way speech is produced. Thus all the stages within the recognition device would appear to have access to knowledge about speech gestures or their resultant effects on acoustic shape.

Implications for the School and Clinic

AUDITORY PROCESSES

In considering the possible implications of this model for school and clinic situations, the discussion will be limited to a consideration of phonetic perception; the macromodel is considered by other contributors to this volume. Whereas the micromodel that was presented is based on the normal speaker/listener, it should apply to certain special populations if the dynamic relationships within it are modified in appropriate ways. Before beginning, however, a few caveats are in order. First, it would be very surprising to find a single person whose language deficit could be attributed solely to the malfunction of a single stage in this model. The model itself would predict that under a long-lasting deficit anomalies would develop that run through the entire system. Second, this model must be discarded in its entirety when applied to the person who has been profoundly deaf since birth. If this individual substitutes sign language for speech, the phonetic aspect of the model may have no relevance. Recent developments in the linguistic analysis of American Sign Language (ASL) have shown it to be quite different from speech (Bellugi and Fischer 1972; Bellugi, Klima, and Siple 1974/1975; Bellugi and Klima 1975; Frishberg 1975; Friedman 1975; Lane, Boyes-Braem, and Bellugi 1976).

The first important stage to be considered is the initial registration of the signal or preliminary auditory analysis. Relevant here are considerations of certain aspects of hearing and deafness. As Danaher, Osberger, and Pickett (1973, p. 439) have noted, "Most persons with sensorineural hearing loss have some impairment of speech discrimination ability," and generally the discrimination of consonants is poorer than that of vowels. This difference could stem from several causes.

There are many kinds of sensorineural deafness. Persons can be classified as having "flat" fairly constant losses across the frequency spectrum or "sloping" losses that increase substantially for each octave above 500 Hertz (Hz) (Danaher, Osberger, and Pickett 1973). Moreover, these pure tone audiograms are not related in any simple way to speech-sound discrimination (Danaher and Pickett 1975). Thus even at this first stage of analysis considerable complexity is found. Part of the resolution must be that the deficit is not restricted to a single stage of processing.

Danaher and Pickett (1975) note three types of masking that appear to reduce the discriminability of second formant transitions. The first is a simple upward spread of masking of peripheral origin. That is, low frequency components of the speech signal mask higher frequency components. These individuals improve their discriminations when the signal level of the first formant is reduced by 10 db or when the lower frequency components of the signal are presented to one ear and the upper frequency components to the other. However, presenting the first formant to one ear and the upper formants to the other, as Broadbent (1955) and others (Rand 1974; Cutting 1976) have done in nonclinical settings, does not alleviate upward spread of masking in all individuals. Dichotic split-formant presentation is of little help in this second group. A third type of masking, as Danaher and Pickett (1975) note, complicates the issue even further. It is a type of simple backward masking in which information in a steady-state portion of a vowel reduces the detectability of initial formant transitions.

Relating these three types of data to the model cannot be straightforward, especially since hard-of-hearing individuals differ in their clinical histories. Nevertheless it appears that whereas decrements due to preliminary auditory analysis probably occur in all three types of masking, differential decrements at other stages may also be involved. The difference between monaural and dichotic upward spread of masking may be attributable to differences in ability to perform auditory feature analysis. In the monaural case the decrement in performance may be attributable to preliminary auditory analysis; in the latter case the decrement may be attributable to that stage as well as to auditory feature analysis. If it is possible that the two groups of people differ in the duration of their hearing loss, the latter group's auditory feature analyzers may have become inoperative through longer-term disuse. (This is only speculation, but the dichotic situation is an unusual one, not experienced outside the laboratory. It may be that these auditory feature analyzers have not been stimulated for a long period of time and have "atrophied" [Eimas 1975], that is, become inoperative through lack of use.)

Speech discrimination loss attributable to backward masking would appear to be of a different kind. Backward masking is a temporal phenomenon, whereas upward spread of masking is one of frequency. Temporal phenomena implicate memories and integrating time windows; part of this loss of information may be attributable, therefore, to limitations on coding information in sensory information store, as well as auditory feature analysis.

This assumption appears to be supported by results of research on quite a different population. Tallal and Piercy (1973, 1974, 1975) have studied children diagnosed as developmental aphasics (see also Tallal, this volume). These children have normal audiograms but have extreme difficulty perceiving temporally patterned auditory signals. In a series of careful studies these researchers found that the developmental aphasic has much more difficulty discriminating consonants than vowels and that cue duration appears to be the primary cause. This result is interesting since normal listeners often report no difficulty in identifying target vowels of even shorter duration in a similar situation (Dorman, et al. 1976).

The results from studies of these children appear to implicate a deficit in sensory information store and perhaps auditory property analysis. As Tallal and Piercy point out, these children appear to have an auditory system whose early stages are more "sluggish" than the norm. Perhaps their integrating time windows are larger. Without remediation of this difficulty, deficits throughout the system, especially at the higher levels, are likely to persist and even spread, as Tallal (this volume) suggests.

The deficits discussed thus far can be considered deficits primarily in auditory analysis rather than deficits of speech perception per se. These deficits clearly have effects higher in the system, but those probably result from anomalies at this lower level rather than the converse.

PHONETIC PROCESSES

Evidence for phonetic impairment without auditory impairment is much more difficult to come by. In fact, given the importance recently attributed to the stage of auditory feature analysis, there may be no such evidence at the present time. Moreover, we are not particularly optimistic about possible applications to the school and clinic. Results from normal listeners that used to be taken as indicative of phonetic processing (see Wood 1975, p. 16) now appear to be indicative of auditory feature analysis (see Cutting, Rosner, and Foard 1976). Evidence found for Stage 2 may be contaminated by the operation of prior stages or even subsequent ones. Instead let us consider some of the problems of such contamination and the more general problems of applying experimental paradigms to field situations in the school and clinic. The research using a dichotic listening paradigm is a case in point.

The standard dichotic listening procedure used in the laboratory during the last fifteen years may not yield decisive information for professionals in the school and clinic. The reasons for this assessment stems from several sources. First, it is difficult, and sometimes unwise, to lump together persons with similar perceptual problems. Given the great complexity of the human organism, these individuals are likely to differ greatly even if they fit into the same diagnostic category, as suggested by the results of Danaher and Pickett (1975). Thus individual-oriented procedure would seem best. But this is precisely where the utility of the dichotic listening procedure and most other laboratory techniques is strained. In dichotic listening a right ear advantage may be attributable to a number of causes including those associated with almost every stage shown in figures 2 and 3. The existence of a right ear advantage would tell the applied researcher that some part of the speech/language system is running properly. Even by using nonsense syllables as stimuli, the applied researcher may gain no more information than that (Dorman and Geffner 1974; Dorman and Porter 1975). The lack of a right ear advantage *may* be indicative of malfunction in one stage, but even in the best of situations i does not necessarily tell the researcher which stage, and in the worst of situations it does not imply that any stage is malfunctioning. The existence of a right ear advantage even among the normal population is only a probabilistic occurrence, and the lack of an ear advantage should not be seen as an abnormality (Shankweiler and Studdert-Kennedy 1975). (We feel strongly about this, if only for the reason that one of us consistently yields no ear advantage in these tasks.) Moreover, left ear advantages for linguistic material are not uncommon and must be interpreted with caution even in the most extreme instances (see Fromkin et al. 1974).

Results from other paradigms may be more interpretable, but they may not indicate much about phonetic processing per se. The application of identification and discrimination paradigms associated with categorical perception, selective adaptation, and tasks requiring selective attention to different dimensions of the speech stimulus may be valuable in terms of knowledge about general auditory processing and may serve as verification (or falsification) of certain aspects of the model proposed here (although again, there is no cause for optimism, in our opinion). In spite of this skepticism, there are some recent results in speech perception combined with reading that may bode an important message for the school and clinic. These findings relate to one of the overriding issues in speech perception—units of analysis.

UNITS OF SPEECH ANALYSIS AND READING

Although all units of speech analysis are relevant to speech perception, not all of them are equally apparent, especially to the child. Conscious access to, or knowledge of, speech parsing routines is important in the acquisition of reading skill. In logographic writing, where words are the primary unit, as in Chinese and in the Japanese Kanji script, knowledge about words is imperative. In syllabaric writing, where syllables are the primary unit, such as in the Japanese use of Kana script (see Sasanuma 1975) and in the Cherokee script (Walker 1974) knowledge of syllables is imperative. In alphabetic writing, as in English, where the phoneme is the principal unit, knowledge of phonemes is imperative.

To the young child, syllables and phonemes are not equally amenable units of analysis. I. Liberman et al. (1974) examined the ability of nursery schoolers, kindergarteners, and first graders to tap out the number of syllables and phonemes in common words. Measured for each child was the number of trials needed to reach a criterion of six consecutive correctly tapped trials without demonstration by the experimenter. Ability to segment by syllable was shown by half of the four-year-olds but none of them could segment by phoneme. By age six, 90 percent of the children could segment by syllable but only 70 percent by phoneme. From these results, Liberman et al. suggested that one of the reasons that children under the age of five or six are not ready to learn to read is because they are not yet consciously aware of the units on which written English is based.

Another important aspect of reading readiness, as Conrad (1972) notes, is the ability to code verbally information into short-term memory. I. Liberman, et al.

(1976) applied this notion to good and poor beginning readers. They suspected that good beginning readers would be those who actively use their newly acquired skill of parsing language into phonemes and develop the appropriate coding in short-term memory based on phonemic structure. Poor readers, on the other hand, might still be baffled by phonemes and therefore be unable to use the parsimonious phoneme code. To these two groups of beginning readers, who had been equated for intelligence, they presented two types of consonant strings to view. One type of string consisted of consonants with rhyming names (B, C, D, G, P, T, V, Z) and the other had nonrhyming names (H, K, L, Q, R, S, W, Y). Subjects were then tested in conditions of immediate recall and delayed recall. In general, the good readers made fewer errors than the poor readers on both rhyming and nonrhyming consonants, although the rhyming consonants were more difficult for all. More importantly, the advantage shown by the good readers disappeared for the rhyming consonants in the delayed recall condition but did not disappear for the nonrhyming consonants. The implication is that the decrement in the delayed condition shown by the good readers, but not by the poor readers, is attributable to the use of phonetic codes in short-term memory by good readers and the possible use of some other code, perhaps a visual one (Conrad 1972), by the poor readers. (See Crowder, this volume, for a further discussion of this research.)

These results imply that for the beginning reader of English reading is in part a derivative of speech perception. Acquisition of reading skill depends on mastery of certain linguistic skills and on the youngster's awareness of the proper units of analysis. The availability of a phonetically organized short-term memory may not be enough to learn to read an alphabetically written language. In addition to memory the children must have phoneme parsing skills. Reading instruction and remediation programs might therefore take as an early goal the teaching of awareness of speech segments.

Summary

The goal of this chapter has been to present one point of view about the perceptua analysis of speech sounds. It is a process that takes time; consists of many stages uses a number of memory stores; is limited in certain ways; and uses all levels o the speech/language system, including phonetic, phonological, lexical, syntactic and semantic components. Evidence has been presented to support the speec perception model from research with adult and infant humans, as well as wit animals. This evidence, of course, does not confirm this model and disprove alte native views; but the evidence, plus important logical considerations, make th view a plausible one. Speech perception and speech production appear to be i extricably intertwined, both on grounds of parsimony, as Lashley (1951) suggeste and on grounds that there is too complex a mapping from auditory features t phonetic features and from phonetic features to phonological features for an alternative method of speech perception to be feasible. Finally, the implicatio of some of the general views and findings of speech perception research have bee applied to problems in the school and clinic. Certain experiments with aphas and hard-of-hearing listeners may help support the proposed view of speech perception. Other findings of speech research, especially those connected with reading, may help the applied researcher in the school and clinic.

References

Albert, M. L., R. W. Sparks, and N. A. Helm. 1973. Melodic intonation therapy for aphasia. *Archives of Neurology, 29,* 130–131.

Allport, D. A. 1968. Phenomenal simultaneity and the perceptual moment hypothesis. *British Journal of Experimental Psychology, 59,* 395–406.

Barclay, J. R. 1972. Noncategorical perception of a voiced stop: A replication. *Perception and Psychophysics, 11,* 269–273.

Bellugi, U., and S. Fischer. 1972. A comparison of sign language and spoken language. *Cognition, 1,* 173–200.

Bellugi, U., and E. S. Klima. 1975. Aspects of sign language structure. In *The role of speech in language,* J. F. Kavanagh and J. E. Cutting (eds.) Cambridge, Mass.: MIT Press. 171–203.

Bellugi, U., E. S. Klima, and P. Siple. 1974/1975. Remembering the signs. *Cognition, 3,* 93–125.

Binnie, C. A., A. A. Montgomery, and P. L. Jackson. 1974. Auditory and visual contributions to the perception of selected English consonants. *Journal of Speech and Hearing Research, 17,* 619–630.

Bjork, R. A. 1975. Short-term storage: The ordered output of a central processor. In *Cognitive theory,* Vol. 1, F. Restle, R. M. Shiffrin, N. J. Castellan, H. Lindman, and D. B. Pisoni (eds.). Hillsdale, N.J.: Erlbaum Associates. 151–172.

Blakemore, C., and F. W. Campbell. 1969. On the existence of neurons in the human visual system selectively sensitive to the orientation and size of retinal images. *Journal of Physiology, 203,* 237–260.

Blechner, M. In press. Right-ear advantage for musical stimuli differing in rise time. *Perception and Psychophysics.*

Blumstein, S., and W. E. Cooper. 1974. Hemispheric processing of intonation contours. *Cortex, 10,* 146–158.

Blumstein, S., and H. Goodglass. 1972. The perception of stress as a semantic cue in aphasia. *Journal of Hearing and Speech Research, 15,* 800–806.

Bransford, J., and K. E. Nitsch. This volume. Coming to understand things we could not previously understand.

Bresnan, J. 1976. Toward a realistic model of transformational grammar. Paper presented at the conference on New Approaches to a Realistic Model of Language, MIT, Cambridge, Mass., March 9.

Broadbent, D. E. 1955. A note on binaural fusion. *Quarterly Journal of Experimental Psychology, 7,* 46–47.

———. 1958. *Perception and communication.* London: Pergamon Press.

———. 1965. Information processing in the nervous system. *Science, 150,* 457–462.

———. 1971. *Decision and stress.* New York: Academic Press.

———. 1975. The magic number seven after fifteen years. In *Studies in long term memory* A. Kennedy and A. Wilkes (eds.). London: Wiley, 3–18.

Brown, R. 1973. *A first language.* Cambridge, Mass.: Harvard University Press.

Brown, R., and D. McNeill. 1966. The "tip of the tongue" phenomenon. *Journal of Verba Learning and Verbal Behavior, 5,* 325–337.

Chomsky, N. 1964. Current issues in linguistic theory. In *The structure of language,* J.A Fodor and J. J. Katz (eds.). Englewood Cliffs, N.J.: Prentice-Hall. 50–118.

Cole, R. A. 1973. Listening for mispronunciations: A measure of what we hear durin speech. *Perception and Psychophysics, 13,* 153–156.

Cole, R. A., and B. Scott. 1974a. The phantom in the phoneme: Invariant cues for sto consonants. *Perception and Psychophysics, 15,* 101–107.

———. 1974b. Toward a theory of speech perception. *Psychological Review, 81,* 348–374.

Collins, A. M., and M. R. Quillian. 1969. Retrieval time from semantic memory. *Journa of Verbal Learning and Verbal Behavior, 8,* 240–248.

Conrad, R. 1972. Speech and reading. In *Language by ear and by eye,* J. F. Kavanagh an I. G. Mattingly (eds). Cambridge, Mass.: The MIT Press. 205–240.

Cooper, F. S. 1972. How is language conveyed by speech? In *Language by ear and by ey* J. F. Kavanagh and I. G. Mattingly (eds.). Cambridge, Mass.: MIT Press. 25–46.

Cooper, W. E. 1975. Selective adaptation to speech. In *Cognitive theory,* Vol. 1, F. Restl R. M. Shiffrin, N. J. Castellan, H. Lindman, and D. B. Pisoni (eds.). Hillsdale, N. J. Erlbaum Associates. 23–54.

Cooper, W. E., and S. E. Blumstein. 1974. A "labial" feature anlayzer in speech percep tion. *Perception and Psychophysics, 15,* 591–600.

de Cordemoy, G. 1668. *A Philosophical discourse concerning speech.* London: J. Martin.

Crowder, R. G. 1971. The sound of vowels and consonants in immediate memory. *Journc of Verbal Learning and Verbal Behavior, 10,* 587–596.

———. 1972. Visual and auditory memory. In *Language by ear and by eye,* J. F. Kavanag and I. G. Mattingly (eds.). Cambridge, Mass.: MIT Press. 251–276.

———. This volume. Language and memory.

Crowder, R. G., and J. Morton. 1969. Precategorical acoustic storage (PAS). *Percepti and Psychophysics, 5,* 365–373.

Cutting, J. E. 1974. Two left-hemisphere mechanisms in speech perception. *Perception an Psychophysics, 16,* 601–612.

———. 1976. Auditory and linguistic processes in speech perception: Inferences from si fusions in dichotic listening. *Psychological Review, 83,* 114–140.

———. 1977. The magical number two and the natural categories of speech and music. I *Tutorial essays in psychology,* N. S. Sutherland (ed.). Hillsdale, N.J.: Erlbaum Associates. 1 33.

Cutting, J. E., and P. D. Eimas. 1975. Phonetic feature analyzers and the processing of speech in infants. In *The role of speech in language*, J. F. Kavanagh and J. E. Cutting (eds.). Cambridge, Mass.: The MIT Press, 127–148.

Cutting, J. E., and B. S. Rosner. 1974. Categories and boundaries in speech and music. *Perception and Psychophysics, 16,* 564–570.

Cutting, J. E., B. S. Rosner, and C. F. Foard. 1976. Perceptual categories for musiclike sounds: Implications for theories of speech perception. *Quarterly Journal of Experimental Psychology, 28,* 361–378.

Danaher, E. M., and J. M. Pickett. 1975. Some masking effects produced by low-frequency vowel formants in persons with sensorineural loss. *Journal of Speech and Hearing Research, 18,* 261–271.

Danaher, E. M., M. J. Osberger, and J. M. Pickett. 1973. Discrimination of formant frequency transitions in synthetic vowels. *Journal of Speech and Hearing Research, 16,* 439–451.

Darwin, C. J., and A. D. Baddeley. 1974. Acoustic memory and the perception of speech. *Cognitive Psychology, 6,* 41–60.

Day, R. S., and C. C. Wood. 1972. Interactions between linguistic and nonlinguistic processing. *Journal of the Acoustical Society of America, 51,* 79(A).

Dorman, M. F., and Geffner, D. S. 1974. Hemispheric specialization for speech perception in six-year-old black and white children from low and middle socio-economic classes. *Cortex, 10,* 171–176.

Dorman, M. F., and R. J. Porter. 1975. Hemispheric lateralization for speech perception in stutterers. *Cortex, 11,* 181–185.

Dorman, M. F., D. Kewley-Port, S. Brady, and M. T. Turvey. In press. Vowel recogniton: Inferences from studies of forward and backward masking. *Quarterly Journal of Experimental Psychology.*

Eimas, P. D. 1974. Auditory and linguistic processing of cues for place of articulation by infants. *Perception and Psychophysics, 16,* 513–521.

———. 1975. Auditory and phonetic coding of the cues for speech: Discrimination of the [r–l] distinction by young infants. *Perception and Psychophysics, 18,* 341–347.

Eimas, P. D., and J. D. Corbit. 1973. Selective adaptation of linguistic feature detectors. *Cognitive Psychology, 4,* 99–109.

Eimas, P. D., W. E. Cooper, and J. D. Corbit. 1973. Some properties of linguistic feature detectors. *Perception and Psychophysics, 13,* 206–217.

Eimas, P. D., E. R. Siqueland, P. W. Jusczyk, and J. M. Vigorito. 1971. Speech perception in infants. *Science, 171,* 303–306.

Erber, N. P. 1969. Interaction of audition and vision in the recognition of oral speech stimuli. *Journal of Speech and Hearing Research, 12,* 423–425.

———. 1975. Auditory-visual perception of speech. *Journal of Speech and Hearing Disorders, 40,* 481–491.

Fant, G., and B. Lindblom. 1961. Studies of minimal speech and sound units. Speech Transmission Laboratory, Royal Institute of Technology, Stockholm, Sweden. *Quarterly Progress Report, 2,* 1–11.

Fodor, J. A., T. G. Bever, and M. F. Garrett. 1974. *The psychology of language.* New York: McGraw-Hill.

Foder, J. A., M. F. Garrett, and S. L. Brill. 1975. Pi ka pu: The perception of speech sounds by prelinguistic infants. *Perception and Psychophysics, 18,* 74–78.

Foulke, E., and T. G. Sticht. 1969. Review of research on the intelligibility and comprehension of accelerated speech. *Psychological Bulletin, 72,* 50–62.

Friedman, L. A. 1975. Space, time, and person reference in American sign language. *Language, 51,* 940–961.

Frishberg, N. 1975. Arbitrariness and iconicity: Historical change in American sign *language, 51,* 696–719.

Fromkin, V. A. 1971. The non-anomalous nature of anomalous utterances. *Language, 47,* 27–52.

Fromkin, V. A., S. Krashen, S. Curtiss, D. Rigler, and M. Rigler. 1974. The development of language in Genie: A case of language acquisition beyond the "critical period." *Brain and Language, 1,* 81–107.

Gerstman, L. J. 1957. Perceptual dimensions for the friction portion of certain speech sounds. Ph.D. dissertation, New York University.

Geschwind, N. 1970. The organization of language and the brain. *Science, 170,* 940–944.

Haber, R. N. 1969. Introduction. In *Information-processing approaches to visual perception,* R. N. Haber (ed.). New York: Holt, Rinehart, Winston. 1–15.

Halperin, Y., I. Nachshon, and A. Carmon. 1973. Shift in ear superiority in dichotic listening to temporal pattern nonverbal stimuli. *Journal of the Acoustical Society of America,* 53, 46–50.

Halwes, T. G. 1969. Effects of dichotic fusion on the perception of speech. Ph.D. dissertation, University of Minnesota. *Dissertation Abstracts International,* 1970, 31, 1565b. (Univ. Microfilms No. 70–15, 736.)

Hirsh, I. J. 1974. Temporal order and auditory perception. In *Sensation and measurement: Papers in honor of S. S. Stevens,* H. R. Moskowitz, B. Sharf, and J. C. Stevens (eds.). Boston: D. Riedel. 251–258.

Howes, D. 1957. On the relation between the intelligibility and frequency of occurrence of English words. *Journal of the Acoustical Society of America 29,* 296–305.

Hughes, G. W., and M. Halle. 1956. Spectral properties of fricative consonants. *Journal of the Acoustical Society of America, 28,* 303–310.

Jakobson, R., G. Fant, and M. Halle. 1951. *Preliminaries to speech analysis.* Cambridge, Mass.: MIT Press (1963 ed.).

Jenkins, J. J., E. Jiménez-Pabón, R. E. Shaw, and J. W. Sefer. 1975. *Schuell's aphasia in adults.* New York: Harper and Row.

Jusczyk, P. W., B. S. Rosner, J. E. Cutting, C. F. Foard, and L. Smith. 1977. Categorical perception of nonlinguistic sounds in the two month old infant. *Perceptions and Psychophysics, 21,* 50–54.

Kaplan, E. L., and G. Kaplan. 1971. The prelinguistic child. In *Human development and cognitive processes,* J. Eliot (ed.). New York: Holt, Rinehart, and Winston. 358–381.

Keith, R. L., and A. E. Aronson. 1975. Singing as therapy for apraxia of speech and aphasia: Report of a case. *Brain and Language, 2,* 483–488.

Kimura, D. 1967. Dual functional asymmetry of brain in visual perception. *Neuropsychologia, 4,* 275–285.

Kozlowski, L. T. in press. Effects of distorted auditory and of rhyming cues on the retrieval of tip-of-the-tongue words by poets and nonpoets. *Memory and Cognition* 69–72.

Kuhn, G. M. 1975. On the front cavity resonance and its possible role in speech perception. *Journal of the Acoustical Society of America, 58,* 428–433.

Lane, H., P. Boyes-Braem, and U. Bellugi. 1976. Preliminaries to a distinctive feature analysis of handshapes in American Sign Language. *Cognitive Psychology, 8,* 263–289.

Lashley, K. 1951. The problem of serial order in behavior. In *Cerebral mechanisms in behavior,* L. Jeffress (ed.). New York: Wiley, 112–136.

Liberman, A. M. 1970. The grammars of speech and language. *Cognitive Psychology, 1,* 301–323.

Liberman, A. M., P. C. Delattre, and F. S. Cooper. 1952. The role of selected stimulus variables in the perception of the unvoiced stop consonants. *American Journal of Psychology, 65,* 497–516.

———. 1958. Some cues for the distinction between voiced and voiceless stops. *Language and Speech, 1,* 153–167.

Liberman, A. M., I. G. Mattingly, and M. T. Turvey. 1972. Language codes and memory codes. In *Coding processes in human memory,* A. W. Melton and E. Martin (eds.). New York: Winston. 307–334.

Liberman, A. M., F. S. Cooper, D. P. Shankweiler, and M. Studdert-Kennedy. 1967. Perception of the speech code. *Psychological Revew, 74,* 431–461.

Liberman, A. M., K. S. Harris, H. S. Hoffman, and B. C. Griffith. 1957. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54,* 358–368.

Liberman, A. M., K. S. Harris, P. D. Eimas, L. Lisker, and J. Bastian. 1961. An effect of learning speech perception: The discrimination of durations of silence with and without phonemic significance. *Language and Speech, 4,* 175–195.

Liberman, I. Y., D. P. Shankweiler, F. W. Fischer, and B. Carter. 1974. Reading and the awareness of linguistic segments. *Journal of Experimental Child Psychology, 18,* 201–212.

Liberman, I. Y., D. P. Shankweiler, A. M. Liberman, C. Fowler, and F. W. Fischer. Forthcoming. Phonetic segmentation and recording in the beginning reader. In *Towards a Psychology of Reading*: Proceedings of the CUNY Conference. A. S. Reber and D. Scarborough (eds.). Hillsdale, N.J.: Erlbaum Associates.

Licklider, J. C. R. 1952. On the process of speech perception. *Journal of the Acoustical Society of America, 24,* 590–594.

Lisker, L. 1975. Is it VOT or a first-formant transition detector? *Jorunal of the Acoustical Society of America, 57,* 1547–1551.

Lisker, L., and A. S. Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20,* 384–422.

———. 1967. Some effects of context on voice onset time in English stops. *Language and Speech, 10,* 1–28.

MacKay, D. G. 1970. Spoonerisms: The anatomy of errors in the serial order of speech. *Neuropsychologia, 8,* 323–350.

Malcolm, N. 1971. The myth of cognitive processes and structures. In *Cognitive development and epistemology,* T. Mischel (ed.). New York: Academic Press. 385–392.

Marslen-Wilson, W. D. 1975. Sentence perception as an interactive parallel process. *Science, 189,* 226–228.

Massaro, D. W. 1975. Language and information processing. In *Understanding language,* D. W. Massaro (ed.). New York: Academic Press. 3–28.

Massaro, D. W., and M. M. Cohen. 1975. Preperceptual auditory storage in speech perception. In *Structure and process in speech perception,* A. Cohen and S. G. Nooteboom (eds.). Heidelberg: Springer Verlag. 226–243.

Mattingly, I. G. 1972. Reading, the linguistic process, and linguistic awareness. In *Language by ear and by eye,* J. F. Kavanagh and I. G. Mattingly (eds.). Cambridge, Mass.: MIT Press. 133–148.

Mattingly, I. G., A. M. Liberman, A. Syrdal, and T. Halwes. 1971. Discrimination in speech and nonspeech modes. *Cognitive Psychology, 2,* 131–157.

Miller, G. A. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81–96.

———. 1962. Decision units in the perception of speech. *IRE Transactions on Information Theory, IT-8,* 81–83.

———. 1972. English verbs of motion: A case study in semantics and lexical memory. In *Coding processes in human memory,* A. W. Melton and E. Martin (eds.). New York: Winston. 335–372.

———. This volume. Lexical meaning.

Miller, G. A., and S. Isard. 1963. Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior, 2,* 217–228.

Miller, G. A., and P. Nicely. 1955. An analysis of some perceptual confusions among some English consonants. *Journal of the Acoustical Society of America, 27,* 338–352.

Miller, G. A., G. A. Heise, and W. Lichten. 1951. The intelligibility of speech as a function of the context of the test materials. *Journal of the Acoustical Society of America, 41,* 329–335.

Miller, J. D., C. C. Wier, R. E. Pastore, W. J. Kelly, and R. J. Dooling. 1976. Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception. *Journal of the Acoustical Society of America, 60,* 411–417.

Miller, J. L. 1975. Properties of feature detectors for speech: Evidence from the effects of selective adaptation in dichotic listening. *Perception and Psychophysics, 18,* 389–397.

Miller, J. L., and P. D. Eimas. 1976. Studies on the selective tuning of feature detectors for speech. *Journal of Phonetics, 4,* 119–127.

Milner, B., L. Taylor, and R. W. Sperry. 1968. Lateralized suppression of dichotically presented digits after commissural section in man. *Science, 161,* 184–185.

Miyawaki, K., W. Strange, R. Verbrugge, A. M. Liberman, J. J. Jenkins, and O. Fujimura. 1975. An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics, 18,* 331–340.

Moffitt, A. R. 1971. Consonant cue perception by twenty- to twenty-four-week-old infants. *Child Development, 42,* 717–731.

Morse, P. A. 1972. The discrimination of speech and nonspeech stimuli in early infancy. *Journal of Experimental Child Psychology, 14,* 477–492.

Morse, P. A., and C. T. Snowden. 1975. An investigation of categorical speech discrimination by rhesus monkeys. *Perception and Psychophysics, 17,* 9–16.

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

Norman, D. A. 1972. The role of memory in understanding language. In *Language by ear and by eye*, J. F. Kavanagh and I. G. Mattingly (eds.). Cambridge, Mass.: MIT Press. 277–288.

O'Neill, J. J. 1954. Contribution to the visual components of oral symbols to speech comprehension. *Journal of Speech and Hearing Disorders, 19*, 429–439.

Orr, D. B., H. L. Friedman, and J. C. Williams. 1965. Trainability of listening comprehension of speeded discourse. *Journal of Educational Psychology, 56*, 148–156.

Parducci, A. 1974. Contextual effects: A range-frequency analysis. In *Handbook of perception*, Vol. 2, E. C. Carterette and M. P. Friedman (eds.). New York: Academic Press. 127–141.

Pisoni, D. B. 1973. Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and Psychophysics, 13*, 253–260.

———. 1975a. Dichotic listening and processing phonetic features. In *Cognitive theory* Vol. 1, F. Restle, R. M. Shiffrin, N. J. Castellan, H. Lindman, and D. B. Pisoni (eds.). Hillsdale, N.J.: Erlbaum Associates. 79–102.

———. 1975b. Information processing and speech perception. In *Speech communication*, Vol. 3, G. Fant (ed.). New York: John Wiley and Sons. 331–337.

———. 1976. Identification and discrimination of relative onset time of two-component tones: Implications for voicing perception in stops. *Quarterly Progress Report*, Research Laboratory of Electronics, MIT, Cambridge, Mass., *118*, 212–229.

———. Forthcoming-a. Mechanisms of auditory discrimination and coding of linguistic information. In *Second auditory processing and learning disabilities symposium*, J. V. Irwin (ed.). Memphis, Tenn.: Memphis State University Press.

———. Forthcoming-b. Speech perception. In *Handbook of learning and cognitive processes*, Vol. 5, W. K. Estes (ed.). Hillsdale, N.J.: Erlbaum Associates.

Pisoni, D. B., and J. R. Sawusch. 1975. Some stages of processing in speech perception. In *Structure and process in speech perception*, A. Cohen and S. Nooteboom (eds.). Heidelberg: Springer-Verlag. 16–34.

Pisoni, D. B., and J. B. Tash. 1975. Auditory property detectors and processing place features in stop consonants. *Perception and Psychophysics, 18*, 401–408.

Polanyi, M. 1964. *Personal knowledge*. New York: Harper and Row.

Pollack, I., and J. M. Pickett. 1963. The intelligibility of excerpts from conversation. *Language and Speech, 6*, 165–171.

———. 1964. Intelligibility of excerpts from fluent speech: Auditory vs. structural context. *Journal of Verbal Learning and Verbal Behavior*, 3, 79–84.

Rand, T. C. 1974. Dichotic release from masking for speech. *Journal of the Acoustical Society of America, 55*, 678–680.

Raphael, L. J. 1972. Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *Journal of the Acoustical Society of America, 51*, 1296–1303.

Sasanuma, S. 1975. Kana and Kanji processing in Japanese aphasics. *Brain and Language, 2*, 369–383.

Savin, H. B. 1963. Word-frequency effect and errors in the perception of speech. *Journal of the Acoustical Society of America, 35,* 200–206.

Sawusch, J. R. 1976. The structure and flow of information in speech perception. Ph.D. dissertation, Indiana University.

Schatz, C. 1954. The role of context in the perception of stops. *Language, 30,* 47–56.

Seamon, J. G., and M. S. Gazzaniga. 1973. Coding strategies and cerebral laterality effects. *Cognitive Psychology, 5,* 249–256.

Shankweiler, D. P., and M. Studdert-Kennedy. 1975. A continuum of lateralization for speech perception. *Brain and Language, 2,* 212–225.

Shiffrin, R. M. 1975. Short-term store: The basis for a memory system. In *Cognitive theory,* Vol. 1, F. Restle, R. M. Shiffrin, N. J. Castellan, H. Lindman, and D. B. Pisoni (eds.). Hillsdale, N.J.: Erlbaum Associates. 193–218.

Shiffrin, R. M., D. B. Pisoni, and K. Castaneda-Mendez. 1974. Is attention shared between the ears? *Cognitive Psychology, 6,* 190–215.

Sinnott, J. M., M. D. Beecher, D. B. Moody, and W. C. Stebbins. 1976. Speech sound discrimination by monkeys and humans. *Journal of the Acoustical Society of America, 60,* 687–695.

Stevens, K. N. 1960. Toward a model for speech perception. *Journal of the Acoustical Society of America, 32,* 47–55.

———. 1972. The quantal nature of speech: Evidence from articulatory data. In *Human communication: A unified view,* E. E. David and D. P. Denes (eds.). New York: McGraw-Hill. 51–66.

———. 1975. The potential role of property detectors in the perception of consonants. In *Auditory analysis and perception of speech,* G. Fant and M. A. A. Tatham (eds.). New York: Academic Press. 303–330.

Stevens, K. N., and A. House. 1972. Speech perception. In *Foundations of modern auditory theory,* Vol. 2, J. V. Tobias (ed.). New York: Academic Press. 1–62.

Stevens, K. N., and D. H. Klatt. 1974. Role of formant transitions in the voiced-voiceless distinction for stops. *Journal of the Acoustical Society of America, 55,* 653–659.

Studdert-Kennedy, M., D. P. Shankweiler, and D. B. Pisoni. 1972. Auditory and phonetic processes in speech perception: Evidence from a dichotic study. *Cognitive Psychology, 2,* 455–466.

Studdert-Kennedy, M., D. P. Shankweiler, and S. Schulman. 1970. Opposed effects of a delayed channel on perception of dichotically and monotically presented CV syllables. *Journal of the Acoustical Society of America, 48,* 579–594.

Studdert-Kennedy, M., A. M. Liberman, K. S. Harris, and F. S. Cooper. 1970. Motor theory of speech perception: A reply to Lane's critical review. *Psychological Review, 77,* 234–249.

Sumby, W. H., and I. Pollack. 1954. Visual contributions to speech intelligibility in noise. *Journal of the Acoustical Society of America, 34,* 58–60.

Tallal, P. This volume. Implications of speech perceptual research for clinical populations.

Tallal, P., and M. Piercy. 1973. Developmental aphasia: Impaired rate of nonverbal processing as a function of sensory modality. *Neuropsychologia, 11,* 389–398.

———. 1974. Developmental aphasia: Rate of auditory processing and selective impairment of consonant perception. *Neuropsychologia, 12*, 83–93.

———. 1975. Developmental aphasia: The perception of brief vowels and extended stop consonants. *Neuropsychologia, 13*, 69–74.

Tartter, V. C., and P. D. Eimas. 1975. The role of auditory feature detectors in the perception of speech. *Perception and Psychophysics, 18*, 293–298.

Tulving, E. 1972. Episodic and semantic memory. In *Organization of memory*, E. Tulving and W. Donaldson (eds.). New York: Academic Press. 381–403.

Walker, W. 1974. Cherokee. In *Studies of Southeastern Indian languages*, J. Crawford (ed.). Athens, Ga.: University of Georgia Press. 189–236.

Waters, R. S., and W. A. Wilson. 1976. Speech perception by rhesus monkeys: The voicing distinction in synthesized labial and velar stop consonants. *Perception and Psychophysics, 19*, 285–289.

Whitfield, I. C. 1965. "Edges" in auditory information processing. In *Proceedings of the XXIII International Congress of Physiological Sciences*, Tokyo, 245–247.

Whorf, B. L. 1940. Linguistics as an exact science. *Technology Review, 43*, 61–63, 80–83. In *Language, thought, and reality*, J. B. Carroll (ed.). Cambridge, Mass.: MIT Press, 1956, 220–232.

Wood, C. C. 1974. Parallel processing of auditory and phonetic information in speech discrimination. *Perception and Psychophysics, 15*, 501–508.

———. 1975. Auditory and phonetic levels of processing in speech perception: Neurophysiological and information-processing analyses. *Journal of Experimental Psychology: Human Perception and Performance, 1*, 3–20.

Wood, C. C., W. R. Goff, and R. S. Day 1971. Auditory evoked potentials during speech perception. *Science, 173*, 1248–1251.

Speech and Language in the Laboratory, School, and Clinic

JAMES F. KAVANAGH
AND WINIFRED STRANGE, EDITORS

1978

The MIT Press, Cambridge, Massachusetts, and London, England