RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 24 (2000)
Indiana University

Speech Perception and Implicit Memory: Evidence for Detailed Episodic Encoding of Phonetic Events[1]

Lorin Lachs, Kipp McMichael and David B. Pisoni[2]

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

[1] This work was supported by NIH RO1 DC00111 and NIH T32 DC00012. Thanks to Luis Hernàndez and Darla Sallee for assistance with the final preparation of this paper.

[2] Also Devault Otologic Research Laboratory, Department of Otolaryngology-Head & Neck Surgery, Indiana University School of Medicine, Indianapolis, IN.

Speech Perception and Implicit Memory: Evidence for Detailed Episodic Encoding of Phonetic Events

Abstract. Recent investigations into the nature of memory for spoken words have demonstrated that detailed, episodic information about the voice of the talker is encoded along with more abstract information about the linguistic content of an utterance. These findings present serious challenges to traditional views of speech perception, in which the process of abstraction plays a major role. We first outline the traditionalist view of speech perception, a theoretical framework largely based on the constructs of formal linguistics. Next, we elaborate the recently emerging "detailed encoding" view and summarize recent evidence from perceptual and memory experiments that demonstrate that "linguistic" and "extra-linguistic" information in spoken language are inseparable in the acoustic signal and in the representation of speech in long-term memory.

Introduction

Nearly every aspect of human speech - our accents, word choice, and even the very language we utter – is influenced by past experience. The perceptual process occurs very quickly and often appears to be carried out almost automatically. For the most part, we rarely, if ever, have any conscious awareness of our linguistic knowledge or our previous experience during speech production or perception. These general observations about speech perception suggest that implicit memory processes may play a pervasive and perhaps indissociable role in both speech perception and production. Yet despite a widespread acceptance by researchers that all behavior is ultimately grounded in prior, long-term experience, the role of implicit memory in speech production and perception has only been the focus of experimental inquiry by cognitive psychologists within the last few years.

To explain why implicit memory research in speech perception has only recently emerged, we begin this chapter with a review and discussion of the theoretical and meta-theoretical notions that underlie the traditional, abstractionist characterization of speech perception. Once we have described the traditional framework, we move on to an emerging view of speech processing and memory where both explicit and implicit effects find a unified, straightforward explanation. Finally, we will expand this emerging view to show that it is highly compatible with a seamless, undichotomized human memory system that incorporates both implicit and explicit memory components.

Speech Perception: The Abstractionist Perspective

At its inception, the field of speech science borrowed many of its constructs and conceptualizations about language from formal linguistics. Perceptual units such as phonemes, allophones, morphemes, and even words themselves were simply direct transplantations from linguistic theory. Even after extensive analysis of speech spectrograms made it clear that speech was nothing like a discrete sequence of idealized segments (Liberman, 1957; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967), researchers continued to maintain that speech was, in essence, a discrete, symbolic signal (Licklider, 1952).

Under this view, then, speech is reduced to spoken text. The viewpoint was so widely accepted that Morris Halle, the noted linguist, went so far as to say:

> 'There can be no doubt that speech is a sequence of discrete entities, since in writing we perform the kind of symbolization just mentioned, while in reading aloud we execute the inverse of this operation; that is, we go from a discrete symbolization to a continuous acoustic signal.' (1956)

Not all views of speech perception have had such a literalist reading of the 'speech as spoken text' hypothesis, but accepted meta-theoretical notions about the discrete, idealized, symbolic nature of speech have been the dominant influence on research in speech perception and production for more than fifty years. Under this view, outlined quite explicitly in early work by phoneticians like Abercrombie (1967), a fundamental distinction is drawn between the *language* and the *medium* that mediates between speech production and reception. For example, the written word is a visible medium that transfers the 'language' produced by a writer to that received by the reader. Likewise, the audible signal generated by the talker's vocal tract during speech production transfers 'language' via acoustic medium to the listener. Because a physical medium has properties that are not related to the communication of language, a dichotomy arises concerning information contained in the physical signal:

> '…all that is necessary for linguistic communication is that the contrasts on which the patterns are based should not be obscured. Usually, therefore, many things about a medium which is being used as vehicle for a given language are not relevant to linguistic communication. Such 'extra-linguistic' properties of the medium, however, may fulfill other functions which may sometimes even be more important than the linguistic communication, which can never be completely ignored.' (Abercrombie, 1967, p.5)

While certainly acknowledging the utility of 'extra-linguistic' variation in the speech signal, this passage illustrates two very important aspects of the traditionalist view of speech and language. First, the passage states clearly that the primary function of any language pattern is the communication of contrast, which can be used to recover the linguistic content of a message. Second, and more importantly, the extra-linguistic properties of the signal are *defined* as the exclusive complement to linguistic properties encoded in the signal. That is, any property of a medium that was *not relevant* to signaling the linguistic content was considered to be extra-linguistic. By this view, extra-linguistic information is simply a source of undesirable 'noise' created in the physiological realization of the idealized speech signal.

Linguistic content, then, is information specifying the underlying, linguistic representation of an utterance, such as segments, phonemes, syllables or other idealized, symbolic units like words. Extra-linguistic content is everything else in the signal. Abercrombie (1967) describes the importance of extra-linguistic content, pointing out that it may contain signs or indices of other, non-linguistically important information about the talker. These *indexical* features of the speech signal –as opposed to linguistic features - might include such things as the talker's gender, dialect, or affect. However, it is precisely the dissociation between linguistic and extra-linguistic information in speech that, in our view, makes this traditional account of spoken language questionable at the present time. Over the last few years, many new findings about the contribution of extra-linguistic information to speech perception have been reported in the literature. These findings suggest that the traditional dichotomy between linguistic and extra-linguistic information in the speech signal may be somewhat misleading and possibly an incorrect characterization of the sensory information that human listeners perceive, encode and store about their language.

Reconstruction and Abstraction

The notion that speech is a noisy and highly degraded signal that fails to perfectly transmit the intended utterance of the speaker led to reconstructionist accounts of speech perception. In the words of

Neisser (1967), 'There must be a kind of filtering, of feature-detection, which precedes the active construction of detail.' (p.196). According to this view, the impoverished acoustic signal is processed extensively to uncover the underspecified linguistic message that is encoded in the speech waveform. Based on rules or schema derived from acquired linguistic knowledge, the speech signal is further processed to construct an accurate perception of the intended utterance. This view of speech was extremely compatible with the information-processing framework of early cognitive psychology (Studdert-Kennedy, 1974; Studdert-Kennedy, 1976), even as J.J. Gibson's approach to perception challenged the notions of underspecification and reconstruction in the field of perception more generally (Gibson, 1966).

The process of speech perception is, according to traditional accounts, a cleaning up or filtering mechanism that uncovers sequences of idealized units such as phonemes, or words. These abstractionist accounts of speech (Pisoni, 1997) make extra-linguistic information unavailable for encoding into memory for speech events – unless some ad hoc reintegration process is proposed before storage. Thus, the long-term memory store of spoken words and knowledge about words - the mental lexicon - necessarily becomes a formalized, idealized, abstract database of linguistic information, a large collection of symbolic representations of words in the language.

This view of speech has motivated a very specific set of research questions and encouraged the development of experimental methodologies that have been prevalent over the last 50 years. Because extra-linguistic variation was thought to obscure the 'real' objects of speech perception–the underlying, abstract, symbolic linguistic units–factors related to the talker's voice, speaking rate, dialect and affect were either eliminated from experimental designs or explicitly controlled so that effects of these 'irrelevant' factors would not obscure the 'interesting' phenomena more directly related to linguistic communication. Hundreds of experiments on speech perception have studied the perception of utterances spoken by a single talker or the perception of highly controlled 'minimal' units of language, like features or phonemes, in CV nonsense syllables using highly controlled synthetic speech signals (see Liberman et al., 1967).

As a consequence, this research paradigm has provided very little information relating to the human listener's remarkable ability to perceive speech accurately and robustly under a wide variety of conditions and circumstances. We take this ability to deal with enormous stimulus variability in the signal to be of paramount importance to the process of spoken word recognition (Klatt, 1989). Indeed, the usefulness of a linguistic system is severely, if not totally, called into question if it is highly susceptible to drastic and unrecoverable interference as a result of the seemingly limitless conditions under which spoken language is used. Ironically, the lack of research into speech variation and variability and the ways in which listeners deal with these perceptual problems is potentially quite damaging to our understanding of spoken communication. In our view, the traditional abstractionist, symbolic, or "symbol-processing" framework can no longer be accepted without serious question as to its utility. We now turn to an alternative theoretical framework in which the importance of stimulus variation is acknowledged and made explicit: the detailed encoding perspective.

Speech Perception: A Detailed Encoding Perspective

The time-varying acoustic signal that impinges upon the ears of the listener is not one that is neatly divided into linguistic and extra-linguistic information. The acoustic signal of speech simultaneously carries information about the linguistic utterance as well as information about the source of the utterance and the listener's communicative circumstances. In other words, linguistic and extra-linguistic information are mixed together and fundamentally inseparable in their initial acoustic form.

In contrast to the traditional views of speech and speech perception, then, one can consider the object of speech processing as a very rich, detailed representation of the original articulatory events that created the signal (Fowler, 1986; Goldinger, 1998). Since this representation incorporates both linguistic and extra-linguistic information, we need not puzzle over how the abstract, idealized, and formalized units of language are first separated from the extra-linguistic information in the speech signal and later recombined for subsequent semantic processing, where information such as gender, dialect or affect become relevant.

Rather than viewed as a filtering or abstracting mechanism, the nature of speech perception and processing in a detailed encoding framework is variation-preserving. Under this novel view, speech processing yields a representation of the speech signal much like the original signal itself: a very rich, interleaved collection of information about the underlying events that generated the acoustic signal, in which linguistic and extra-linguistic variation are both inextricably linked.

Detailed Encoding and Stimulus Variability

The acoustic signal that carries linguistic and extra-linguistic information provides a rich and very detailed source of information about the speaker, speaking environment, and the linguistic message. This proposal is nicely illustrated in Figure 1, a schematic diagram taken from Hirahara and Kato (1992). The figure describes some of the encoding processes that take place when an incoming acoustic signal is processed by the nervous system. Of particular interest to the present discussion is the top level of the figure, where the composite form of linguistic and extra-linguistic information is illustrated by two transformations of the incoming signal. On the left, particular frequencies in the signal are represented using a Bark scale. These frequencies correspond to the resonances of the vocal tract and can be grouped into three primary clusters, commonly referred to as formants. The location and absolute frequency of these formants provide the distinctive cues to talker identification, an 'extra-linguistic' feature of the signal. On the right, the same acoustic signal is transformed and represented on a Bark difference scale, showing the relationships between these formants. These relative differences are necessary and sufficient for vowel identification, which is based on the 'linguistic' features of the signal. Thus, different analyses of the same acoustic signal yield two distinct sources of information about its production. It is important to point out here that both of these analyses are based on a frequency analysis of the components of the acoustic signal. It is not that there are two different sources of information buried in the signal for each of these sets of attributes. Rather, the two properties of speech—the linguistic and indexical—are carried simultaneously and in parallel by the same acoustic signal.

By conceptualizing the speech signal as a rich source of information, we adopt an ecological approach to speech perception (Fowler, 1986; Gaver, 1993). According to this view, speech is neither under specified nor noisy; potentially, all of the variation in the acoustic signal can be utilized during the process of speech perception and spoken word recognition. Variation is assumed to be lawful and highly informative about the articulatory events that underlie the production of speech (Fowler, 1986; Pisoni, 1997).

The initial stages of speech perception within such a framework are then stages of information detection and recognition rather than reconstruction and rule application. Essentially, the detailed-encoding framework embraces the fact that any dichotomy between linguistic and extra-linguistic information in the speech signal is arbitrary. The distinction between linguistic and extra-linguistic information becomes merely a convenient way of discussing the different kinds of tasks that can be carried out on an acoustic speech signal by a listener. Moreover, the increased emphasis on processing of the variation in the speech signal intuitively explains the retention of this information in memory – without the need for re-integration of separate sources of linguistic and extra-linguistic information.

Because extra-linguistic information is not lost or filtered from the incoming signal, it is encoded in memory and available for use at later levels of processing.

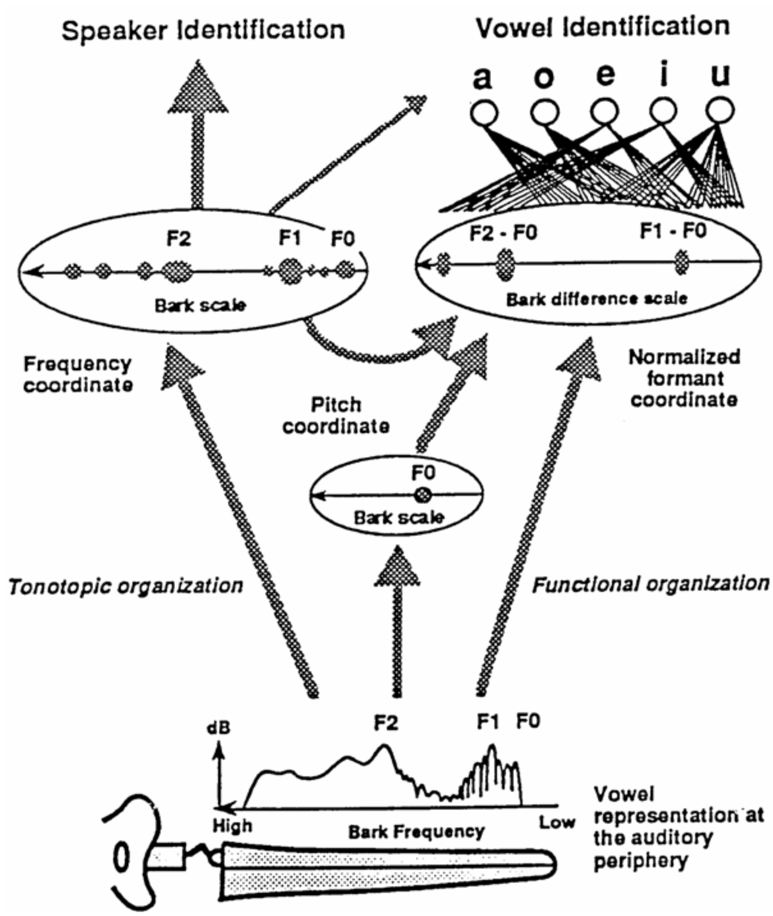

Figure 1. Information for talker and vowel identity is carried in parallel by an acoustic signal. Absolute frequencies contain information useful in talker identification, represented here along a Bark scale. Simultaneously, relative frequencies provide information useful in vowel identification (Taken from Hirahara & Kato, 1992).

This fundamental reconceptualization of the distinctive information available in the speech signal is not simply convenient or philosophically intriguing. This emerging view was necessitated by the results of a variety of novel experiments conducted over the last 10 years. In the next section, we summarize some of these findings and place them in a somewhat broader framework. We consider perceptually based phenomena in speech and describe how they affect both implicit and explicit memory processes.

Processing Dependencies: Effects of Stimulus Variation in Speech Perception

Although early work in speech perception suggested that the effects of extra-linguistic variation on the perception of spoken language were minimal (Creelman, 1957; Peters, 1955), these conclusions must be interpreted in light of the meta-theoretical notions that influenced the research agenda of the day. For example, although Creelman (1957) found that the accuracy of spoken word recognition under three different signal-to-noise ratios decreased with an increased number of talkers, he dismissed this effect as 'relatively minor'—only a difference in performance of 7%. A difference of this size probably seemed like a small effect back in the middle 1960s. In an analogue, magnetic audiotape era, large corpora of stimuli spoken by multiple talkers were difficult to create and use in behavioral experiments with human listeners. The complex presentation schemes required to uncover effects of stimulus variation were likewise virtually intractable before the advent of computer controlled experiments. Computer control and the digital audio format have provided the tools to examine and understand the nature of stimulus variation in speech processing and encoding.

Of course, much of this discussion on the composite nature of linguistic and extra-linguistic information in memory for speech would be moot if there were not evidence that the two forms of information show demonstrable effects upon each other during processing. Early studies showed that simply changing the voice of the talker from one trial to the next affected the identification of vowels (Verbrugge, Strange, Shankweiler, & Edman, 1976), consonants (Fourcin, 1968), and words (Creelman, 1957; Mullennix, Pisoni, & Martin, 1989). In addition, changes in the talker's voice also affect speed of processing. In one study, Cole, Coltheart and Allard (1974) had participants make same-different judgments on pairs of syllables. The items in each pair were spoken either by the same talker or by different talkers. Despite the fact that the task required access to and use of what would traditionally be called 'linguistic' information, Cole et al. found that reaction times were slower when different talkers spoke the two syllables in a pair than when the same talker was used. Obviously, then, variation along an extra-linguistic dimension affects the performance in even the simplest of linguistic tasks like determining if a pair of words is the same or different.

A more detailed investigation was carried out by Mullennix and Pisoni (1990) to assess the codependencies of processing linguistic and extra-linguistic information. Using a Garner speeded classification task (Garner, 1974), they constructed sets of stimuli that varied along two dimensions. One dimension, the 'word' dimension, varied the cues to phonetic categorization of the initial segment of a word (e.g., 'b' vs. 'p'). The other dimension, the 'voice' dimension, varied the cues to the identity of the talker uttering the word (e.g., 'male' and 'female'). Mullennix and Pisoni asked subjects to make several judgments about the stimuli using one dimension at a time, while manipulating the variation along the irrelevant dimension. In the 'orthogonal' conditions, the irrelevant dimension was varied randomly from trial to trial. Subjects were asked to classify stimuli as either 'b' or 'p', while the stimuli varied in terms of the gender of the talker speaking the token. For example, in this condition, sometimes the 'b' token would be spoken by the male talker, and sometimes by the female talker. In the 'correlated' conditions, the irrelevant dimension varied consistently along with the relevant dimension. In other words, a male talker might always speak the 'b' tokens, while a female talker would always speak the 'p' tokens. Finally, in the 'control' conditions, the irrelevant dimension was always held constant, while subjects made judgments about the relevant dimension (i.e., the male or the female spoke all the tokens). Response latencies were collected so that patterns of processing speed could be assessed across these different conditions.

Mullennix and Pisoni found consistent differences in reaction time that depended on the variation in the irrelevant dimensions. Response times were fastest in the correlated conditions, slower in the control conditions and slowest in the orthogonal conditions. Correlated variation along the irrelevant dimension produced a 'redundancy gain' and facilitated classification times, while orthogonal variation along the irrelevant dimension inhibited classification and slowed down response times. The pattern of speeded classification data was consistent with the proposal of mutually dependent processing of the two stimulus dimensions. In other words, the perceptual aspects of a spoken word that are associated with phonetic information and those attributes that are associated with talker information are not analyzed independently, but rather are perceived and processed in a mutually dependent fashion. Interestingly, Mullennix and Pisoni also manipulated the 'extent' of variation along each dimension in several additional experiments in which the number of response alternatives along each dimension was varied from 2 to 4, 8 or 16. While the general pattern of results was similar across all four conditions, they found that the amount of interference in the orthogonal condition increased as a function of stimulus variability. The results of Mullennix and Pisoni's study provide further evidence that increases in stimulus variation produce reliable effects on perceptual processing time and suggest that fine details of the stimulus patterns are not lost or discarded in a speeded classification task.

Thus, stimulus variability has an effect on speech processing. More importantly, the information about a talker's voice in an acoustic signal is processed in a dependent or contingent fashion along with the information specifying the linguistic content of the message. But precisely what kind of information about a talker's voice is available, and how does that information contribute to speech perception? In a measurement study of the acoustic correlates of talker intelligibility, Bradlow, Torretta and Pisoni (1996) found that while global characteristics of speech such as fundamental frequency and speaking rate had little effect on speech intelligibility, detailed changes in the acoustic-phonetic properties of a talker's voice, such as the amount of vowel space reduction and the degree of 'articulatory precision', were strong indicators of overall speech intelligibility. Their findings suggest that the indexical properties of a talker may be completely intermixed with the phonetic realization of an utterance and there may be no real dissociation between the two sources of information in the speech signal itself.

More direct evidence for the parallel encoding of linguistic and extra-linguistic information in the speech signal comes from other studies using sinewave replicas of speech. Sinewave speech is created by generating independent sinusoidal signals that trace the center frequencies of the three lowest formants in naturally produced utterances. The resulting pattern sounds perceptually unnatural, but the signal can be perceived by listeners as speech and the original linguistic message can be recovered (Remez, Rubin, Pisoni, & Carrell, 1981). Indeed, not only is the linguistic content of the utterance perceptible, but specific aspects of a talker's unique individual identity and speaking style are also preserved in sinewave replicas of speech.

Remez, Fellowes, and Rubin (1997) reported that listeners could explicitly identify specific familiar talkers from sinewave replicas of their utterances. Their findings on familiar talker recognition are remarkable because sinewave speech patterns preserve none of the traditional 'speech cues' that were thought to support the perception of vocal identity, such as fundamental frequency, or the average long-term spectrum. In creating sinewave speech patterns, an utterance is essentially stripped of all of the redundant acoustic information in an utterance except the time-varying properties of the vocal resonances generated by articulatory motion. While these skeletonized versions of speech have been shown to be sufficient for accurate identification of the linguistic content of a message, the new findings by Remez and colleagues demonstrates that sinewave speech patterns are also sufficient for the accurate identification of extra-linguistic information about familiar voices as well. These time-varying sinewave speech patterns preserve individual, talker-specific cues needed for voice recognition.

Thus, even in its most basic forms, linguistic and extra-linguistic sources of information appear to be inextricably bound to one another. Because sinewave speech patterns preserve little of the original signal other than the acoustic variation corresponding to the kinematics of articulatory motion, we suggest that the link between linguistic and extra-linguistic information derives from the common underlying articulatory events and movements of the speech articulators that produce speech. As we have argued, these links produce consistent effects on speech perception. But do the links between linguistic and extra-linguistic sources of information affect the memory processes that are so crucial to spoken word recognition and lexical access? We suggest they do in the next section.

Detailed Encoding Effects in Implicit and Explicit Memory

The integration of linguistic and extra-linguistic attributes in the speech signal and the mutually dependent perceptual processes that encode and process these cues has several important implications for the representation of speech in memory. According to the detailed encoding perspective, the mental representation of speech preserves the same sorts of information found in the original speech event (Goldinger, 1998). Rather than a static word-store of idealized, abstract, formalized units, Goldinger (1998) has proposed that the mental lexicon should be viewed as an extremely detailed set of instance-specific episodes. Extra-linguistic and linguistic information are preserved in the lexicon just as they are encoded in the auditory signal - in an integrated, holistic composite of linguistic and extra-linguistic properties. Evidence supporting this 'episodic' view of the lexicon comes from a series of recent memory experiments that show effects of extra-linguistic variation, even when the specific task only requires access to and use of linguistic information alone. The specific memory demands of the task—whether the task measures or assesses explicit or implicit memory—should not matter. If the basic representation of speech events in memory is highly detailed and episodic in nature, then any behavior that requires access to these memory representations should show contingent effects of these detailed composite representations.

While written words have been the primary focus in implicit memory research (Bowers, 2000), *spoken* words have received much less attention in the implicit memory literature. In the next section, we review some of the recent work that has been done on implicit effects of extra-linguistic variation in speech. We take as our starting point the operational definition summarized by Schacter (1987) that 'implicit memory is revealed when previous experience facilitates performance on a task which does not require conscious or intentional recollection of those experiences.' (p.501). The results of experiments using different memory paradigms are important in establishing the generality of these findings. Thus, we summarize experiments that examine the role of variability in both implicit and explicit memory for speech events.

Effects of Stimulus Variation on Implicit Memory

In a perceptual identification experiment conducted by Goldinger (1992), several groups of subjects were first asked to repeat words spoken to them in the quiet over headphones. The original set of stimuli was spoken by pools of 2, 6 and 10 talkers. Subjects then returned 5 minutes, 1 day or 1 week following the initial exposure and again identified spoken words in the quiet. Goldinger found that subjects were faster and more accurate in repeating words spoken by old talkers who were used at the time of the initial presentation than new talkers. Figure 2 shows the difference between test phase and study phase accuracy for words in the three talker pools across the three delay periods. Overall, Goldinger's data show evidence for a 'repetition effect'. That is, the identification of words was more accurate when those words were repeated in the same voice that spoke them at the time of study than in a novel voice. In addition, the advantage conferred by a repeated voice did not significantly decline as the delay between training and testing increased from 5 minutes to 1 day to 1 week. Goldinger's findings demonstrate that long-term memory representations of speech events not only include extra-linguistic information, but also preserve these instance-specific details for long periods of time. For talker similarity to have any effects on repetition accuracy, a record or memory trace of the extra-linguistic attributes of the talkers' speech had to persist in memory along with the more abstract, symbolic linguistic information encoded in the signal.

Goldinger also found that the differences in repetition accuracy for old and new voices were related to the perceptual similarity of the talker's voices. Words produced by talkers who had perceptually distinctive voices resulted in larger repetition effects for repeated talkers than words produced by talkers who had less distinctive voices. These latter lists showed smaller, but still significant, effects for repeated talkers. The 'graded' effects that similarity had on the repetition effect was interpreted by Goldinger as evidence for an episodic view of memory for spoken items. Even when subjects were not instructed to attend to the voices of the stimuli, their memories for these speech events were detailed enough for the relative similarity among the talkers' voices to produce differential effects on performance in this task. Such findings are inherently compatible with exemplar models of categorization, in which similarity is represented continuously as distance in a perceptual space (Nosofsky, 1986; Nosofsky, 1987). Because points in the perceptual space represent individual tokens and not prototypical, idealized, abstract categories, the exemplar view of the lexicon provides a representational basis for making predictions sensitive to the graded similarity between voices.

In another study, Schacter and Church (1992) found consistent effects of voice information on implicit memory for words. In their experiments, subjects completed a study phase in which they made simple judgments about the enunciation or intonation of lists of words spoken by multiple talkers. Subjects then completed several implicit and explicit memory tasks in a test phase. Test stimuli were composed of tokens from the original study phase and 'new' tokens derived from study stimuli by changing the voice, intonation, or fundamental frequency of old items used in the study phase. In both an auditory identification task and a stem completion task, Schacter and Church found that study-to-test changes in all three of these stimulus attributes yielded significant reductions in subjects' accuracy.

The impairment in performance observed in the implicit memory tasks from this experiment is particularly interesting because performance on explicit recall and recognition tasks showed little, if any, effects of study-to-test changes. Similar effects had previously been observed by Church and Schacter (1994) for word identification when stimuli were presented in white noise and for stem completion tasks when stimuli were presented in the clear. As with the Goldinger experiment reported above, the findings of Schacter and Church show that extra-linguistic variation in speech is encoded and retained in memory for speech events and is an important enough component of this representation to produce reliable implicit effects on the recognition of spoken words even when such tasks do not mention these extra-linguistic dimensions at the time of initial encoding or even call attention to these attributes of the stimulus materials.

The long-term storage of extra-linguistic information about a talker's voice in implicit memory has also been demonstrated in a series of studies that examined the learning of unfamiliar voices (Nygaard & Pisoni, 1998; Nygaard, Sommers, & Pisoni, 1995). In one experiment, Nygaard and Pisoni (1998) trained participants to identify a set of novel talkers from their voices alone. Once the participants had learned the names of the voices using a set of training stimuli, Nygaard and Pisoni found that the knowledge of talker characteristics obtained under these conditions also generalized to new stimuli that were never used in training. More importantly, Nygaard and Pisoni found that the perceptual learning of the trained voices transferred to a novel task: words spoken by familiar voices were recognized in noise more accurately than words spoken by unfamiliar voices. Thus, performance on the transfer task was facilitated by prior experience with the voices of the talkers with whom the participants were trained. Because there was no explicit reference to previous episodes or to recognizing words during training, Nygaard and Pisoni's findings provide evidence for the implicit encoding and use of information about a talker's voice in an entirely different task - recognizing spoken words.

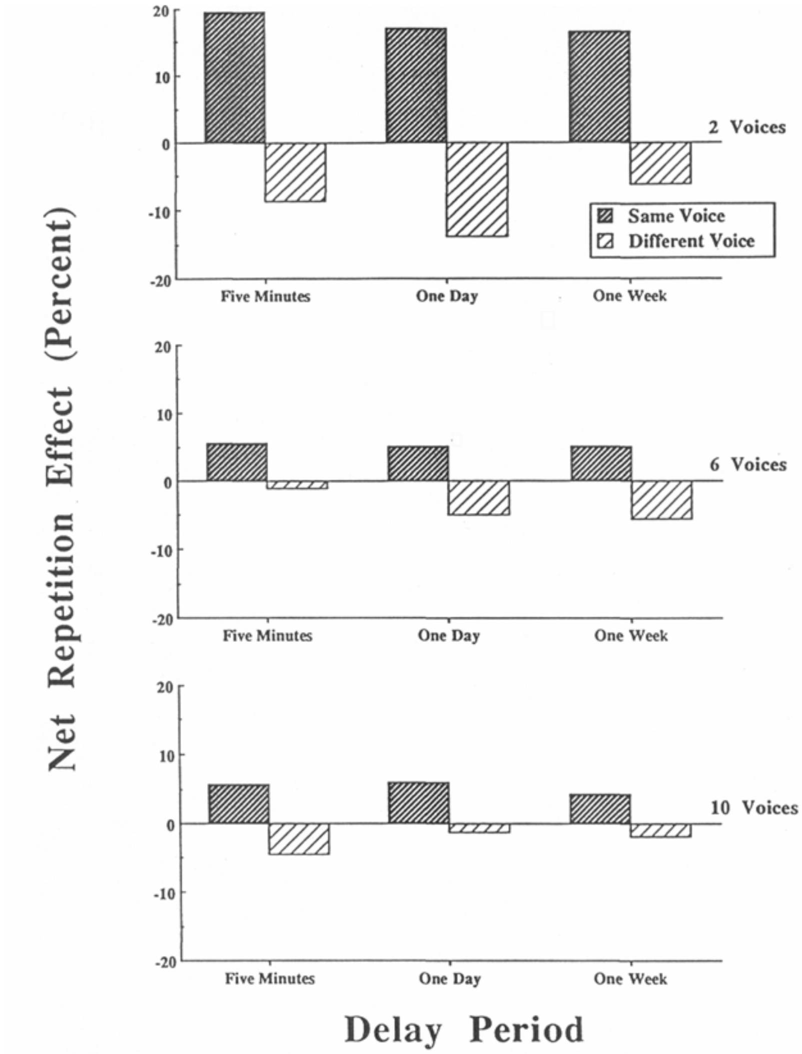

Figure 2. Net repetition effects observed in perceptual identification as a function of delay between sessions and repetition voice. The y-axis shows the difference in word identification accuracy for the original and subsequent presentation of the word. There was a benefit to word identification if the word was repeated in the same voice as it was presented with originally. This effect did not dissipate over time. Increasing the number of voices in the experiment decreased the effect somewhat, due to a decrease in the perceptual distinctiveness of the voices used (from Goldinger, 1992).

Further evidence for implicit encoding and storage of extra-linguistic variation in speech perception comes from a study on the learning of English [r] and [l] by native speakers of Japanese. In a series of perceptual learning experiments in which Japanese listeners were trained to recognize the English /l/ and /r/ distinction, Logan, Lively, and Pisoni (1991) showed implicit effects for variation in the original training sets. In a follow-up study, Lively, Pisoni, Yamada, Tohkura and Yamada (1992) found that the English /l/ and /r/ contrast was better retained by Japanese listeners when they were exposed to a large corpus of stimuli spoken by many different talkers during training. Compared to a group of listeners who had been trained using stimuli spoken by a single talker, listeners who had been trained using tokens produced by multiple talkers were better able to distinguish /l/ and /r/ in the speech of entirely new speakers. Although not originally designed to study implicit memory effects, these perceptual learning results satisfy the standard definition of implicit memory since it is unlikely that subjects explicitly recalled their earlier training experience when required to identify novel speech samples. Moreover, the subjects were never explicitly told to attend to the different voices used in the training phases of the experiment. All they were required to do was categorize each word they heard as having an /r/ or /l/ in it.

Effects of Stimulus Variation on Explicit Memory

Although research findings on implicit memory for speech are limited, a variety of other experimental paradigms have uncovered effects that parallel the results of these implicit memory experiments. These experiments do not measure implicit memory in the standard sense laid out above. Although most of these experiments did require subjects to consciously recall their previous experiences, the results from these experiments are important because they demonstrate the same kinds of effects of encoding detail in speech memory that were revealed in implicit memory research.

In one study, Goldinger, Pisoni, and Logan (1991) examined the effects of talker variability on explicit memory for spoken words using a serial recall task. They manipulated the number of talkers used to create the stimulus lists and the rate at which items within a list were presented. They measured recall accuracy for items at the various serial positions in the list. The results showed that presentation rate interacted with the number of talkers used in the stimulus lists. At the fastest presentation rates, talker variability caused a decrease in accuracy across all serial positions in the list. Recall of single talker lists was better than recall of multiple talker lists. But, as the presentation rate decreased and the items were presented more slowly, however, the original pattern reversed. At the slowest presentation rates, subjects were more accurate at recalling lists of words spoken by multiple talkers than by a single talker, especially in the earlier portions of the list. Goldinger et al. concluded that information about voices must be incidentally encoded in memory at the time of presentation. At faster rates, this incidental encoding of voice features interferes with the perceptual encoding of items, leading to lower recall performance. At slower presentation rates, however, multiple talker lists contain additional distinctive cues that can be used to retrieve items from memory, thus yielding higher recall scores at test.

In another experiment that examined the encoding of extra-linguistic information in memory, Palmeri, Goldinger, and Pisoni (1993) used a continuous recognition memory procedure in which subjects listened to long lists of words spoken by multiple talkers. In this recognition task (see Shepard & Teghtsoonian, 1961), participants are required to judge each stimulus item in a list as ‘old’, if they have previously experienced the stimulus in the list, or ‘new’, if they have not. By varying the lag between the initial stimulus and its subsequent presentation, the effects of time and decay can be measured. In their experiment, Palmeri, et al. added a variant to the standard recognition memory paradigm by repeating old words in either the same voice or in a different voice from the initial presentation.

Palmeri et al.’s results were consistent with the findings we have reported thus far. Subjects showed the highest recognition accuracy for words that were presented in a repeated voice. Interestingly, subjects also showed the worst performance when talkers of a different gender repeated the words, indicating that highly dissimilar voices (as in cross-gender talker changes) were unable to function as reliable ‘cues’ to recognition of the words. The effects of lag between study and test in this experiment were surprising. As expected, recognition accuracy decreased overall with increasing lags between initial and subsequent presentations of the stimulus. However, the advantage for ‘same voice’ repetition did not interact with the lag between initial and subsequent presentations of an item. In other words, the encoding of extra-linguistic information facilitated the recognition of words regardless of the time between the initial encoding of the word and test. This pattern of results indicates that extra-linguistic information is preserved in memory to the same extent that linguistic information is preserved. Although the memory trace for a word may decay over time, many of the fine details of the memory representation are not lost over time and can be used to facilitate subsequent recognition.

In another study, Lightfoot (1989) reported that subjects who had previously been trained to identify a set of talkers using common names showed better cued recall scores for lists of words when the words were spoken by multiple talkers compared to single talkers. Unlike the interaction observed by Goldinger et al. (1991), however, Lightfoot found that multiple talkers helped recall even at relatively fast presentation rates. Because the listeners in Lightfoot's experiment had been explicitly trained for several days to learn the voices of the talkers to a criterion beforehand, they were more familiar with these voices than participants in Goldinger's experiment. Both experiments provide support for the same conclusion. Detailed information about the talker's voice is encoded in memory along with the more abstract, symbolic linguistic content of the signal, and these instance-specific attributes facilitate the later recall and recognition of spoken words.

Explicit memory research using sentences has also revealed effects that suggest that detailed encoding of linguistic and extra-linguistic information is retained in memory. Geiselman and Bellezza (1977) presented one group of subjects with a set of sentences spoken by a male and a female talker. They also presented a control group with a set of sentences spoken only by the male talker or only by the female talker. Subjects were instructed either to attend only to the content of the sentences ('incidental gender encoding' condition) or to remember *both* the content and gender of the sentences ('explicit gender encoding' condition) for a subsequent memory test. Geiselman and Bellezza found that recall of the sentences was not significantly different for the experimental and control groups even though experimental subjects remembered the gender of sentences at higher than chance levels under both incidental and explicit gender encoding instructions.

Geiselman and Bellezza considered two possible explanations for how gender information could be encoded without detrimental effects on encoding of the linguistic content of the sentences. According to their 'voice-connotation' hypothesis, the meaning of a sentence may be encoded such that information about the speaker's voice is automatically encoded without increasing demands on processing resources. In contrast, their 'dual-hemisphere parallel-processing hypothesis' explained the encoding of gender information without increased processing costs by positing that both content and gender information are encoded in parallel by the left and right hemispheres of the brain. In subsequent research, Geiselman and Bellezza replicated their initial results (unpublished experiment mentioned in Geiselman & Bellezza, 1977) and found support for the voice-connotation hypothesis: the 'voice attribute is not 'attached' to the code for the item in memory. Rather, it may become an integral part of the code itself…' (Geiselman & Bellezza, 1977). These findings are important because they show that the composite encoding in memory of linguistic and extra-linguistic information is not constrained to isolated spoken words, but generalizes to larger linguistic units, like sentences.

Recently, McMichael and Pisoni (2000) obtained additional evidence for implicit encoding and retention of voice information in sentence-length stimuli. In a series of four discrete recognition memory experiments, they presented listeners with a list of 40 sentences in a 'study' phase. In the 'Intentional encoding' conditions, subjects were specifically told that their memory for sentences would be tested following the study phase. In the 'Incidental encoding' conditions, subjects received a surprise recognition memory test. During the study phase, 5 male and 5 female talkers spoke the list of sentences. In the test phase, listeners were asked to judge a list of 80 sentences as 'old' (i.e., heard at the time of study) or 'new' (i.e., not heard at the time of study). The forty 'old' sentences were spoken by either the same talker used during study ('Repeated Voice') or an entirely new talker that had not been heard during the study phase ('Non-repeated Voice'). The forty 'new' sentences were also spoken by either a talker that had been heard during the study phase or by an entirely new talker.

McMichael and Pisoni also manipulated the encoding task during the study phase, in order to determine whether instructions focusing attention on voice attributes would affect recognition memory performance. In one task, subjects simply hit the 'enter' key on a keyboard after hearing each sentence. In the other task, subjects indicated the gender of the speaker by typing in 'm' or 'f' after each sentence was played. Both study tasks were run under either 'incidental encoding' instructions or 'intentional encoding' instructions, producing four combinations of instructions and study task.

Figure 3 shows the recognition memory results from these experiments. Each set of four columns within a panel represents the probability of a correct response for the four different types of sentences at the time of test. The pattern of results shows consistent repetition effects based on the voice of the talker. Sentences were recognized more accurately as 'old' when they were presented at test in the same voice that was used at study ('Old/Repeated') than when they were presented in a different voice ('Old/Non-repeated'). This voice repetition effect was statistically significant across all four experiments, showing that even for sentences, voice information is encoded and stored in memory along with linguistic information.

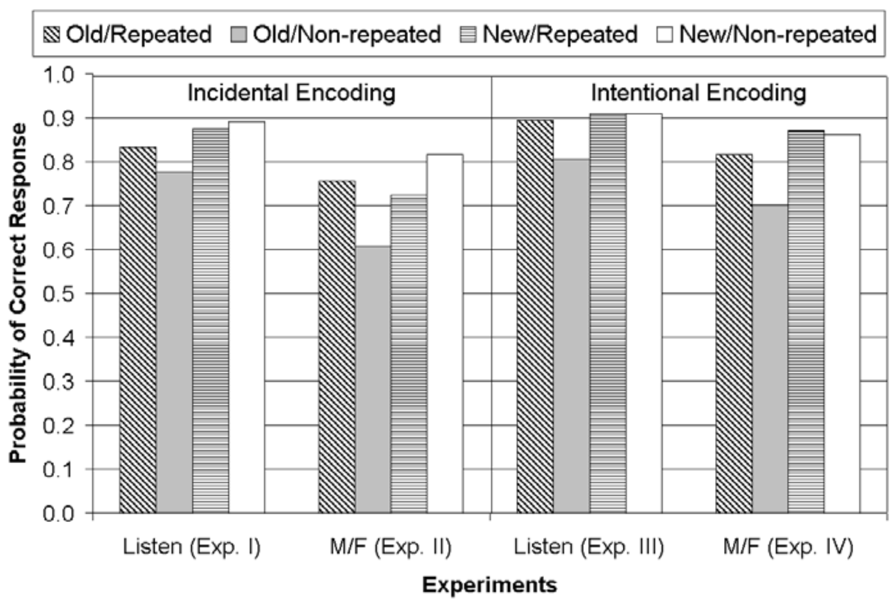

Figure 3. Probability of correct response from four recognition memory experiments using sentences. Each group of columns represents the results from a different experiment. In the 'Intentional Encoding' experiments, participants were instructed that there would be a test of recognition memory following the 'study' task. In 'Incidental Encoding' experiments, participants received a surprise recognition test after study. The four columns associated with each experiment show the various conditions under which the sentences were presented. 'Old' sentences are items that were presented during the study phase; 'New' sentences were not presented during the study phase. 'Repeated' voices were voices used during the study phase; 'Non-repeated' voices were novel voices that were not presented during study. The results show that across all four experiments, old sentences were more accurately identified as old when they were repeated in the same voice as original presentation than when they were presented in a new voice (adapted from McMichael & Pisoni, 2000).

Surprisingly, new sentences—those that were never presented during study—also showed a voice repetition effect in one of the four experiments. Under the "incidental encoding" instructions, when the study task involved identifying the gender of the speaker in Experiment II, new sentences spoken in non-repeated voices ('New/Non-repeated') were more accurately identified as new than new sentences spoken by repeated voices ('New/Repeated'). These recognition memory results suggest that listeners did encode information about the specific attributes of the talkers used at the time of study—even when the instructions used at the time of study never mentioned there would be a subsequent recognition memory test for the sentences.

These results indicate that specific details regarding the voice of the talker were encoded at the time of study and this information was later retrieved and utilized in the test phase recognition task. Additionally, the fact that these effects for voice information were observed for both sets of instructions - one that explicitly focused attention on the talker's voice at study, and one that did not - demonstrates that these recognition memory effects were not simply a result of instructions that encouraged voice information encoding. Even without the explicit intention of the listeners to encode voice information, details of the talker's voice were encoded that were sufficient to facilitate performance in the recognition memory task. Thus, the earlier results obtained with isolated words clearly generalized to sentence length materials as well.

The effects of stimulus variation uncovered in these experiments, although obtained under what would traditionally be called explicit memory paradigms, are notable because subjects were never explicitly instructed to pay attention to the talker's voice during study or test. Although subjects were required to explicitly recall or recognize the words or sentences they heard, they might not have been consciously aware of incidental variation in voice information while performing these tasks. To the extent that specific details of the original speech events affected performance in these explicit memory tasks, the results have a clear and direct relationship with the implicit memory effects summarized above.

The range of stimuli investigated in these experiments runs from isolated phonemes to words to sentence length speech samples. The fact that comparable effects of extra-linguistic variation are observed across these different stimulus materials suggests that similar perceptual and memory processes may be involved in the encoding and storage of phonemes, words and sentences. The similarity of the effects of talker variability in these experiments suggests a close link between implicit and explicit memory and raises important theoretical issues about how to explain and account for the pattern of these findings in a unified and coherent fashion.

Some Final Thoughts About Implicit Memory and Detailed Encoding

Although the history of research on speech perception has been dominated by abstractionist, information-processing approaches that consider extra-linguistic information irrelevant to the primary task of uncovering the idealized linguistic signal from beneath a wide range of noisy transformations, an emerging line of research suggests that extra-linguistic information and variation actually plays an important role in the process of speech perception. Several results of this line of research are particularly important to emphasize here:

Scale Invariance. Whether the experiment used phonemes, words or sentences, similar effects of variation in linguistic and extra-linguistic information have been obtained with units of differing lengths. The similarity of these findings suggests that all levels of speech representation may rely on a common substrate that incorporates both linguistic and extra-linguistic information. If the mental lexicon is conceived of as an abstract word-store that encodes only word 'types' and not word 'tokens', then current accounts of lexical memory will have great difficulty in explaining why units of differing length show effects of extra-linguistic variation. Effects for stimuli that are shorter and longer than words are difficult to explain since it is unclear how an abstracted store of word information could generate episodic effects for phoneme or sentence length stimuli if the fine instance-specific details of speech events are lost or discarded from memory at the time of initial encoding via the process of normalization.

Rather, a conceptualization of the mental lexicon as an integrated, functionally-identical subsection of a larger, more general memory system, in which all experience with speech is encoded and preserved in a detail-rich, complex and multidimensional representation, seems more appropriate as a way to account for these results.

Parallel Transmission. In contrast to the traditional view of speech, in which linguistic and extra-linguistic sources of information were viewed as separate components of the speech signal (Abercrombie, 1967), the research summarized in this chapter suggests that these two sets of attributes may be inseparable. In both speech perception and memory tasks, subjects are consistently affected by variation in both sources of information even when they are not explicitly instructed to attend to one set of attributes or the other.

It is important to keep in mind that the dissociation between linguistic and extra-linguistic information in speech is arbitrary and has been handed down to speech scientists from pre-existing meta-theoretical notions inherited from the study of linguistics, where human performance had been explicitly ruled as irrelevant to the study of language by Chomsky's competence/performance distinction (Chomsky, 1965).

The finding that human listeners encode and retrieve both linguistic and extra-linguistic information is not surprising—after all, how else could we learn to recognize and identify the voices of our friends, or the slight nuances of affect that allow us to negotiate the complex rules of pragmatic discourse, unless we encode and retain very detailed extra-linguistic information in memory. It is precisely the inseparable relationship between linguistic and extra-linguistic information that is important - that is, variation in linguistic and extra-linguistic information is not simply a helpful source of information when listeners happen to have access to it. It is rather an integral part of understanding and remembering the meaning and intent of speech events. Variation in speech is so important that even when listeners are explicitly instructed to ignore differences in linguistic or extra-linguistic information, their performance in speech perception and memory tasks appears to be influenced by all aspects of the original signal, including attributes not relevant to the specific task at hand. The processing of extra-linguistic detail without conscious awareness sounds much like the obligatory or mandatory processing needed for module status under Fodor's modularity hypothesis (Fodor, 1981). However, we do not wish to imply that speech processing is undertaken by a specialized module. Rather, we think it more reasonable to claim that the conjoint processing of linguistic and extra-linguistic attributes follows naturally from their simultaneous and inextricable production by the vocal articulators.

Parallels in Explicit and Implicit Memory. In both implicit and explicit memory paradigms, similar effects of stimulus variation in the speech signal have been obtained across a variety of tasks. Utterances spoken in the same voice as earlier presentations increase accuracy in explicit memory for words and reduce response latencies in implicit tasks such as word identification.

That the same variation in the original speech signal can have parallel effects in both explicit and implicit tasks suggests that these two memory systems rely on the same types of representations for speech events. These representations are not based on abstract, idealized, contrastive linguistic units, but rather carry with them detailed episodic, instance-specific information about the circumstances of vocal articulation that produce speech. Furthermore, the similarity of these effects suggests that the traditional separation of these two types of memory may not be a valid conceptualization of memory for spoken language.

A detailed encoding, or ‘exemplar,’ perspective provides an alternative view that can account for findings. With regard to scale invariance, the specific, rich detail with which speech events are retained in memory preserves information about the dialect, gender or other indexical properties of the talker at any scale, whether the units are phonemes, words, sentences, or even units like discourse segments. Just as speech is perceived and produced in a consistent manner across scales from words to extended discourse, the memory representation for speech may be similar across different sized chunks of speech. Whether these units are phonemes or sentences, the detail of these speech events in memory would allow for the observed effects of stimulus variability.

The approach advocated here is consistent with a composite form of mental representations for speech. Since speech is produced and perceived as a unitary event that carries both linguistic and extra-linguistic information, the composite representation of a detailed memory representation falls out naturally from the physics of speech motor control and behavior. The intended message of a speaker preserves a form of parity with the production of that message via vocal articulation. Rather than being a source of noise, however, the complex interaction of the speech articulators lawfully varies the speech signal in ways that are informative and distinctive to the listener. We need not posit that different ‘entries’ for information about linguistic content, gender, dialect and affect are stored in a complex associative memory system. Rather, the fact that this information arrives encoded and packaged in a unitary speech signal provides a *de facto* explanation of its storage together in speech memory. Since our memories for speech events are integrated, unitary composites of both linguistic and extra-linguistic information, behavioral tasks that assess these memories may also be affected by the rich, redundant information stored therein. Just as the detailed encoding perspective questions the validity of a distinction between linguistic and extra-linguistic information, so too does this perspective challenge the distinction between implicit and explicit memory.

For the purposes of speech research, implicit and explicit memory have largely been distinguished based on the kinds of speech information relevant to each memory system. Extra-linguistic information such as speaking rate or gender has been the traditional focus of implicit memory for speech experiments (Church & Schacter, 1994; Goldinger, 1992; Schacter & Church, 1992) while explicit memory research has focused on the more abstract, idealized linguistic information such as phonemes or words (Liberman, 1957; Peters, 1955). The result of this divergence of research has been the tacit assumption that implicit and explicit memory for speech events reflect the operation of functionally distinct, separate memory systems that deal with different types of speech information.

In a detailed encoding perspective, however, there is no valid distinction between linguistic and extra-linguistic information. Without this information-based distinction, the difference between implicit and explicit memory for speech events begins to blur. If the same memory representation underlies behavior in both implicit and explicit memory tasks and if behavior in these tasks shows similar effects of variation in the speech signal, then we can rightly question whether these two memory systems are, in fact, separate and distinct.

To apprehend the meaning of a given speech event and to recover the talker’s intended message, it is necessary for the listener to know who is speaking, what they said, and under what conditions the articulatory events that produced an utterance occurred. The traditional perspective on speech perception, as well as the accepted distinction between implicit and explicit memory, assumes that the information in these representations is processed, stored, and accessed separately. The episodic view of speech perception, which is intimately tied to a description of the underlying events and their consequences, takes the integration of this information as an important constraint on the way speech events are processed and stored in memory. This approach incorporates both implicit and explicit memory phenomena as reflecting aspects of a single complex memory system that retains highly detailed, instance-specific information in a perceptual record containing all of our experiences—speech and otherwise.

References

Abercrombie, D. (1967). *Elements of general phonetics*. Chicago, IL: Aldine Publishing Company.

Bowers, J.S. (2000). In defense of abstractionist theories repetition priming and word identification. *Psychonomic Bulletin & Review, 7*, 83 - 99.

Bradlow, A.R., Torretta, G.M., & Pisoni, D.B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication, 20*, 255 - 273.

Chomsky, N. (1965). *Aspects of a theory of syntax*. Cambridge, MA: MIT Press.

Church, B.A., & Schacter, D.L. (1994). Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 521 - 533.

Cole, R.A., Coltheart, M., & Allard, F. (1974). Memory of a speaker's voice: Reaction time to same- or different-voiced letters. *Quarterly Journal of Experimental Psychology, 26*, 1 - 7.

Creelman, C.D. (1957). Case of the unknown talker. *Journal of the Acoustical Society of America, 29*, 655.

Fodor, J.A. (1981). The mind-body problem. *Scientific American, 224*.

Fourcin, A.J. (1968). Speech-source interference. *IEEE Transactions in Audio Electroacoustics, ACC-16*, 65 - 67.

Fowler, C.A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics, 14*, 3 - 28.

Garner, W.R. (1974). *The processing of information and structure*. Potomac, MD: Erlbaum.

Gaver, W.W. (1993). What in the world do we hear?: An ecological approach to auditory event perception. *Ecological Psychology, 5*(1), 1 - 29.

Geiselman, R.E., & Bellezza, F.S. (1977). Incidental retention of speaker's voice. *Memory and Cognition, 5*, 658 - 665.

Gibson, J.J. (1966). *The Senses Considered as Perceptual Systems*. Boston, MA: Houghton Mifflin.

Goldinger, S.D. (1992). Words and voices: Implicit and explicit memory for spoken words, *Research on Speech Perception Technical Report No. 7* . Bloomington, IN: Indiana University.

Goldinger, S.D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*, 251 - 279.

Goldinger, S.D., Pisoni, D.B., & Logan, J.S. (1991). On the nature of talker variability effects on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 17*, 152-162.

Halle, M. (1956). *For Roman Jakobson: essays on the occasion of his sixtieth birthday, 11, Oct 1956*. The Hague: Mouton.

Hirahara, T., & Kato, H. (1992). The effects of F0 on vowel identification. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech perception, production, and linguistic structure* (pp. 89 - 112). Tokyo: Ohmsha Publishing.

Klatt, D.H. (1989). Review of selected models of speech perception. In W.D. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 201 - 262). Cambridge, MA: MIT Press.

Liberman, A.M. (1957). Some results of research on speech perception. *Journal of the Acoustical Society of America, 29*, 117 - 123.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*, 431 - 461.

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

Lightfoot, N. (1989). Effects of talker familiarity on serial recall of spoken word lists, *Research on Speech Perception, Progress Report No. 15*. Bloomington, IN: Indiana University.

Lively, S.E., Pisoni, D.B., Yamada, R.A., Tohkura, Y., & Yamada, T. (1992). Training Japanese listeners to identify English [r] and [l]: III. Long-term retention of the new phonetic categories, *Research on Speech Perception Progress Report No. 18* (pp. 185-216). Bloomington, IN: Indiana University.

Logan, J.S., Lively, S.E., & Pisoni, D.B. (1991). Training Japanese listeners to identify the English [r] and [l]: A first report. *Journal of the Acoustical Society of America, 89*, 874 - 886.

McMichael, K., & Pisoni, D.B. (2000). Talker-specific encoding effects on recognition memory for spoken sentences. Manuscript in preparation.

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. (1967). *Cognitive Psychology*. New York: Appleton-Century-Crofts.

Nosofsky, R.M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology : General, 115*, 39 - 57.

Nosofsky, R.M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 87-108.

Nygaard, L.C., & Pisoni, D.B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics, 60*, 355 - 376.

Nygaard, L.C., Sommers, M.S., & Pisoni, D.B. (1995). Effects of stimulus variability on perception and representation of spoken words in memory. *Perception & Psychophysics, 57*, 989 - 1001.

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*, 309 - 328.

Peters, R.W. (1955). *The relative intelligibility of single-voice and multiple-voice messages under various conditions of noise* (Joint Project Report No. 56). Pensacola, FL: U.S. Navel School of Aviation Medicine.

Pisoni, D.B. (1997). Some thoughts on "Normalization" in speech perception. In K. Johnson & J.W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 9-32). San Diego: Academic Press.

Remez, R.E., Fellowes, J.M., & Rubin, P.E. (1997). Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 651 - 666.

Remez, R.E., Rubin, P.E., Pisoni, D.B., & Carrell, T.D. (1981). Speech perception without traditional speech cues. *Science, 212*, 947 - 950.

Schacter, D.L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 501 - 518.

Schacter, D.L., & Church, B.A. (1992). Auditory priming: Implicit and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 915-930.

Shepard, R. N., & Teghtsoonian, M. (1961). Retention of information under conditions approaching a steady state. *Journal of Experimental Psychology, 62*, 302 - 309.

Studdert-Kennedy, M. (1974). The perception of speech. In T.A. Sebeok (Ed.), *Current trends in linguistics* (Vol. XII, pp. 2349 - 2385). The Hague: Mouton.

Studdert-Kennedy, M. (1976). Speech perception. In N.J. Lass (Ed.), *Contemporary issues in experimental phonetics* (pp. 243 - 293). New York: Academic Press.

Verbrugge, R.R., Strange, W., Shankweiler, D.P., & Edman, T.R. (1976). What information enables a listener to map a talker's vowel space? *Journal of the Acoustical Society of America, 60*, 198-212.

This page left blank intentionally.