

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 19 (1993-1994)
Indiana University

**Sources of Variability Affecting Speech Perception
and Spoken Word Recognition¹**

David B. Pisoni

*Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405*

¹ This research was supported by NIDCD Research Grant DC-00111-18 to Indiana University in Bloomington, IN. A shorter version of this paper was presented at the National Academy of Sciences, CHABA Meeting, June 3, 1993, in Washington, DC. I thank Steve Goldinger, Scott Lively, Lynne Nygaard, Mitchell Sommers, Thomas Palmeri and John Karl for their help and collaboration in various phases of this research program.

Abstract

This paper reviews recent studies on the perception, encoding and retention of stimulus variability in speech perception. Experiments on talker variability, speaking rate and perceptual learning provide evidence for the encoding of very fine perceptual details of the speech signal. Listeners apparently encode specific attributes of the talker's voice and speaking rate into long-term memory. The process of perceptual normalization in speech perception therefore appears to involve the encoding of specific instances or "episodes" of the stimulus input and the processing operations used in perceptual analysis. The present set of findings is consistent with non-analytic accounts of perception, memory and cognition which emphasize the contribution of episodic or exemplar-based encoding in long-term memory. The results also raise questions about the long-standing dissociation in phonetics between the linguistic and indexical properties of speech. Listeners apparently do encode and retain non-linguistic information in long-term memory about the speaker's gender, dialect, speaking rate and emotional state, attributes of speech signals that are not traditionally considered part of phonetic or lexical properties of words. The findings reported here have important implications for current theoretical accounts of how the nervous system encodes speech signals and what kinds of information are stored in the mental lexicon.

Sources of Variability Affecting Speech Perception and Spoken Word Recognition

For the last several years we have been interested in the interface between speech perception and spoken language comprehension and, in particular, problems of lexical access and the structure and organization of sound patterns in the mental lexicon (Pisoni, Nusbaum, Luce & Slowiaczek, 1985). Findings from a variety of recent studies carried out at Indiana suggest that very fine details in the speech signal are preserved in the human memory system for relatively long periods of time (see Pisoni, 1990; 1992a,b, 1993; Goldinger, 1992). This information appears to be used in several ways to facilitate perceptual encoding, retention and retrieval of information from memory. Many of our recent investigations have been concerned with assessing the effects of different sources of variability in speech perception (Sommers, Nygaard & Pisoni, 1992, 1994; Nygaard, Sommers & Pisoni, 1992a,b). The results of these studies have encouraged us to reassess our beliefs about several long-standing theoretical issues in speech perception such as acoustic-phonetic invariance and the problems of perceptual normalization (Pisoni, 1992a,b).

In the sections below, I will summarize the results from several recent studies that deal with the encoding of stimulus variability in speech perception experiments. These findings have raised a number of important new questions about the traditional dissociation between the linguistic and indexical properties of speech signals and the role that different sources of variability play in speech perception and spoken word recognition. For many years, linguists have considered attributes of the talker's voice-- what Ladefoged refers to as the "personal" characteristics of speech-- to be independent of the linguistic content of the talker's message (Ladefoged, 1975; Laver & Trudgill, 1979). The dissociation of these two parallel sources of information in speech may have served a useful function in the formal linguistic analysis of language when viewed as an idealized abstract system of symbols. However, the artificial separation has at the same time created some difficult problems for researchers who wish to gain a detailed understanding of how the nervous system encodes speech signals and how real speakers and listeners deal with the enormous amount of acoustic variability in the speech signal.

Experiments on Talker Variability in Speech Perception

Several novel experiments have been carried out to study the effects of different sources of variability on speech perception and spoken word recognition. We consider these studies to be novel because instead of reducing or eliminating variability in the stimulus materials, as most researchers have routinely done in the past, we specifically introduced variability from different talkers and different speaking rates to study their effects on perception (Pisoni, 1992a). Our research on talker variability began with the observations of Mullennix et al. (1989) who found that the intelligibility of isolated spoken words presented in noise was affected by the number of talkers that were used to generate the test words in the stimulus ensemble. In one condition, all the words in a test list were produced by a single talker; in another condition, the words were produced by 15 different talkers, including both male and female voices. The results which are shown in Figure 1 were very clear. Across three signal-to-noise ratios, identification performance was always better for words that were produced by a single talker than words produced by multiple talkers. Trial-to-trial variability in the speaker's voice apparently affects word recognition performance. This pattern was observed for both high-density (i.e., confusable) and low-density (i.e., non-confusable) words. Our findings in this study replicated results originally reported by Peters (1955) and Creelman (1957) back in the 1950's and suggested to us that the perceptual system must engage in some form of on-line "recalibration" each time a new voice is encountered during the set of test trials.

Insert Figure 1 about here

In a second experiment, we measured naming latencies to the same words presented in both test conditions (Mullennix et al., 1989). Table I provides a summary of the major results. We found that subjects were not only slower to name words from multiple-talker lists but they were also less accurate when their performance was compared to naming words from single-talker lists. Both sets of findings were again surprising to us at the time because all the test words used in the experiment were highly intelligible when presented in the quiet. The intelligibility and naming data from these two experiments immediately raised a number of additional questions about how the various perceptual dimensions of the speech signal are processed by the human listener. At the time, we naturally assumed, as most people did in the past, that the acoustic attributes used to perceive voice quality were independent of the linguistic properties of the signal. However, to our knowledge no one had ever tested this assumption directly.

Insert Table I about here

In another series of experiments, we used a speeded classification task (Garner, 1974) to assess whether attributes of a talker's voice were perceived independently of the phonetic form of the words (Mullennix & Pisoni, 1990). Subjects were required to attend selectively to one stimulus dimension (i.e., voice) while simultaneously ignoring another stimulus dimension (i.e., phoneme). Figure 2 shows the main findings. Across all conditions, we found increases in interference from both dimensions when the subjects were required to attend selectively to only one of the stimulus dimensions. The pattern of results suggested that words and voices were processed as integral dimensions; the perception of one dimension (i.e., phoneme) affects classification of the other dimension (i.e., voice) and vice versa. Subjects apparently cannot selectively ignore irrelevant variation on the non-attended dimension. If both perceptual dimensions were processed separately and independently, as we had originally assumed, we should have found little if any interference from the non-attended dimension, which could be selectively ignored without affecting performance on the attended dimension. Not only did we find mutual interference suggesting that the two sets of dimensions, voice and phoneme, are perceived in a mutually dependent manner but we also found that the pattern of interference was asymmetrical. It was easier for subjects to ignore irrelevant variation in the phoneme dimension when their task was to classify the voice dimension than it was to ignore the voice dimension when they had to classify the phonemes in these stimuli.

Insert Figure 2 about here

The results from these perceptual experiments were surprising given our prior assumption that the indexical and linguistic properties of speech were perceived independently. To study this problem further, we carried out a series of memory experiments to assess the mental representation of spoken words in long-term memory. Experiments on serial recall of lists of spoken words by Martin et al. (1989) and Goldinger et al. (1991) demonstrated that specific details of a talker's voice are also encoded into long-term memory along with the to-be-remembered items. Using a continuous recognition memory procedure, Palmeri et al. (1993) found that detailed episodic information about a talker's voice is also encoded in memory and is

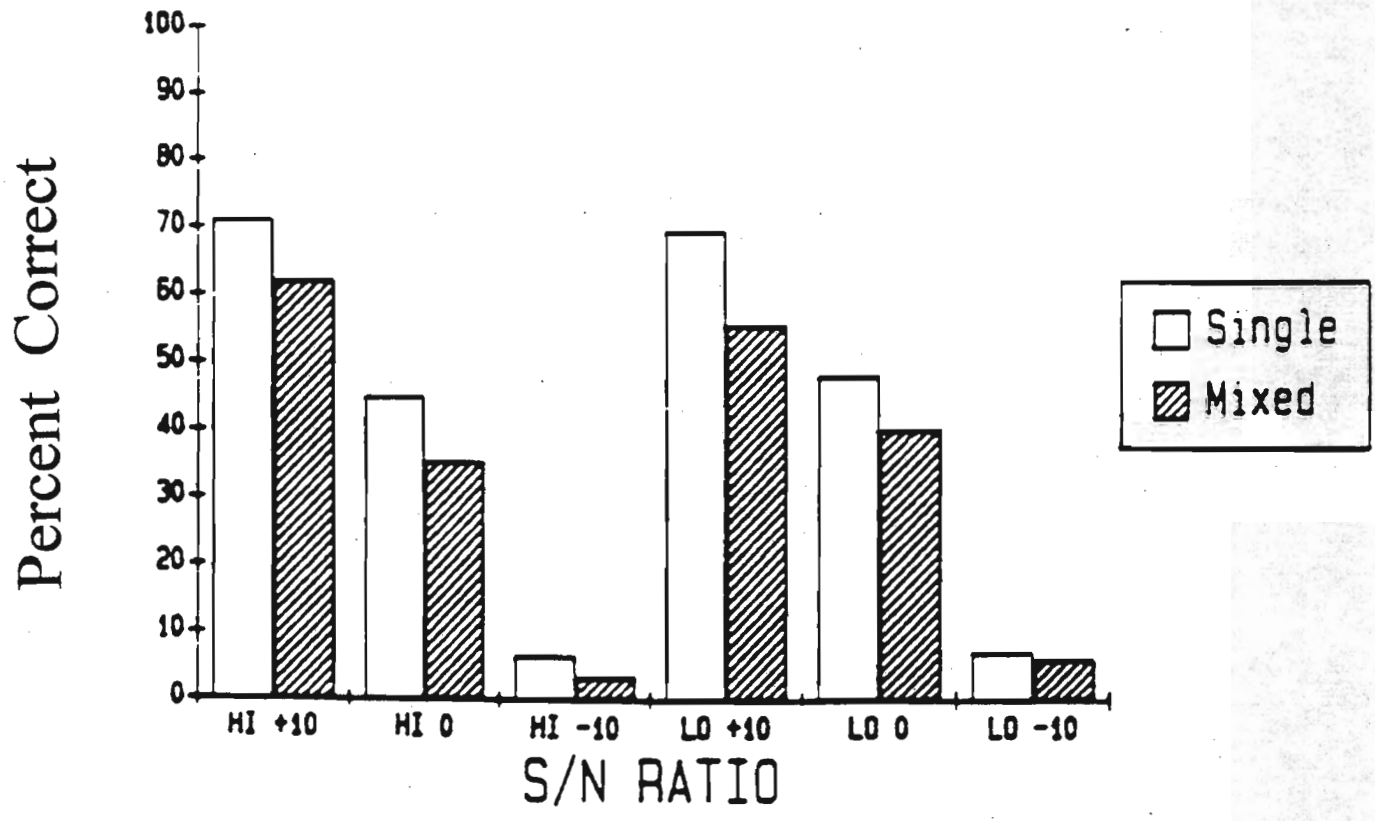


Figure 1. Overall mean percent correct performance collapsed over subjects for single- and mixed-talker conditions as a function of high- and low-density words and S/N ratio (from Mullennix et al., 1989).

Table I

Mean response latency (ms) for correct responses for single- and mixed-talker conditions as a function of lexical density (from Mullennix et al., 1989).

	Density	
	High	Low
Single talker	611.2	605.7
Mixed talker	677.2	679.4

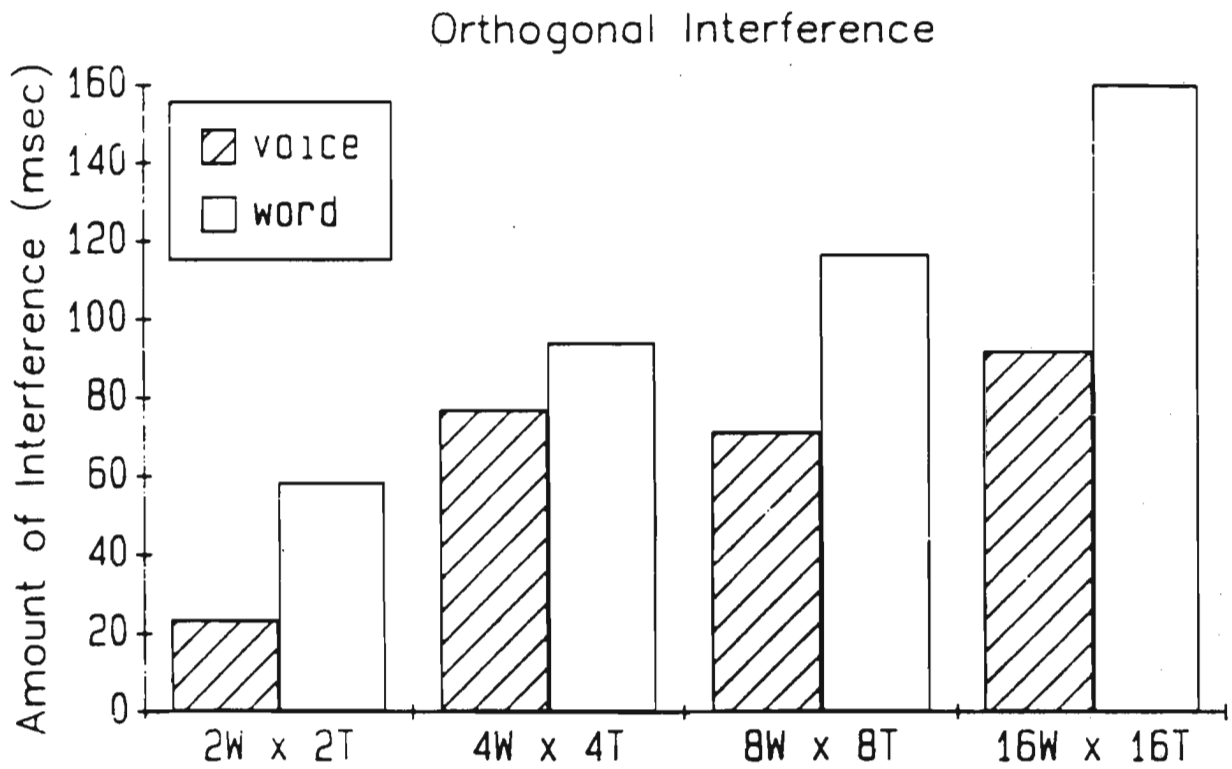


Figure 2. The amount of orthogonal interference (in milliseconds) across all stimulus variability conditions as a function of word and voice dimensions (from Mullennix and Pisoni, 1990).

available for explicit judgments even when a great deal of competition from other voices is present in the test sequence. Palmeri et al.'s results are shown in Figure 3. The top panel shows the probability that an item was correctly recognized as a function of the number of talkers in the stimulus set; the bottom panel shows the probability of a correct recognition across different stimulus lags of intervening items. In both cases, the probability of correctly recognizing a word as "old" (filled circles) was greater if the word was repeated in the same voice than if it was repeated in a different voice of the same gender (open squares) or a different voice of a different gender (open triangles).

In another set of memory experiments, Goldinger (1992) found very strong evidence of implicit memory for attributes of a talker's voice which persist for a relatively long period of time after perceptual analysis has been completed. His results are shown in Figure 4. Goldinger also found that the degree of perceptual similarity affects the magnitude of the repetition effect in memory for identical voices suggesting that the perceptual system encodes very detailed talker-specific information about spoken words in episodic memory representations.

Additional support for the proposal that detailed information about the talker's voice is encoded in memory comes from a recent experiment on sentence recall by Karl and Pisoni (1994). In this study, a cued recall procedure was used to study the retrieval of spoken sentences from long-term memory. After subjects transcribed lists of sentences, they were given a probed recall test with cues presented either visually or auditorily. Recall accuracy depended on the probe cues; when the probe words matched the study conditions, recall was highest, suggesting that detailed information about a talker's voice is encoded in long-term memory and that with the appropriate probe cue at the time of retrieval, this information can be accessed and used to recall the entire sentence.

Insert Figures 3 and 4 about here

Taken together, our recent findings on the effects of talker variability in perception and memory provide support for the proposal that detailed perceptual information about a talker's voice is preserved in long-term memory. At the present time, it is not clear whether there is one "composite" representation in memory or whether these different sets of attributes are encoded in parallel in separate representations (Eich, 1982; Hintzman, 1986). It is also not clear whether spoken words are encoded and represented in memory as a sequence of abstract symbolic "phoneme-like units" along with much more detailed episodic information about specific instances and the processing operations used in perceptual analysis. These are important questions for future research on spoken word recognition.

Experiments on the Effects of Speaking Rate

We also carried out another set of experiments to examine the effects of speaking rate on perception and memory. These studies, which were designed to parallel the earlier experiments on talker variability, have also shown that the perceptual details associated with differences in speaking rate are not lost or discarded as a result of perceptual analysis. In one experiment, Sommers et al. (1994) found that words produced at several different speaking rates (i.e., fast, medium and slow) were identified more poorly than the same words produced at only one speaking rate. These results were compared to another condition in which differences in amplitude were varied randomly from trial to trial in the test sequences. In this case, identification performance was not affected at all by variability in overall signal level. The results from both conditions are shown in Figures 5 and 6.

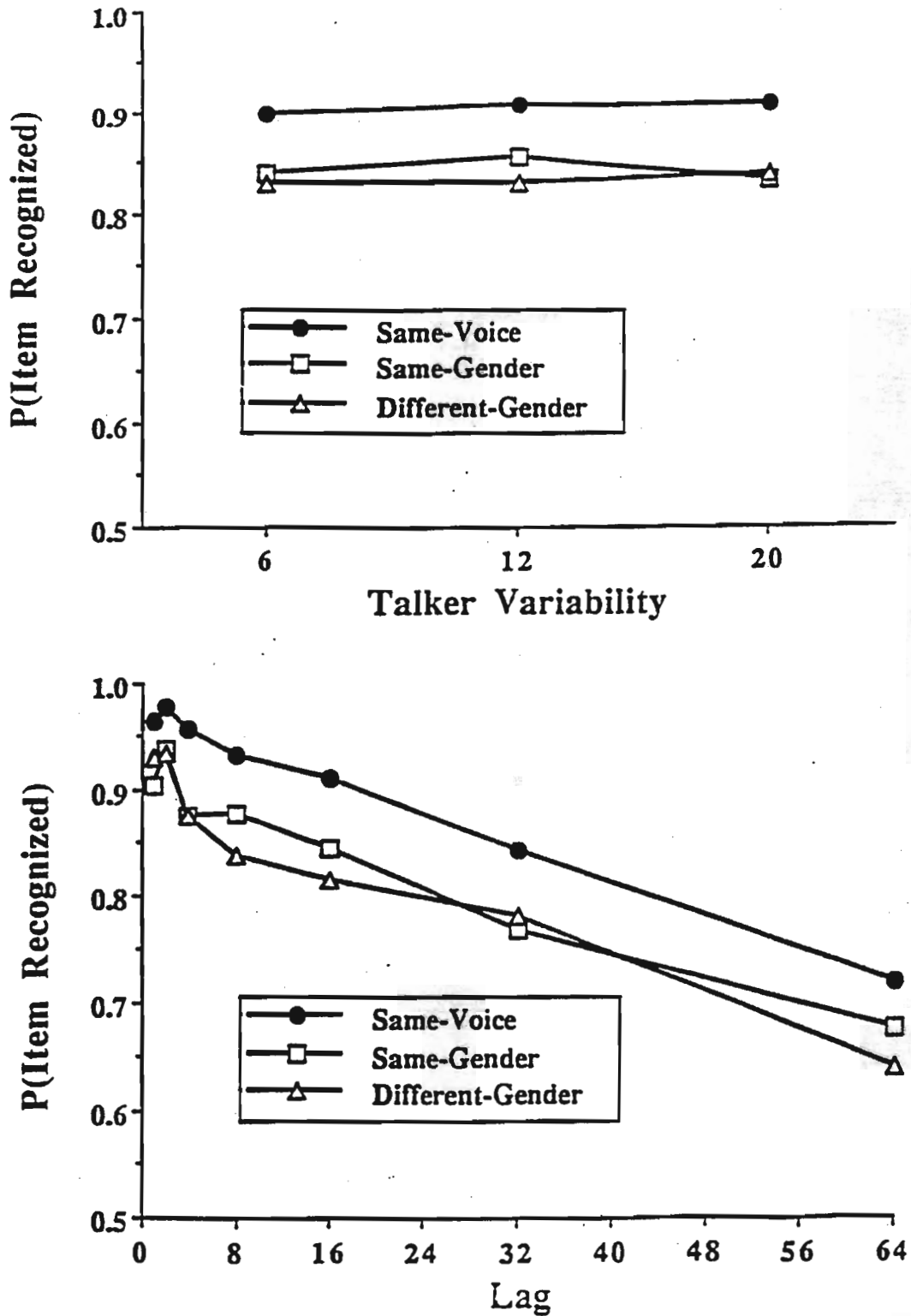


Figure 3. Probability of correctly recognizing old items in a continuous recognition memory experiment. In both panels, recognition for same-voice repetitions is compared to recognition for different-voice/same-gender and different-voice/different-gender repetitions. The upper panel displays item recognition as a function of talker variability, collapsed across values of lag; the lower panel displays item recognition as a function of lag, collapsed across levels of talker variability (from Palmeri et al., 1993).

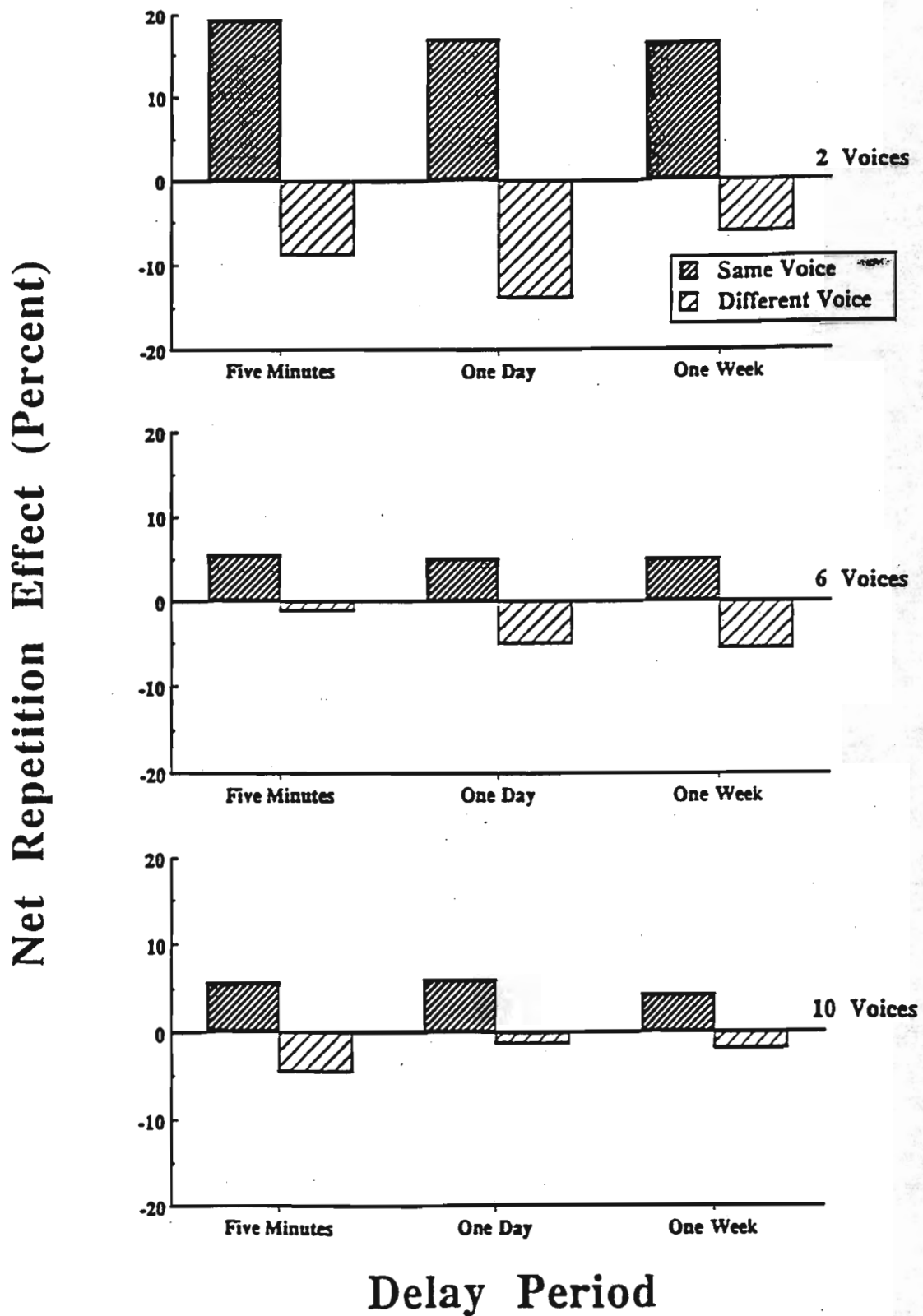


Figure 4. Net repetition effects observed in perceptual identification as a function of delay between sessions and repetition voice (from Goldinger, 1992).

Insert Figures 5 and 6 about here

Other experiments on serial recall have also been completed to examine the encoding and representation of speaking rate in memory. Nygaard et al. (1992a,b) found that subjects recalled words from lists produced at a single speaking rate better than the same words produced at several different speaking rates. Interestingly, the differences appeared in the primacy portion of the serial position curve suggesting greater difficulty in the transfer of items into long-term memory (Luce, Feustel & Pisoni, 1983). Differences in speaking rate, like those observed for talker variability in our earlier experiments, suggest that perceptual encoding and rehearsal processes, which are typically thought to operate on only abstract symbolic representations, are also influenced by low-level perceptual sources of variability. If these sources of variability were "filtered out" or normalized by the perceptual system at relatively early stages of analysis, differences in recall performance would not be expected in memory tasks like the ones used in these experiments.

Taken together with the earlier results on talker variability, the findings on speaking rate suggest that details of the early perceptual analysis of spoken words are not lost as a result of perceptual analysis but instead become an integral part of the neural representation of spoken words in memory. We have also found that in some cases increased stimulus variability in an experiment may actually help listeners to encode items into long-term memory (see Goldinger et al., 1991; Nygaard et al., 1992a,b). Listeners encode speech signals in multiple ways along many perceptual dimensions, and the human memory system apparently preserves these perceptual details much more precisely than researchers believed in the past.

Experiments on Perceptual Learning of Voices

We have also been interested in perceptual learning, specifically the tuning or adaptation that occurs when a listener becomes familiar with the voice of a specific talker (Nygaard, Sommers & Pisoni, 1994). This particular kind of perceptual learning has not received very much attention in the past despite the obvious relevance to problems of speaker normalization, acoustic-phonetic invariance and the potential application to automatic speech recognition and speaker identification (Takehi, 1992; Fowler, In Press). Our search of the research literature on talker adaptation revealed only a small number of studies on this topic and all of them appeared in obscure technical reports from the mid 1950's. Thus, we decided to carry out a perceptual learning experiment of our own to see how knowledge of a talker's voice affects speech perception.

To determine how familiarity with a talker's voice affects the perception of spoken words, we had listeners learn to explicitly identify a set of unfamiliar voices over a nine day period using common a set of names (i.e., Bill, Joe, Sue, Mary). After the subjects learned to recognize the voices explicitly, we presented them with a set of novel words mixed in noise at several signal-to-noise ratios; half the listeners heard the words produced by talkers that they were previously trained on (i. e., the familiar voices) and half the listeners heard the words produced by new talkers that they had not been exposed to previously (i.e., the novel voices). In this phase of the experiment, which was designed to measure speech intelligibility, subjects were now required to identify the words rather than explicitly recognize the voices as they had done in the earlier phase of the experiment.

The results of the speech intelligibility experiment are shown in Figure 7 for the two groups of subjects. We found that identification performance for the trained group was reliably better than the control

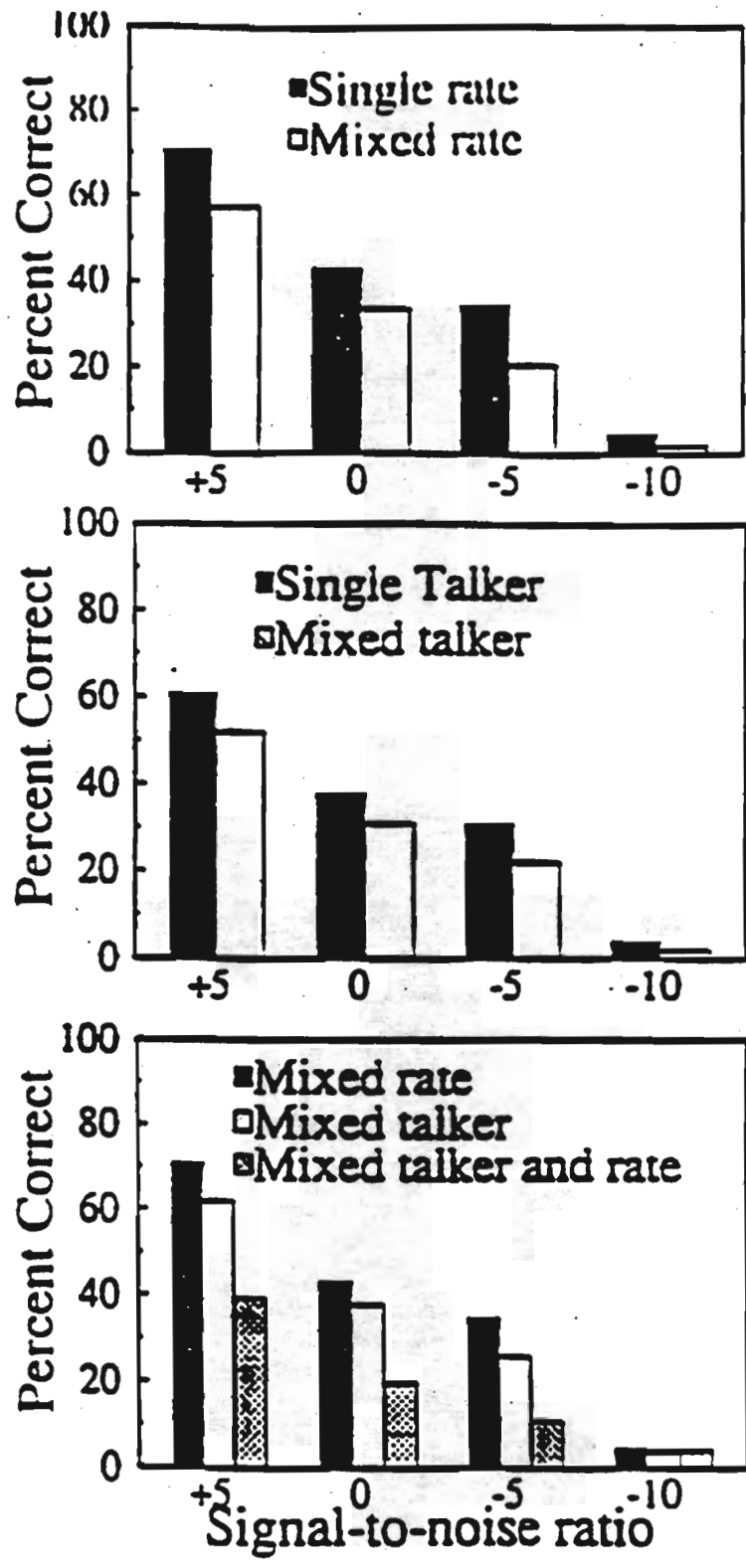


Figure 5. Effects of talker, rate, and combined talker and rate variability on perceptual identification (from Sommers et al., 1992, 1994).

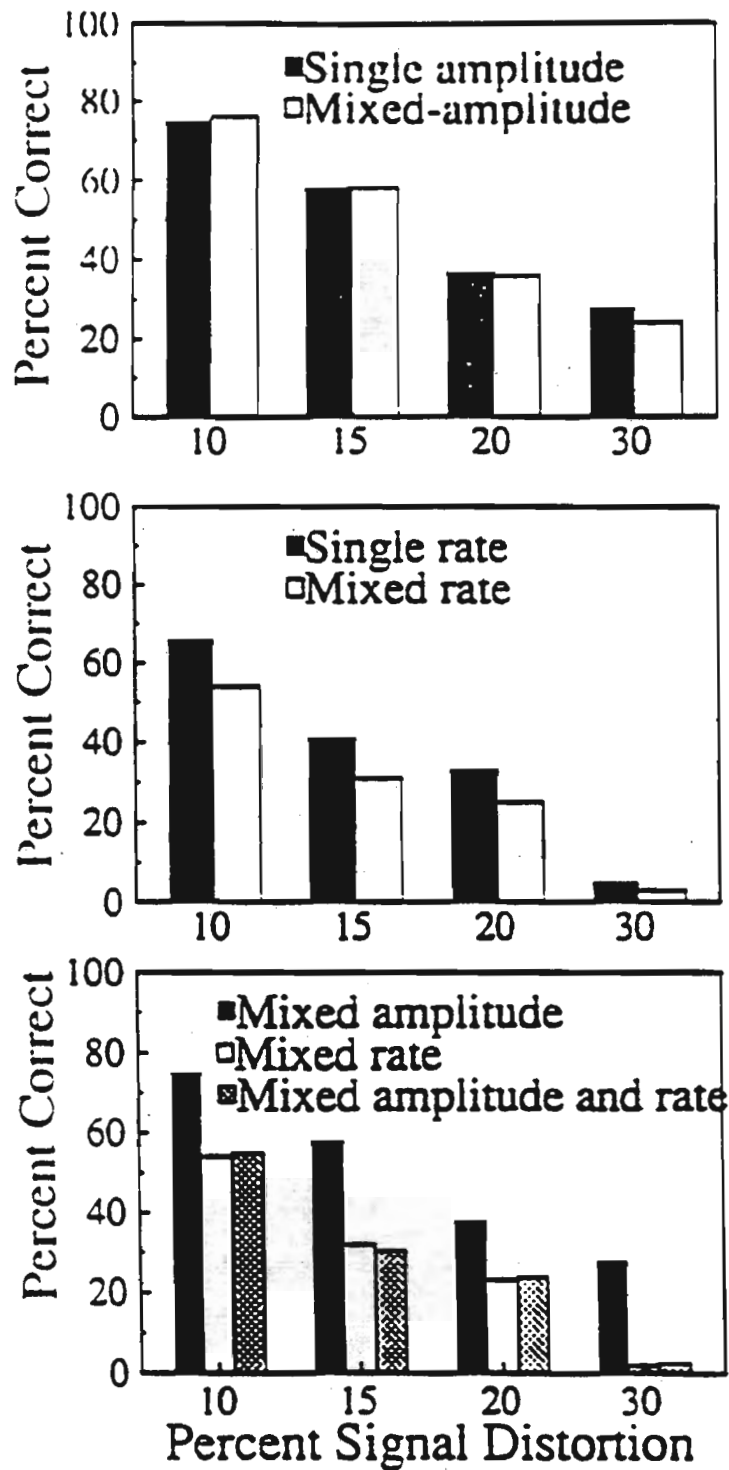


Figure 6. Effects of amplitude, rate, and combined amplitude and rate variability on perceptual identification (from Sommers et al., 1992, 1994).

group at each of the signal-to-noise ratios tested. The subjects who had heard novel words produced by familiar voices were able to recognize words in noise more accurately than subjects who received the same novel words produced by unfamiliar voices. Two other groups of subjects were also run in the intelligibility experiment as controls; however, these subjects did not receive any training and were therefore not exposed to any of the voices prior to hearing the same set of words in noise. One control group received the set of test words presented to the trained experimental group; the other control group received the test words that were presented to the trained control subjects. The performance of both of the control groups was not only same but was equivalent to the intelligibility scores obtained by the trained control group. Only subjects in the experimental group who were explicitly trained on the voices showed the advantage in recognizing novel words produced by familiar talkers.

Insert Figure 7 about here

The findings from this perceptual learning experiment demonstrate that exposure to a talker's voice facilitates subsequent perceptual processing of novel words produced by a familiar talker. Thus, speech perception and spoken word recognition draw on highly specific perceptual knowledge about a talker's voice that was obtained in an entirely different experimental task-- explicit voice recognition as compared to a speech intelligibility test in which novel words were mixed in noise, and subjects identified the items explicitly from an open response set.

What kind of perceptual knowledge does a listener acquire when he listens to a speaker's voice and is required to carry out an explicit name recognition task like our subjects did in this experiment? One possibility is that the analysis procedures or "perceptual operations" (Kolers, 1973) used to recognize the voices are retained in some type of "procedural memory" and these same processing routines are invoked again when the same voice is encountered in a subsequent intelligibility test. This kind of "procedural knowledge" might increase the efficiency of the perceptual analysis for novel words produced by familiar talkers because detailed analysis of the speaker's voice would not have to be carried out again. Another possibility is that specific instances-- "perceptual episodes" or exemplars of each talker's voice are stored in memory and then later retrieved during the process of word recognition when new tokens from a familiar talker are encountered (Jacoby & Brooks, 1984).

Whatever the exact nature of this information or knowledge turns out to be, the important point here is that prior exposure to a talker's voice facilitates subsequent recognition of novel words produced by the same talker. Such findings demonstrate a form of implicit memory for a talker's voice that is distinct from the retention of the individual items used and the specific task that was employed to familiarize the listeners with the voices (Schacter, 1992; Roediger, 1990). These results provide additional support for the view that the neural representation of spoken words encompasses both a phonetic description of the utterance, as well as information about the structural description of the source characteristics of the specific talker. Thus, speech perception appears to be carried out in a "talker-contingent" manner; indexical and linguistic properties of the speech signal are apparently closely interrelated and are not dissociated in perceptual analysis as many researchers previously thought (see Nygaard, et al., 1994). We believe these talker-contingent effects may provide a new way to deal with some of the old problems in speech perception that have been so difficult to resolve in the past.

Intelligibility of Words in Noise

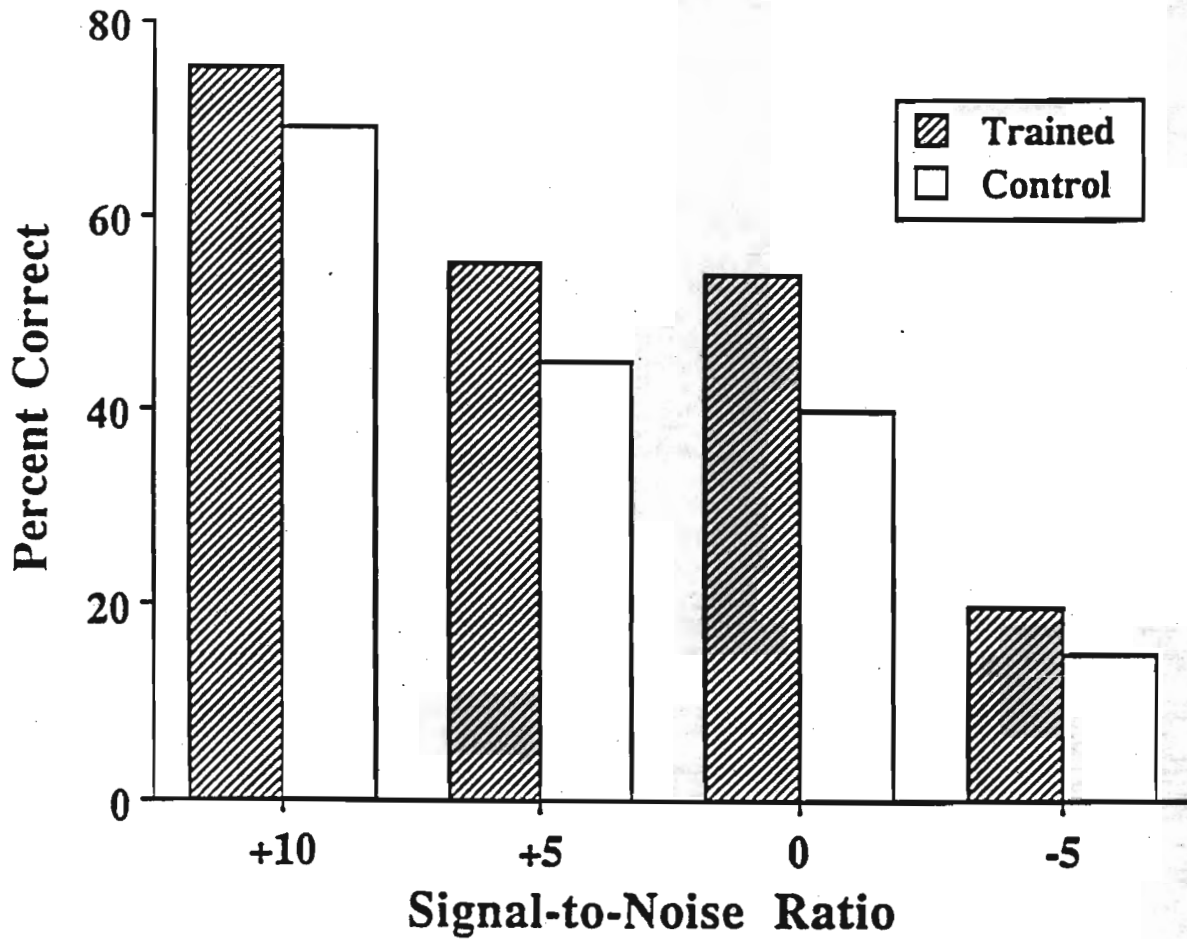


Figure 7. Mean intelligibility of words mixed in noise for trained and control subjects. Percent correct word recognition is plotted at each signal-to-noise ratio (from Nygaard et al., 1994).

Abstractionist vs. Episodic Approaches to Speech Perception

The results we have obtained over the last few years raise a number of important questions about the theoretical assumptions that have been shared for many years by almost all researchers working in the field (Pisoni & Luce, 1986). Within cognitive psychology, the traditional approach to speech perception can be considered among the best examples of what have been called "abstractionist" accounts of categorization and memory (Jacoby & Brooks, 1984). Units of perceptual analysis in speech were assumed to be equivalent to the abstract idealized categories proposed by linguists in their formal analyses of language structure and function. The goal of speech perception studies was to find the physical invariants in the speech signal that mapped onto the symbolic phonetic categories of speech (Studdert-Kennedy, 1976). Emphasis was directed at separating stable, relevant features from the highly variable, irrelevant features of the signal. An important assumption of this traditional approach to perception and cognition was the process of abstraction and the reduction of information in the signal to a more efficient and economical symbolic code (Posner, 1969; Neisser, 1976). Unfortunately, it became apparent very early on in speech research that idealized linguistic units, such as phonemes or phoneme-like units, were highly dependent on the surrounding phonetic context and moreover that a wide variety of factors influenced their physical realization in the speech signal (Stevens, 1971; Klatt, 1986). Nevertheless, the search for acoustic invariance has continued in one way or another and still remains a central problem in the field today.

Recently, a number of studies on categorization and memory in cognitive psychology have provided evidence for the encoding and retention of episodic information and the details of perceptual analysis (Estes, 1994; Jacoby & Brooks, 1984; Brooks, 1978; Tulving & Schacter, 1990; Schacter, 1990). According to this approach, stimulus variability is considered to be "lawful" and "informative" to perceptual analysis (Elman & McClellan, 1986). Memory involves encoding specific instances, as well as the processing operations used in recognition (Kolers, 1973, 1976). The major emphasis of this view of perception and memory is on particulars, rather than abstract generalizations or symbolic coding of the stimulus input into idealized categories. Thus, the problems of variability and invariance found in speech perception can be studied in a different way by non-analytic accounts of perception and memory with its emphasis on encoding of exemplars and specific instances of the stimulus environment rather than the search for physical invariants for abstract symbolic categories.

We believe that the findings from studies on nonanalytic cognition can be generalized to theoretical questions about the nature of perception and memory for speech signals and to assumptions about abstractionist representations based on formal linguistic analyses. When the criteria used for postulating non-analytic representations are examined carefully, it immediately becomes clear that speech signals display a number of distinctive properties that make them especially good candidates for this approach to perception and memory (Jacoby & Brooks, 1984; Brooks, 1978). These criteria are summarized below and can be applied directly to speech perception and spoken language processing.

High Stimulus Variability. Speech signals display a great deal of variability primarily because of factors related to the production of spoken language. Among these are within- and between-talker variability; changes in speaking rate and dialect; differences in social contexts; syntactic, semantic and pragmatic effects; as well as a wide variety of effects due to the ambient environment such as background noise, reverberation and microphone characteristics (Klatt, 1986). These diverse sources of variability consistently produce large changes in the acoustic-phonetic properties of speech and they need to be accommodated in theoretical accounts of speech perception.

Complex Category Relations. The use of phonemes as perceptual categories in speech perception entails a set of complex assumptions about category membership which are based on formal linguistic criteria involving principles such as complementary distribution, free variation and phonetic similarity. The relationship between allophones and phonemes acknowledges explicitly the context-sensitive nature of the category relations that are used to define classes of speech sounds that function in similar ways in different phonetic environments.

Incomplete Information. Spoken language is a highly redundant symbolic system which has evolved to maximize transmission of information. In speech perception, research has demonstrated the existence of multiple speech cues for almost every phonetic contrast. While these speech cues are, for the most part, highly context-dependent, they also provide partial information that can facilitate comprehension of the intended message when the signal is degraded. This feature of speech perception permits high rates of information transmission even under poor listening conditions.

High Analytic Difficulty. Speech sounds are inherently multidimensional in nature. They encode a large number of quasi-independent articulatory attributes that are mapped on to the phonological categories of a specific language. Because of the complexity of speech categories and the high acoustic-phonetic variability, the category structure of speech is not amenable to simple hypothesis testing. As a consequence, it has been extremely difficult to formalize a set of explicit rules that can successfully map speech cues onto a set of idealized phoneme categories. Phoneme categories are also highly automatized. The category structure of a language is learned in a tacit and incidental way by young children. Because the criterial dimensional structures of speech are not typically available to consciousness, it has been difficult to make many aspects of speech perception explicit.

Three Domains of Speech. Among category systems, speech appears to be unique in several respects because of the mapping between production and perception. Speech exists simultaneously in three very different domains: the acoustic domain, the articulatory domain and the perceptual domain. While the relations among these three domains is complex, they are not arbitrary because the sound contrasts used in a language function within a common linguistic signaling system that is assumed to encompass aspects of both production and perception. Thus, the phonetic distinctions generated in speech production by the vocal tract are precisely those same acoustic differences that are important in perceptual analysis (Stevens, 1972). Any theoretical account of speech perception must also take into consideration aspects of speech production and acoustics. The perceptual spaces mapped out in speech production have to be very closely correlated with the same ones used in speech perception. In learning the sound system of a language, the child must not only develop abilities to discriminate and identify sounds, but he/she must also be able to control the motor mechanisms used in articulation to generate precisely the same phonetic contrasts in speech production that he/she has become attuned to in perception. One reason that the developing perceptual system might preserve very fine phonetic details as well as characteristics of the talker's voice would be to allow a young child to accurately imitate and reproduce speech patterns heard in the surrounding language learning environment (Studdert-Kennedy, 1983). This skill would provide the child with an enormous benefit in acquiring the phonology of the local dialect from speakers he/she is exposed to early in life.

General Discussion

It has become common over the last 25 years to argue that speech perception is a highly unique process that requires specialized neural processing mechanisms to carry out perceptual analysis (Lieberman, Cooper, Shankweiler & Studdert-Kennedy, 1967). These theoretical accounts of speech perception have

typically emphasized the differences in perception between speech and other perceptual processes. Relatively few researchers working in the field of speech perception have tried to identify commonalities among other perceptual systems or draw parallels with speech perception. Our recent findings on the encoding of different sources of variability in speech and the role of long-term memory for specific instances are compatible with a rapidly growing body of research in cognitive psychology on implicit memory phenomena and non-analytic modes of processing (Jacoby & Brooks, 1984; Brooks, 1978).

Traditional memory research has been concerned with "explicit memory" in which the subject is required to consciously access and manipulate recently presented information from memory using "direct tests" such as recall or recognition. This line of memory research has had a long history in experimental psychology and it is an area that most speech researchers are familiar with. In contrast, the recent literature on "implicit memory" phenomena has provided new evidence for unconscious aspects of perception, memory and cognition (Schacter, 1992; Roediger, 1990). Implicit memory refers to a form of memory that was acquired during a specific instance or episode and it is typically measured by "indirect tests" such as stem completion, cued recall, priming or changes in perceptual identification performance (Roediger, 1990; Roediger & McDermott, 1993). In these types of memory tests, subjects are not required to consciously recollect previously acquired information. In fact, in many cases, especially in processing spoken language, subjects may be unable to access the information deliberately or even bring it to consciousness (Studdert-Kennedy, 1974).

Studies of implicit memory have uncovered important new information about the effects of prior experience on perception and memory. In addition to traditional abstractionist modes of cognition which tend to emphasize symbolic coding of the stimulus input, recent experiments have provided evidence for a parallel non-analytic memory system that preserves specific instances of stimulation as perceptual episodes or exemplars which are also stored in memory. These perceptual episodes have been shown to affect later processing activities. We believe that it is this implicit perceptual memory system that encodes the indexical information in speech about a talker's gender, dialect and speaking rate. And, we believe that it is this memory system that encodes and preserves the perceptual operations or procedural knowledge that listeners acquire about specific voices that facilitates later recognition of novel words produced by familiar speakers.

Our findings demonstrating that spoken word recognition is talker-contingent and that familiar voices are encoded differently than novel voices, raises a new set of questions concerning the long-standing dissociation between the linguistic properties of speech-- the abstract, symbolic features, phonemes and words used to convey the linguistic message-- and the indexical properties of speech-- those personal or paralinguistic attributes of the speech signal which provide the listener with information about the form of the message, such as the speaker's gender, dialect, social class, and emotional state among other things. In the past, these two sources of information were separated for purposes of linguistic analysis of the message. The present set of findings suggest this may have been an incorrect assumption.

Relative to the research carried out on the linguistic properties of speech, which has a history dating back to the late 1940's, much less is known about perception of the acoustic correlates of the indexical or paralinguistic functions of speech (Ladefoged, 1975; Laver & Trudgill, 1979). While there have been a number of recent studies on explicit voice recognition and identification by human listeners (Papcun, Kreiman & Davis, 1989), very little research has been carried out on problems surrounding the "implicit" or "unconscious" encoding of attributes of voices and how this form of memory might affect the recognition process associated with the linguistic attributes of spoken words (Goldinger, 1992; Lively, 1994). A question that naturally arises in this context is whether or not familiar voices are processed

differently than unfamiliar or novel voices. Perhaps familiar voices are simply recognized more efficiently than novel voices and are perceived in fundamentally the same way by the same neural mechanisms as unfamiliar voices? The available evidence in the literature has shown, however, that familiar and unfamiliar voices are processed differentially by the two hemispheres of the brain and that selective impairment resulting from brain language can affect the perception of familiar and novel voices in very different ways (see Kreiman & VanLancker, 1988; VanLancker, Cummings, Kreiman & Dobkin, 1988; VanLancker, Kreiman & Cummings, 1989).

Most researchers working in speech perception have adopted a common set of theoretical assumptions about the units of linguistic analysis and the goals of perceptual processing of speech signals. The primary objective was to extract the speaker's message from the acoustic waveform without regard to the source (Studdert-Kennedy, 1974). The present set of findings suggests that while the dissociation between indexical and linguistic properties of speech may have been a useful dichotomy for linguists who approach language as a highly abstract formalized symbolic system, the same set of assumptions may no longer be useful for speech scientists who are interested in describing and modeling how the human nervous system encodes speech signals and represents this information in long-term memory.

Our recent findings on variability suggest that fine phonetic details about the form of the signal are not lost as a consequence of perceptual analysis as widely assumed by researchers in the past. Attributes of the talker's voice are also not lost or normalized, at least not immediately after perceptual analysis has been completed. In contrast to the theoretical views that were very popular a few years ago, the present findings have raised some new questions about the problems of variability, invariance and perceptual normalization. For example, there is now sufficient evidence from perceptual experimentation to suggest that the fundamental perceptual categories of speech-- phonemes and phoneme-like units-- are probably not as rigidly fixed or well-defined physically as theorists once believed. These perceptual categories appear to be highly variable and their physical attributes have been shown to be strongly affected by a wide variety of contextual factors (Klatt, 1979). It seems very unlikely after some 45 years of research on speech that very simple physical invariants for phonemes will be uncovered from analysis of the speech signal. If invariants are uncovered they will probably be very complex time-varying cues that are highly context-dependent.

Many of the theoretical views that speech researchers have held for a long time about language were motivated by linguistic considerations of speech as an idealized symbolic system essentially free from physical variability. Indeed, variability in speech was considered by many researchers to be a source of "noise"-- an undesirable set of perturbations on what was otherwise supposed to be an idealized sequence of abstract symbols arrayed linearly in time. Unfortunately, it has taken many years for speech researchers to realize that variability is an inherent characteristic of all biological systems including speech. Rather than view variability as noise, some theorists have recognized that variability might actually be useful and informative to human listeners who are able to encode speech signals in variety of different ways depending upon the circumstances and demands of the listening task (Elman & McClellan, 1986). The recent proposals in the human memory literature for multiple memory systems suggest that the internal representation of speech is probably much more detailed and much more elaborate than previously believed from simply an abstractionist linguistic point of view. The traditional views about features, phonemes and acoustic-phonetic invariance are no longer adequate to accommodate the new findings that have been uncovered concerning context effects and variability in speech perception and spoken word recognition. In the future, it may be very useful to explore the parallels between similar perceptual systems such as face recognition and voice recognition. There is, in fact, some reason to suspect that parallel neural mechanisms may be employed in each case despite the obvious differences in modalities.

Conclusions

The results summarized in this paper on the role of variability in speech perception are compatible with non-analytic or instance-based views of cognition which emphasize the episodic encoding of specific details of the stimulus environment. Our studies on talker and rate variability and our new experiments on perceptual learning of novel voices have provided important information about speech perception and spoken word recognition and have served to raise a set of new theoretical questions for future research. In this section, I simply list the major conclusions.

First, our findings raise questions about previous views of the neural representation of speech. In particular, we have found that detailed instance-specific information about the source characteristics of a talker's voice are encoded into long-term memory. Whatever the internal representation of speech turns out to be, it is clear that it is not isomorphic with the linguist's description of speech as an abstract idealized sequence of segments. Mental representations of speech are much more detailed and more elaborate and they contain several sources of information about the talker's voice.

Second, our findings suggest a different approach to the problem of acoustic-phonetic variability in speech perception. Variability is not a source of noise; it is lawful and informative and provides potentially useful knowledge about the characteristics of a talker's voice and speaking rate as well as the phonetic context. These sources of information appear to be accessed when a listener hears novel words or sentences produced by a familiar talker. Variability provides important talker-specific information that affects encoding fluency and processing efficiency in a variety of tasks.

Third, our findings provide additional evidence that speech perception is highly sensitive to context and that details of the input signal are not lost or filtered out as a consequence of perceptual analysis. These results are consistent with recent proposals for the existence of multiple memory systems and the role of perceptual representation systems (PRS) in memory and learning. The present findings also suggest a somewhat different view of the process of perceptual normalization which has generally focused on abstraction and stimulus reduction in categorization of speech sounds.

Finally, the results described here suggest several new directions for models of speech perception and spoken word recognition. These models are motivated by a different set of criteria than traditional abstractionist approaches to perception and memory. Exemplar-based or episodic models of categorization which emphasize instance-specific encoding provide a viable new theoretical alternative to the problems of invariance, variability and perceptual normalization that have been difficult to resolve with current models of speech perception that were inspired by formal linguistic analyses of language. We believe that many of the current theoretical problems in the field of speech perception can be approached in quite different ways when viewed within the general framework of non-analytic or instance-based models of cognition which have alternative methods of dealing with the problems of stimulus variability, context effects and perceptual learning phenomena which have been the hallmarks of human speech perception for many years.

REFERENCES

- Brooks, L. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch and B. Lloyd (Eds.), *Cognition and Categorization*. Hillsdale, NJ: Erlbaum.
- Creelman, C.D. (1957). Case of the unknown talker. *Journal of the Acoustical Society of America*, **29**, 655.
- Eich, J.E. (1982). A composite holographic associative memory model. *Psychological Review*, **89**, 627-661.
- Elman, J.L. & McClellan, J.L. (1986). Exploiting lawful variability in the speech wave. In J. S. Perkell and D. H. Klatt (Eds.), *Invariance and Variability in Speech Processes*. Hillsdale, NJ: Erlbaum, pp. 360-380.
- Estes, W.K. (1994). *Classification and Cognition*. New York: Oxford University Press.
- Fowler, C.A. (In Press). Listener-talker attunements in speech. In T. Tighe, B. Moore, and J. Santrock (Eds.), *Human Development and Communication Sciences*. Hillsdale, NJ: Erlbaum.
- Garner, W.R. (1974). *The Processing of Information and Structure*. Potomac, MD: Erlbaum.
- Goldinger, S.D. (1992). Words and Voices: Implicit and Explicit Memory for Spoken Words. *Research on Speech Perception Technical Report No. 7*, Indiana University, Bloomington, IN.
- Goldinger, S.D., Pisoni, D.B. & Logan, J.S. (1991). On the locus of talker variability effects in recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **17**, 152-162.
- Hintzman, D.L. (1986). Schema abstraction in a multiple-trace memory model. *Psychological Review*, **93**, 411-423.
- Jacoby, L.L. & Brooks, L.R. (1984). Nonanalytic cognition: Memory, perception, and concept learning. In G. Bower (Ed.), *The Psychology of Learning and Motivation, Vol. 18*, New York: Academic Press, pp. 1-47.
- Takehi, K. (1992). Adaptability to differences between talkers in Japanese monosyllabic perception. In Y. Tohkura, E. Vatikiotis-Bateson and Y. Sagisaka (Eds.), *Speech Perception, Production and Linguistic Structure*. Tokyo, Japan: IOS Press, Inc.
- Karl, J. R. & Pisoni, D. B. (1994). The role of talker-specific information in memory for spoken sentences. *Journal of the Acoustical Society of America*, **95**, 2873.
- Klatt, D.H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics*, **7**, 279-312.

- Klatt, D.H. (1986). The problem of variability in speech recognition and in models of speech perception. In J.S. Perkell and D.H. Klatt (Eds.), *Invariance and Variability in Speech Processes*. Hillsdale, NJ: Erlbaum.
- Kolers, P.A. (1973). Remembering operations. *Memory & Cognition*, **1**, 347-355.
- Kolers, P.A. (1976). Pattern analyzing memory. *Science*, **191**, 1280-1281.
- Kreiman, J. & VanLancker, D. (1988). Hemispheric specialization for voice recognition: Evidence from dichotic listening. *Brain and Language*, **34**, 246-252.
- Ladefoged, P. (1975). *A Course in Phonetics*. New York: Harcourt Brace Jovanovich, Inc.
- Laver, J. & Trudgill, P. (1979). Phonetic and linguistic markers in speech. In K.R. Scherer and H. Giles (Eds.), *Social Markers in Speech*. Cambridge: Cambridge University Press, pp. 1-31.
- Liberman, A.M., Cooper, F.S., Shankweiler, D.P. & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, **74**, 431-461.
- Lively, S.E. (1994). Preserving the Perceptual Record: Retention of Voice Information in Long-Term Memory. *Research on Speech Perception Technical Report No. 9*, Indiana University, Bloomington, IN.
- Luce, P.A., Feustal, T.C. & Pisoni, D.B. (1983). Capacity demands in short-term memory for synthetic and natural word lists. *Human Factors*, **25**, 17-32.
- Martin, C.S., Mullennix, J.W., Pisoni, D.B. & Summers, W.V. (1989). Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **15**, 676-684.
- Mullennix, J.W. & Pisoni, D.B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics*, **47**, 379-390.
- Mullennix, J. W., Pisoni, D.B. & Martin, C.S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America*, **85**, 365-378.
- Neisser, U. (1976). *Cognitive Psychology*. New York: Appleton-Century-Crofts.
- Nygaard, L.C., Sommers, M.S. & Pisoni, D.B. (1992). Effects of speaking rate and talker variability on the recall of spoken words. *Journal of the Acoustical Society of America*, **91**, 2340.
- Nygaard, L.C., Sommers, M.S. & Pisoni, D.B. (1992). Effects of speaking rate and talker variability on the representation of spoken words in memory. *Proceedings 1992 International Conference on Spoken Language Processing, Banff, Canada, 12-17 October 1992*.
- Nygaard, L.C., Sommers, M.S. & Pisoni, D.B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, **5**, 42-46.

- Palmeri, T.J., Goldinger, S.D. & Pisoni, D.B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **19**, 1-20.
- Papcun, G., Kreiman, J. & Davis, A. (1989). Long-term memory for unfamiliar voices. *Journal of the Acoustical Society of America*, **85**, 913-925.
- Peters, R.W. (1955). The relative intelligibility of single-voice and multiple-voice messages under various conditions of noise. *Joint Project Report No. 56, U.S. Naval School of Aviation Medicine*, pp. 1-9. Pensacola, FL.
- Pisoni, D.B. (1990). Effects of talker variability on speech perception: Implications for current research and theory." *Proceedings of 1990 International Conference on Spoken Language Processing*, Kobe, Japan, pp. 1399-1407.
- Pisoni, D.B. (1992a). Some comments on talker normalization in speech perception. In Y. Tohkura, E. Vatikiotis-Bateson and Y. Sagisaka (Eds.), *Speech Perception, Production and Linguistic Structure*. Tokyo, Japan: IOS Press, Inc.
- Pisoni, D.B. (1992b). Some comments on invariance, variability and perceptual normalization in speech perception. *Proceedings 1992 International Conference on Spoken Language Processing, Banff, Canada, 12-17 October 1992*.
- Pisoni, D.B. (1993). Long-term memory in speech perception: Some new findings on talker variability, speaking rate and perceptual learning. *Speech Communication*, **13**, 109-125.
- Pisoni, D.B. & Luce, P.A. (1986). Speech reception: Research, theory, and the principal issues. In E.C. Schwab and H.C. Nusbaum (Eds.), *Pattern Recognition by Humans and Machines*. New York: Academic Press, pp. 1-50.
- Pisoni, D.B., Nusbaum, H.C., Luce, P.A. & Slowiaczek, L.M. (1985). Speech perception, word recognition and the structure of the lexicon. *Speech Communication*, **4**, 75-95.
- Posner, M.I. (1969). Abstraction and the process of recognition. In J.T. Spence and G.H. Bower (Eds.), *The Psychology of Learning and Motivation: Advances in Learning and Motivation*. New York: Academic Press.
- Roediger, H.L. (1990). Implicit memory: Retention without remembering. *American Psychologist*, **45**, 1043-1056.
- Roediger, H.L. & McDermott, K.B. (1993). Implicit memory in normal human subjects. In F. Boller and J. Grafman (Eds.) *Handbook of Neuropsychology*. New York: Elsevier Publishing.
- Schacter, D.L. (1990). Perceptual representation systems and implicit memory: Toward a resolution of the multiple memory systems debate. In A. Diamond (Ed.), *Development and Neural Basis of Higher Cognitive Function. Annals of the New York Academy of Sciences*, **608**, 543-571.

- Schacter, D.L. (1992). Understanding implicit memory: A cognitive neuroscience approach. *American Psychologist*, **47**, 559-569.
- Sommers, M.S., Nygaard, L.C. & Pisoni, D.B. (1992). Stimulus variability and the perception of spoken words: Effects of variations in speaking rate and overall amplitude. *Proceedings 1992 International Conference on Spoken Language Processing, Banff, Canada, 12-17 October 1992*.
- Sommers, M.S., Nygaard, L.C. & Pisoni, D.B. (1994). Stimulus variability and spoken word recognition. I. Effects of variability in speaking rate and overall amplitude. *Journal of the Acoustical Society of America*, **96**, 1314-1324.
- Stevens, K.N. (1971). Sources of inter- and intra-speaker variability in the acoustic properties of speech sounds. *Proceedings of the Seventh International Congress of Phonetic Sciences*. The Hague: Mouton.
- Stevens, K.N. (1972). The quantal nature of speech: Evidence from articulatory acoustic data. In E.E. David, Jr. and P.B. Denes (Eds.), *Human Communication: A Unified View*. New York: McGraw-Hill.
- Studdert-Kennedy, M. (1974). The perception of speech. In T.A. Sebeok (Ed.) *Current Trends in Linguistics*, The Hague: Mouton, pp. 2349-2385.
- Studdert-Kennedy, M. (1976). Speech perception. In N.J. Lass (Ed.), *Contemporary Issues in Experimental Phonetics*. New York: Academic Press.
- Studdert-Kennedy, M. (1983). On learning to speak. *Human Neurobiology*, **2**, 191-195.
- Tulving, E. & Schacter, D.L. (1990). Priming and human memory systems. *Science*, **247**, 301-306.
- VanLancker, D.R., Cummings, J.L., Kreiman, J. & Dobkin, B. (1988). Phonagnosia: A dissociation between familiar and unfamiliar voices. *Cortex*, **24**, 195-209.
- VanLancker, D.R., Kreiman, J. & Cummings, J. (1989). Voice perception deficits: Neuroanatomical correlates of phonagnosia. *Journal of Clinical and Experimental Neuropsychology*, **11**, 665-674.