RESEARCH ON SPEECH PERCEPTION

Technical Report No. 7

May 6, 1992

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

Supported by:
Department of Health and Human Services
U.S. Public Health Service
National Institutes of Health
Research Grant No. DC-00111-16
and Training Grant No. DC-00012-13

WORDS AND VOICES: IMPLICIT AND EXPLICIT MEMORY FOR SPOKEN WORDS

Stephen D. Goldinger

Submitted to the faculty of the Graduate School
in partial fulfillment of the requirements
of the degree
Doctor of Philosophy
in the Department of Psychology
Indiana University
May, 1992

Acknowledgments

At the time of this writing, I can think of two distinctly pleasurable aspects of completing a dissertation. The first, of course, is getting it over with. The second is taking advantage of the conventional forum for thanking important people for all their support. I take particular pleasure in this latter aspect; I owe much to many people. Obviously, the acknowledgements conferred on this page do not begin to dispense of my personal or collegial debts, but it's all most of you are likely to get.

On the strictly personal side, my parents, Leonard and Carolyn Goldinger, deserve more appreciation than I can adequately express. Their continual love, support, and tolerance have pulled me through my studies, along with everything else to date. Tamiko Azuma has been patient and understanding (especially patient!) throughout the course of this project. Denise Beike has been my close friend, and more recently a colleague, since before my studies began.

On the personal and professional side, my colleagues at Indiana have earned my respect, friendship, and gratitude. David Pisoni has provided an extraordinary blend of guidance and opportunity to ensure my safe passage through graduate school. Paul Luce has been my favorite and most insightful collaborator. Lisa Burgin worked tirelessly to collect most of the data in this report; I could not have completed it without her help. And my fellow epsilons: Scott Lively, John Logan, Keith Johnson, Lynne Nygaard, Tom Palmeri, Luis Hernandez, and Mitch Sommers. During my time here, you all have constituted the spirit and intelligence of the lab. Scott has been the perfect blend of friend, colleague, and basketball fan. I would not have enjoyed or benefited from my experience here without the environment you all help to create.

This research was supported by NIH Research Grant DC-00111-16 and by NIDCD Training Grant DC-00012-13 to Indiana University, Bloomington, Indiana.

Abstract

Contemporary theories of lexical representation and processing derive from an information-processing approach to psychology. By this view, abstract, canonical representations are stored in long-term memory, and specialized processes match new stimuli to these representations. In speech perception and spoken word recognition, such processes are assumed to resolve perceptual "problems," such as contextual variability and speaker variability. After idiosyncratic perceptual details are normalized, the input is compared to idealized, symbolic representations in the lexicon. The present investigation assessed implicit and explicit memory for the fine perceptual details of spoken words. In particular, the experiments assessed memory for spoken words and attributes of speakers' voices. In tests of implicit and explicit memory across variable delays and across levels of processing, it was found that specific perceptual details of spoken words were encoded into memory and were accessible long after recognition was complete. In addition, the degrees of perceptual similarity between repeated words and their original presentations were strongly related to the magnitudes of repetition effects. In all conditions, implicit measures displayed greater sensitivity to surface details than explicit measures: Effects of voice repetition were observed across longer delays and deeper levels of processing in implicit tests than in explicit tests. Fluent perception of word/voice repetitions suggested that detailed, episodic traces subserve later word recognition. Taken together, the results support a view of a multiple-trace mental lexicon. Every word in the mental lexicon may be represented by multiple, partially redundant, episodic traces. These traces retain fine perceptual details that are often considered "noise" in a recognition system; spoken word recognition apparently occurs against this background of perceptual episodes.

Table of Contents

Words and Voices: Implicit and Explicit Memory for Spoken Words

Chapter I: Introduction

> *Experience is never limited, and it is never complete; it is an immense sensibility, a kind of huge spider-web of the finest silken threads suspended in the chamber of consciousness, and catching every airborne particle in its tissue.*
>
> Henry James, *Partial Portraits*

In the first decade of the twentieth century, a German scientist named Richard Semon published a book describing a theory of memory (Semon, 1909/1923). Semon's book anticipated many of the concerns and theoretical innovations of contemporary theories of memory. His work received accolades from important scholars of his day, such as Kurt Koffka and Bertrand Russell. Nevertheless, his work was apparently forgotten for 70 years. Only recently Schacter, Eich, and Tulving (1978; see also Hintzman, 1986) re-discovered Semon's theory and introduced it back into the mainstream literature. Semon's book was translated into English in 1923. Had it been translated more recently, with more contemporary usage, most cognitive psychologists would never realize the ideas were nearly 90 years old.

In modern parlance, we would call Semon's theory an *exemplar* or *multiple-trace* theory of memory. Semon's theory assumes that every sensory experience leaves a unique trace, or *engram,* in memory. Upon presentation of a new stimulus, all traces are accessed in parallel; each trace is activated according to its similarity to the stimulus. The most activated trace or traces come to consciousness and the stimulus is thus "recognized." The sum total of memory participates in all recognition; perceptual similarity is the central governing concept. As discussed below, these basic assumptions are reflected in many contemporary theories of memory and categorization (e.g., Hintzman, 1986; Nosofsky, 1986; Whittlesea, 1987).

Semon believed that every perceptual episode created a unique and permanent engram in memory. He also believed these engrams were in constant use during perception and recognition. Unlike other theorists of his time, he believed repetitions of a stimulus created new engrams, rather than strengthening old ones. These assumptions allowed Semon's theory to explain the permanence of specific memories. Given the *zeitgeist* of his era, however, Semon was also an introspectionist, and he realized that conscious expressions of memory are typically generic or abstract. For example, when remembering a friend, we are likely to recall an overall impression, rather than a particular encounter. The challenge to Semon's theory was to create abstraction from countless, partially redundant engrams.

The resolution of this problem apparently derived from Galton's work on photographic composites of faces. Galton discovered that blending many faces into a single composite created the image of a "generic" face. When discussing memory, Galton applied this phenomenon as a metaphor: "Whenever a single cause throws different groups of brain elements simultaneously into excitement, the result must be a blended memory" (Galton, 1883, pg 229).

According to Galton, abstraction occurs when multiple memory traces are accessed in parallel. Hintzman (1986) summarizes Semon's application of this logic:

> If several traces having common components were activated simultaneously, they were said to be in a state of *nondifferentiating homophony*-- a kind of resonant state in which the common properties of the traces stand out and their distinctive properties are masked, so that what appears in consciousness is an abstraction, rather than the content of a particular memory trace. The term *homophony* suggests music sung or played in unison, and Semon illustrated the notion with a chorus of voices metaphor. Homophony played a central role in Semon's theory of memory, and an entire chapter of his book was devoted to its role in abstraction (page 426).

Taken together, Semon's ideas about memory covered a broad range of topics with a small set of principles. His basic claims regarding multiple traces in memory and parallel access are commonly repeated today. His ideas were lost, however, for a variety of reasons (see Schacter et al., 1978, for speculation). After his death, the study of mind changed in emphasis with the rise of behaviorism. Eventually, when simultaneous advances in linguistic theory and other areas pushed cognition back into the foreground of research, the popular dogma had changed.

In the late 1950s, many psychologists grew dissatisfied with the behaviorist perspective of mind. The resurgence of mentalism arose from issues such as the "poverty of the stimulus argument:" Perception involves more information than the stimulus provides (Neisser, 1967; Fodor, 1983). Perceptual objects in the world convey more than their features; to the mind they convey their names, their category memberships, etc. The solution was to posit that mental representations exist in the mind and interact with the senses. These notions quickly led to the metaphor of "computer as mind." Computers operate on symbolic information, they process information in stages, and they embody the functional dualism of hardware and software. All of these qualities quickly appeared in psychological theory.

Another powerful influence in the cognitive revolution was development of Chomsky's generative grammar. In his well-know review of Skinner's *Verbal Behavior*, Chomsky (1959) argued that language is critically dependent on mental representation. Language is a creative, open-ended system. Semantics, syntax, and morphology are related in complex hierarchies. Chomsky argued that none of these critical properties of language can derive from a behaviorist theory. Since the development of Chomsky's theory, language has typically been characterized in terms of cognitive operations that access and manipulate linguistic symbols in memory. These symbols are assumed to embody the principle of *cognitive economy*-- an infinite set of grammatical utterances can produced or perceived via a finite set of symbolic units and combinatorial rules.

With the influence of linguistic theory and the computer metaphor of mind, the ideas found in Semon's theory were no longer in fashion. The emphasis in theory was on minimalist representation. Perception was assumed to work by information reduction; successive stages of information processing derive progressively abstract representations (e.g., Posner, 1964). Whereas Semon's theory emphasized the proliferation of redundant representations, later theories emphasized abstraction. Especially in theories of language, the decoding of specific perceptual episodes into canonical representations was a key assumption. This remains true of virtually all theories of language processing today.

In more recent years, however, theories of memory and categorization have evolved to their starting point; they have come full circle to Semon's theory. Indeed, with advances in their formal expression, modern exemplar theories of memory and categorization improve upon Semon's theory. Nevertheless, the similarities are striking and the issues are still the same. In the memory literature, many *global memory models* have been recently proposed. Global models assume, like Semon's theory, that countless memory traces share space in memory and are accessed in parallel.

Recent global memory models include the SAM model of recognition and recall (Gillund & Shiffrin, 1984), Hintzman's MINERVA 2 model (Hintzman, 1986; 1988), Ratcliff's random-walk model (1978), Murdock's TODAM model (1982), and Metcalfe's CHARM model (Eich, 1982). Although these models differ widely in their assumptions and implementations, all assume, like Semon's theory, that every experience leaves a unique trace in memory. They also all assume, like Semon's theory, that recognition occurs by parallel comparison of a new item to all the traces in memory.

Other models that bear resemblance to Semon's theory are exemplar models of categorization. In the context model of categorization (Medin & Schaffer, 1978; Nosofsky, 1984; 1986; 1987), a categorization or identification response is generated by comparing the similarity of a probe to all exemplars in memory. Recently, Nosofsky (1988b; 1991) has described the relation between the context model and other models of recognition memory. Finally, a wide class of connectionist, or *distributed memory* models of memory and categorization also embody the multiple-trace assumption (e.g., Knapp & Anderson, 1984; McClelland & Rumelhart, 1985), although these models do not assume separate traces for every experience. In short, the basic concepts of Semon's theory have established themselves as cornerstones of contemporary theory.

Memory for the Specific-- Some Theoretical Considerations

The present investigation was designed to examine memory for specific exemplars of spoken words. Memory for specific exemplars constitutes an important tool for theoretical analysis in areas beyond memory: Theoretical constructs in speech perception, spoken word recognition, and categorization all depend critically on their representational assumptions. Semon's primary motivation in developing a multiple-trace theory of memory was to account for the retention of specific information. Accordingly, it is natural to examine whether people actually retain specific information, or whether memory is more abstract. As it turns out, the literature on this issue is vast and the evidence is convincing.

Underwood (1969) described a theory of memory representation based on the retention of "attributes" of experiences. Underwood lists many attributes that are preserved in memory, but his basic claim was independent of the particular attributes observed in the laboratory. Memory traces are assumed to be multidimensional reflections of specific processing episodes: Anything perceptual, contextual, or that "comes to mind" (e.g., affect) during processing of a stimulus may become a represented attribute in memory. Underwood conceded that some attributes may be less stable in memory than others, but he maintained that episodic specificity was a cornerstone of representation. As discussed above, these assumptions are embedded in many newer theories of memory as well.

The opposing view to a theory like Underwood's is an abstractionist view. By this view, the specific perceptual features of a stimulus are encoded into a sensory store, but such "low-level" details are discarded as the input is subjected to higher levels of processing. A key assumption of the abstractionist view is that stimuli are *normalized* with respect to canonical mental representations. In order to recognize a stimulus as an example of an abstract category in memory (e.g., a word), the information processing system must somehow resolve and discard variable surface information that is tangential to the identity of the stimulus (e.g., Bruder & Silverman, 1974; Posner, 1969; Green, Kuhl, Meltzoff, & Stevens, 1991). In it's strongest form, therefore, the normalization hypothesis predicts that surface features of stimuli will be absent from long-term memory representation (see Besner & Coltheart, 1975).

Given the theoretical importance of claims about specific memory, numerous researchers have applied themselves to its assessment. The research reviewed below constitutes only a portion of the vast body of data such researchers have collected. The majority of studies reviewed were typical recognition memory experiments that revealed many important facts about memory for specific events. In recent years, however, a great deal of attention has been dedicated to comparing *implicit* and *explicit* measures of memory (also called *indirect* and *direct* measures; see Richardson-Klavehn & Bjork, 1991).

In recent years, the distinction between implicit and explicit memory has become a major topic in its own right (in 1991, fully one-fourth of all articles published in *Journal of Experimental Psychology: Learning, Memory, and Cognition* compared these memory systems). A full discussion of the topic is beyond the scope of this review. Much of the research reviewed below, as well as the present investigation, compared implicit and explicit memory, so a brief description is in order:

Explicit measures of memory entail a conscious recollection by the subject. Such measures have dominated memory research; free recall, cued recall, and recognition are the major paradigms of memory research. Implicit measures of memory typically entail a facilitation of task performance without conscious recollection (for examples and review, see Johnson & Hasher, 1987; Musen & Treisman, 1990; Parkin, Reid, & Russo, 1990; Richardson-Klavehn & Bjork, 1988; Roediger, 1990; Schacter, 1987; 1990; Tulving & Schacter, 1990; Tulving, Schacter, & Stark, 1982). The most common measure of implicit memory is the so-called *repetition effect*, or the change in recognition of an item with repeated presentations (e.g., Cofer, 1967; Jacoby & Dallas, 1981; Roediger & Blaxton, 1987). [1] In a repetition effect experiment, subjects typically perform some task on the same set of stimuli twice, and the difference between the first and second sessions is measured. Repetition effects are considered implicit memory measures because performance in the second session requires no conscious recollection of the first. Evidence of memory is often found in an implicit test, even when explicit tests reveal no memory at all. As an example, Jacoby and Dallas (1981) found that perceptual identification of words improved if the words had been previously read. This was true even when subjects had no recollection of previously reading the words.

The implicit/explicit memory distinction is relevant to the present investigation for several reasons. First, as reviewed by Schacter (1987), memory for specific details of

---

[1] A common usage in the literature on implicit and explicit memory is to refer to repetition effects as ``priming." Following Jacoby and Brooks (1984; see also Jacoby, Marriott, and Collins, 1990), I prefer to avoid this label for such effects. As Jacoby and Brooks note, ``priming" implies that the activation of an abstract unit, such as a logogen (Morton, 1969) has been temporarily heightened by the presentation of a stimulus. The term therefore implies a theoretical predisposition toward abstractionism. Moreover, it is not clear that repetition effects observed a day or a week after study can be considered ``priming" in the logogen sense.

stimuli is frequently revealed by implicit measures, but is not found by explicit measures. This implies, as argued below, that the specific details in question affect automatic perceptual processes more than conscious memory retrieval. Second, the specific perceptual details revealed by implicit and explicit memory measures display different levels of sensitivity to task demands. For example, the effects of specific details appear to rapidly disappear from explicit memory, but last for long periods in implicit memory (Tulving, Schacter & Stark, 1982). Also, the level of processing of a stimulus during study (Craik & Lockhart, 1972) has a strong effect on explicit memory for surface details, but has almost no effect on implicit memory for the same information (Graf, Mandler, & Haden, 1982; Graf & Mandler, 1984). Both of these task manipulations were examined in the present research.

Memory for the Specific-- Some Empirical Findings

Memory for the specific details of experiences has been investigated in many domains, including memory for pictures, faces, voices, social interactions, and the specific visual features of text. The weight of the evidence suggests that, with sufficiently sensitive measures, the specific details of experiences can be found in memory.

Memory for Faces

Some of the most striking evidence of memory for specific detail has been reported in studies of memory for faces (see Bruce, 1988). Face recognition has been shown to be highly accurate over surprisingly long periods of time. Bahrick and his colleagues (Bahrick, Bahrick, & Wittlinger, 1975) studied adults' recognition of faces from their high-school yearbooks. Although subjects were not always able to generate names from pictures, they were able to discriminate photographs of their former classmates from foils. This was true even for subjects who had graduated high school 50 years previously. Recently, Bruck, Cavanagh, and Ceci (1991) extended this finding by showing adult subjects photographs from their high school yearbooks 25 years after their graduation. In this experiment, however, subjects did not merely classify the faces as classmates and foils; they were required instead to match the old photographs to new ones taken at their 25th reunion (none of the subjects attended the reunion). Once again, performance was well above chance.

Goldstein, 1977; Chance, Goldstein, & McBride, 1975) presented new faces to subjects in a study phase and tested their recognition memory after a delay. Memory for faces seen just once was excellent (above 90%) even after a 6-month delay, attesting to the longevity of face-specific memories.

Memory for Pictures

Research on the perception and recognition of pictures has received considerable attention as well. Rock and Engelstein (1959; see also Rock, Halper, & Clayton, 1972) showed subjects novel figures once during a study session; weeks later, subjects could accurately select these figures from sets of similar distractors. Shepard (1967) reported accuracy of recognition memory for pictures above 98%. Nickerson (1965) showed subjects 600 pictures in a continuous recognition paradigm and observed recognition accuracy above 95%. He also tested his subjects a day later, showing the 200 old and 200 new pictures in a recognition memory test; accuracy was 92% after 1 day. A year later, recognition accuracy was still an impressive 63%. In an even more striking result, Standing, Conezio, and Haber (1970) showed subjects 2,560 color slides, one slide every 10 seconds, and found recognition accuracy above 90% after several days.

Increasing the presentation speed to one slide per second did not reduce accuracy; showing slides with left-right reversal at test also did not reduce accuracy.

More recently, Kroll and Potter (1984) examined repetition effects for pictures and their corresponding object labels and found that significant facilitation of picture classification required preservation of physical format; switching from pictures at study to words at test resulted in attenuated repetition effects. Similar findings have been reported by Kersteen-Tucker (1991) with symmetric polygons. Biederman and Cooper (1991; see also Roediger & Srinivas, 1992) reported repetition effects for exact matches of visual patterns; even slight variants of patterns seen at study led to reduced repetition effects at test.

Musen and Treisman (1990) compared implicit and explicit memory of novel visual patterns immediately after study or one week later. Implicit memory for specific patterns (measured via perceptual identification) was evident after the one-week delay, although explicit recognition was significantly reduced. In a similar investigation, Cooper, Schacter, Ballesteros, and Moore (1992) compared implicit and explicit memory of unfamiliar 3-dimensional objects. Implicit memory for specific objects was robust despite changes in size or rotation, but explicit recognition was markedly impaired by such changes.

Memory for Music

Highly specific memory has been observed for auditory patterns as well as faces and other visual patterns. For example, Halpern (1989) recently found that subjects displayed retention for the absolute pitch of familiar songs. These results were observed for musicians and non-musicians alike. Crowder, Serafine, and Repp (1990) found that memory for the melody and lyrics of songs are dependent upon one another; their association in perception necessitates their association for recall.

Memory for Social Interactions

Memory for the specific has also been investigated in social cognition. Smith (1990; Smith and Zarate, 1992) has recently argued in favor of an exemplar-based model of social judgment; he cites numerous studies in which memory for specific social interactions affects later behavior. For example, Lewicki (1986) had subjects participate in a questionnaire study. As some subjects were working on their questionnaires, an experimenter entered the room and insulted them. Subjects were instructed to give their completed questionnaires to either of two assistants in another room; both assistants were free at all times, so the choice was arbitrary. Subjects who had been insulted by the experimenter tended to avoid the assistant whose hairstyle resembled the hairstyle of the rude experimenter. Control subjects displayed no preference.

In a similar experiment, White and Shapiro (1987) had subjects converse with confederates over the telephone. Half of the subjects were led to believe that the person on the other end of the line physically resembled one of their close friends. All of the conversations were equivalent in content. After the conversation, misled subjects rated the confederates as similar to their friend along important personality dimensions. Control subjects showed no such pattern. In a more compelling experiment, Anderson and Cole (1990) had subjects list attributes of someone they knew well, attributes of a well-known social stereotype (e.g., redneck), and attributes of an abstract social category (e.g., honest person). Later, in an unrelated memory experiment, subjects read descriptions of fictitious people and their recognition memory for these descriptions was tested. During test, when the description of a fictitious person included attributes from the subject's description of a specific friend, subjects were likely to falsely report that the attribute had been part of the description presented at study. This false alarm tendency was much weaker for either the stereotypes or the abstract categories. Anderson and Cole concluded that representations of specific individuals in memory constitute more powerful bases of inference than general categories. Smith (1990) maintains that we make social inferences in this manner every day (e.g., we irrationally dislike a stranger who resembles an unpleasant colleague.

Memory for Physical Dynamics

For many years, researchers have attached luminous patches to the limbs of walkers to study the physical dynamics of the human gait. The procedure is the following: Small patches are attached to the major joints of a person's body. The person is then filmed walking across a dark room, so all that appears in the display are the glowing spots. When the film is held frozen, the spots appear independent and disjointed. When the film is played without interruption, however, the coordinated movement of the spots is easily recognized as a person walking (Johansson, 1975).

Johansson (1973; 1975) refined this point-light display technique to isolate the regularities of movements. Since then, it has been found that invariant relational aspects of movements define the human gait (Cutting, Proffitt, & Kozlowski, 1978), and that individual characteristics, such as gender, are easily perceived from point-light displays (Cutting & Kozlowski, 1977).

From the perspective of the present review, perhaps the most interesting finding is that people can recognize their friends and acquaintances from the dynamics of a point-light display. Cutting and Kozlowski (1977) filmed six undergraduate students in a point-light procedure. Two months later, the students returned and tried to identify themselves and each other from the films. Accuracy was well above chance, even though these subjects had little prior exposure to each others' gaits. Cutting and Kozlowski suggested that recognition for truly familiar gaits (e.g., gaits of close friends or relatives) would be nearly perfect.

Thus, even in the absence of cues typically considered synonymous with identification (e.g., face, body shape, hairstyle), subjects can easily identify other people from the dynamics of their physical movements. These findings suggest that specific perceptual details of dynamics are encoded and stored in long-term memory. Upon presentation of a similar dynamic display, recognition occurs without effort.

Memory for Frequency

Underwood (1969) cited frequency as one of the major attributes retained in memory (Underwood & Ekstrand, 1968; Winograd, 1968). Following Underwood's claim, memory for frequency was been widely investigated. Indeed, only four years after publication of Underwood's article, Howell (1973) published a review article summarizing the data on memory for frequency. Many studies have shown that frequency of exposure affects performance in many ways; the term "memory for frequency" is intended here to denote a subjects' ability to estimate the number of times a stimulus was presented in an experimental session.

Although the data on memory for frequency are vast, the basic findings indicate that memory for frequency is quite accurate, and that frequency is automatically encoded (Hasher & Zacks, 1979). Hintzman and Block (1971) found nearly perfect frequency discrimination (above 90%) of the same words across lists. Hockley (1984) had 4 subjects generate over 46,000 absolute frequency judgments and found a strong relation between estimates and actual frequency. Zacks, Hasher, and Sanft (1982) found that frequency discrimination is unaffected by practice, test expectations, or competing demands, implying that frequency is automatically encoded.

Hintzman and his colleagues (e.g., Hintzman & Block, 1971; Hintzman, Grandy, and Gold, 1981; Hintzman, 1988) have maintained for more than 20 years that memory for frequency is best explained by a multiple-trace memory model. The automatic nature of frequency encoding, and it's non-propositional nature support Hintzman's interpretation. Hintzman, Nozawa, and Irmscher (1992) showed subjects pictures paired with digits a variable number of times. Estimated frequencies of picture presentations were quite accurate and showed no intrusion from the digits. When frequency information for picture-digit pairs was communicated verbally, strong interference effects were observed. Hintzman et al. concluded that multiple, analog representations of separate presentations were the basis of frequency estimation in their first experiment, independent of propositional information regarding the association of pictures and digits.

Supporting evidence for the multiple-trace theory of frequency memory comes from recent studies of semantic associates and memory for typeface. Begg, Maxwell, Mitterer, and Harris (1986) investigated accuracy of frequency estimates to new associative pairs and found that, as recall improved with superior cues, frequency estimates improved as well. This was true even when words were paired with multiple associates during study. Begg et al. concluded that frequency estimates contact all prior episodes. More recently, Johnson, Peterson, Yap, and Rose (1989) showed subjects 16-square grids containing 7 letters and digits in various typefaces. When subjects were later asked to estimate the number of times any given letter or digit appeared in any given typeface, their performance was quite accurate and was independent of test expectations. Hintzman's (1988) multiple-trace model of memory is consistent with all of these findings.

Memory for Modality

More in line with the present investigation, many experiments have shown that subjects can remember the modality through which they first experienced a verbal stimulus. Underwood (1969) also identified modality as a persistent attribute in memory (Murdock, 1967). Hintzman, Block and Inskeep (1972; see also Hintzman, Block, & Summers, 1973) presented subjects with lists of words followed by surprise recognition tests. In one condition, half of the words presented at study were visual and half were auditory; assignments of half the words to modalities were changed at test. Recognition was superior for words that were presented in the same modality for both sessions.

Other researchers have directly inquired about memory for modality. Light, Stansbury, Rubin, and Linde (1973) showed subjects pictures and words and later found high performance on a memory test for original presentation mode. Bray and Batchelder (1972) presented subjects with visual and auditory words; recognition of original mode of input was well above chance. This held true whether the recognition test was expected or not, and whether the test was given immediately or after a 15-minute delay.

Kirsner and his colleagues investigated memory for modality in several experiments. In two continuous recognition experiments (cf. Shepard & Teghtsoonian, 1961), Kirsner (1974) found that recognition was superior for words presented and repeated in the same modality (auditory vs. visual). He also found that subjects could explicitly remember the modality of a word's original presentation for several minutes. In a similar experiment, Kirsner and Smith (1974; see also Kirsner, Milech, & Standen, 1983) presented words and nonwords in a "continuous lexical decision" task (i.e., words and nonwords were presented and later repeated after variable delays). Subjects performed lexical decisions on stimuli from these lists, and were told to ignore the fact that half of the words were visual and half were auditory. Repetition facilitated lexical decisions for words and nonwords presented and later repeated in the same modality; cross-modality repetition facilitated lexical decisions slightly for words and not at all for nonwords.

Kirsner (1974) suggested that information about modality is encoded directly and automatically, as an attribute of each exemplar. Other investigations have assessed the automatic nature of modality retention. Lehman (1982) investigated memory for modality across age groups (grades 2, 3, 6, and college), expected vs. surprise tests, list length, and relatedness of words in the list (words were grouped into categories or were random). All manipulations affected overall recall of words, but none affected the retention of modality information, which was well above chance. Lehman also concluded that modality is encoded automatically (cf. Hasher & Zacks, 1979).

More recently, Schacter and Graf (1989) investigated the modality specificity of explicit and implicit memory for new associations. Subjects performed fragment completion to either visual or auditory word fragments. All fragments were paired with complete words at study, so subjects could derive associations between the words. When items were repeated later, implicit, fragment completion was very sensitive to modality shifts, but explicit memory for associations was not. In a similar experiment, Kelley, Jacoby, and Hollingshead (1989) compared memory for modality using both explicit and implicit memory measures. Kelley et al. found that the tests were related to each other via perceptual fluency of repeated items: Same-modality repetitions increased perceptual fluency, thereby increasing perceptual identification in the implicit test, and serving as a familiarity cue in the explicit memory test. When other mnemonic cues were provided, subjects relied less on perceptual fluency and the tests displayed more independence. Like Kirsner (1974), Kelley et al. suggested that modality is encoded directly into exemplars which are accessed during later perception. Because exact matches increase perceptual fluency and stimulus familiarity, subjects display savings in a perceptual test and sensitivity to encoding conditions in an explicit test.

Memory for Sentence and Discourse Details

The findings on memory for modality reveal only one aspect of specific memory for verbal material. Another line of research has shown that subjects can also remember specific details about text. Two major sources of evidence have been provided: memory for location of information in text, and memory for exact sentence wording.

With respect to location of information in text, Rothkopf (1971) evaluated the common intuition among students that recall of substantive information from text is related to recall of information about location within text. Rothkopf had subjects read a 3000-word text, then he administered surprise recall tests for substantive information and the location of information in the text. Memory for location was above chance and correlated with memory for substance. As Rothkopf notes,

> The results of this experiment indicated that readers did have some incidental memory for the location of information on a page. Anecdotal reports about this apparently have some basis in fact. (page 612).

Rothkopf's findings prompted others to investigate the phenomenon of location memory in text. Zechmeister and McKillip (1972) replicated Rothkopf's findings, and Zechmeister, McKillip, Pasko, and Bespalec (1975) reported that location memory was not substantially improved by warning subjects it would be tested. In a more extensive study, Lovelace and Southall (1983) investigated the relation between memory for location and substantive information. They found that depriving readers of spatial cues by having them read from a continuous scroll significantly reduced word recall. Reinstating spatial cues improved word recall. Moreover, providing cues to aid word recall was found to incidentally benefit location recall. Lovelace and Southall argued that words are encoded as a constellation of features; the particular features encoded are not determined entirely by linguistic relevance.

In addition to memory for spatial location of information, people display memory for the exact wording of sentences. Begg (1971; Begg & Wickelgren, 1974) presented auditory and visual sentences in a continuous recognition memory experiment. Subjects were instructed to judge both the meaning and the wording of each sentence as "old" or "new." Both forms of judgment were well above chance. Accuracy of meaning judgments decreased as a function of lag; accuracy of wording judgments was not correlated with either lag or accuracy of meaning judgments. Finally, whereas meaning judgments were not affected by modality, wording judgments were more accurate for visual sentences.

Other demonstrations of memory for specific wording have been reported by Levy (1983) who observed improved proofreading speed and accuracy for familiar text. Memory for wording is improved when sentences require pragmatic inferences, (Gibbs, 1981), and also when wording is particularly complex (McDaniel, 1981). Recognition memory for the specific wording of conversation has been found to be extremely accurate 30 hours afterward (Keenan, MacWhinney, & Mayhew, 1977), and even 72 hours afterward (MacWhinney, Keenan, & Reinke, 1982).

Masson (1984) investigated memory for sentence wording further by comparing implicit and explicit measures. Masson showed subjects sentences during study and tested their recognition either immediately, two days later, or one week later. During test, the sentences were repeated either in their original form, with new wording, with typographic transformation (see below), or with both new wording and typography. As explicit measures, subjects gave recognition judgments about meaning, wording, and typography. As implicit measures, savings for original surface forms were assessed from recognition judgments and reading times (Kolers & Ostry, 1974). Masson observed memory for surface form that remained robust up to a week after original presentation: All measures displayed significant effects of surface form. Furthermore, memory for surface form was independent of the deterioration of content memory over time, and was independent of the depth of semantic processing during study. Masson concluded that precise surface details of sentences are preserved in episodic memory; memory of first reading and fluency of re-reading depend on activation of specific prior episodes.

Memory for Transformed Text

Perhaps the most well-known findings on memory for specific episodes come from Kolers' experiments on transformed text. His research focused on memory for the specific visual features of text; he studied reading times and explicit recognition to assess memory for specific reading episodes. Most of Kolers' experiments examined memory for transformed text, although he also examined bilinguals' memory for specific language and wording for sentences (e.g., Kolers, 1974a; 1978). Transformed text, or rotated text, is typically created by inverting normal text on its X-axis, so that it appears as a mirror image of normal text, although other transformations have been studied as well.

Over a period of about five years, Kolers collected an impressive body of data on memory for transformed text; throughout all reports, he maintained that human memory permanently encodes the "irrelevant" surface features of text. Kolers (1973) had subjects read sentences in eight different transformations; later subjects classified sentences as "old" or "new." During test, half of the sentences were presented in the old typography and half were presented in a new typography. Memory for "old" sentences was better if the original typography was repeated. This advantage was not related to the content of the sentences or the amount of time taken to decode the transformed text. Kolers (1975a; 1975b; Kolers & Magee, 1978; Kolers, Palef, & Stelmach, 1980) found that, as reading time of inverted text decreased with practice, memory for inverted text decreased as well. Also, Kolers found that reading time for old sentences was decreased more by repetition if the original typography was repeated than if the typography was changed (Rudnicky & Kolers, 1984).

In related research, Kolers examined memory for transformed text over longer delays. Kolers and Ostry (1974; Kolers, 1974b) tested subjects' memory and reading time for transformed sentences over variable delays. Subjects read normal and transformed sentences during study; during test the typography of half of each kind of sentence was switched. Subjects read the sentences again (in the context of foils) at test and judged them "old" or "new." The results were striking: Even after 32 days delay, subjects' reading times were reduced and recognition was increased for sentences repeated in their original typography. Extending this research still further, Kolers (1976a; 1976b) timed subjects reading pages of inverted text during a study session. Subjects returned to the laboratory 13 to 15 months later and were timed again while reading the old passages and new passages. A small but reliable savings was observed for the old passages, even though more than a year had elapsed.

From all these findings, Kolers maintained that the *cognitive operations* performed during the reading of a text form an integral part of the memory of the text. Kolers espoused a procedural view of mind (Kolers & Roediger 1984), wherein the operations directed at a stimulus become associated with the stimulus. By extension, the pattern analyzing operations used in reading transformed text became part of the memory representation for the text; reinstating the typography induced similar operations during re-reading. Therefore, reading speed and recognition memory were improved.

Kolers argued that the savings for old passages in their original typography was independent of memory for lexical or semantic content. This claim, however, was contested in later research. Tardif and Craik (1989) reported no evidence of memory for specific perceptual information after a delay of only one week. Instead, they reported strong transfer for the general skill of reading any transformed text, and memory for the semantic content of old sentences. It should be noted, however, that Craik and Gemar (1990) later reported finding evidence for perceptual savings in re-reading transformed text without effects of semantics. Horton (1985) found that re-reading times for sentences presented twice in the same typography were no faster than re-reading times for sentences presented in two different typographies. Horton concluded that semantic memory was more important than perceptual memory (see also Graf & Levy, 1984). In later research, however, Horton (1989) found that perceptual features clearly affect reading time if semantic cues to recall are not provided. Graf (1982) found that generating words had the same effect on memory as reading words in rotated form; he concluded that any distinctive cognitive operations yield increased perceptual memory. Masson (1986), however, found that facilitation of reading isolated words in transformed text was word-specific; no transfer of general skill was observed. Masson argued that fluent reading of transformed words is based on memory for specific instances of earlier words sharing the same visual features. Later, from evidence of context-dependent repetition effects, Masson and Freedman (1990) argued that fluent identification of repeated words is based on long-term episodic memory for physical features *and* semantic interpretation.

It is clear that both data and theory about the transformed text paradigm remain controversial. No doubt further research is currently underway to resolve the true nature of the beneficial effects of pattern analyzing operations. Related research, like Masson's (1986) study of individual word recognition, has provided support for a view similar to Kolers' view. Instead of arguing in favor of unique perceptual operations as the basis of improved memory; the findings reviewed below implicate memory for specific instances as the source of perceptual facilitation.

Memory for Visual Word Details

Although the findings reviewed above attest to the persistence of perceptual details in memory, perhaps the most compelling evidence to date comes from investigations of the recognition of isolated words. Evidence from visual word recognition is summarized here; comprehensive reviews are also provided by Kirsner and Dunn (1985; also Kirsner, Dunn, and Standen, 1987) or Jacoby and Brooks (1984). Evidence from auditory word recognition is summarized below.

Hintzman and Summers (1973) discuss the early motivations to study memory for the specific visual attributes of words: Research on visual memory had demonstrated robust long-term retention of specific visual scenes, complex figures, and faces (e.g., Shepard, 1967). Theories of lexical processing, however, maintained that the visual features of words are quickly forgotten after more "stable" symbolic memory codes are derived (e.g., Morton, 1969; 1979). Why should visual memory for some stimuli last so much longer than visual memory for others?

Hintzman and his colleagues (Hintzman et al., 1972; Hintzman & Summers, 1973) conducted several experiments to directly assess long-term memory for the visual details of words. Subjects saw words in uppercase block letters and lowercase script letters. In a subsequent recognition test, accuracy was higher for words repeated in their original format. Hintzman and Summers (1973) conducted a similar experiment measuring latencies of correct recognitions. As in the earlier accuracy data, they found that recognition decisions were made faster for words tested in the same format as their original presentation. More recently, Graf and Ryan (1990) also reported format-specific repetition effects in both implicit and explicit tasks.

Another early demonstration of visual memory for words was provided by Kirsner (1973) in a continuous recognition experiment. Words were first presented either in uppercase or lowercase letters. When repeated later, half of the words retained their original format and half were switched. In another experiment, the same design was used, but nonword letter strings were presented. In both experiments, same-case repetitions were more likely to be recognized than different-case repetitions, although the effect was larger for nonwords than for words. In both experiments, subjects discriminated well above chance whether "old" stimuli were presented in their original case. These differences in memory between same- and different-case repetitions was found across all values of lag; memory for case lasted at least 2-3 minutes.

Roediger and his colleagues have reported several investigations of implicit and explicit memory that demonstrate the persistence of memory for perceptual details. For example, Roediger and Blaxton (1987) had subjects complete word fragments that were either typed or handwritten. Afterward, they presented the same fragments for completion again, but the formats of half of the stimuli were reversed. Repetition effects, measured implicitly via fragment completion, were greater for the fragments that were repeated in their original format. This advantage for same-format repetitions was observed with an immediate test and also after a one-week delay. Recently, very similar results were reported by Manso de Zuniga, Humphreys, and Evett (1991), although their delay was shorter.

Jacoby and his colleagues have contributed a great deal of research and theory to the study of implicit and explicit memory, the role of specific episodic traces in perception, and the nature of memory for perceptual details. Regarding memory for perceptual details, Jacoby and Hayman (1987) found that changing the typeface of words between study and repetition significantly reduced accuracy of perceptual identification. They concluded that visual details of words persist in memory, and they suggested that perceptual identification relies on memory for prior perceptual episodes.

Despite the findings reviewed above, many investigators have observed only small and insignificant effects of changing format in repetition paradigms. That is, performance for words repeated in a new modality or typeface equals performance for exact repetitions. For example, Carr, Brown, and Charalambous (1989) had subjects read paragraphs aloud two times. Between readings, the formats of half of the texts were changed, switched from typed to handwritten format or vise-versa. Magnitudes of repetition effects were equivalent for the exact repetitions and the new-format repetitions. Carr et al. argued that the repetition benefits derived from priming of abstract word units, such as logogens, that are insensitive to surface variability. Similarly, Scarborough, Cortese, and Scarborough (1977; see also Brown, Sharma, and Kirsner, 1984) found no difference in repetition benefits when words presented in a lexical decision task were repeated in an old or new typeface. Clarke and Morton (1983) observed no effects of study-test typography differences on a recognition test. Feustel, Shiffrin, and Salasoo (1983) found only small effects of typeface in a repetition paradigm and suggested that abstract and episodic representations simultaneously mediate repetition effects.

More recently, Levy and Kirsner (1989) and Graf and Ryan (1990) also reported that visual details such as typeface effect the magnitude of repetition effects in some experimental conditions, but not in all conditions. Many investigators (e.g., Masson & Freedman, 1990) have asserted that effects of specific visual details are only observed when subjects must focus their attention at the word level during reading. Effects should be observed, for example, when the presentation formats used are highly distinctive or require excessive processing, such as rotated text. When relatively normal texts are read in a natural comprehension task, no memory of visual detail should be observed (see Masson, 1989).

From all these null findings, the role of specific perceptual episodes in later perception and memory may seem diminished. Carr et al. (1989) and Feustel et al.

(1983) proposed more abstract representations to account for their findings. A study by Woltz (1990), however, directly compared the time courses of semantic priming effects and repetition effects. Reliable effects of typography were observed and the magnitudes of the differences in repetition between same- and different-typography trials were equivalent across delays. In contrast, facilitation of performance due to conceptual repetition decreased steadily across delays. Woltz (1990) suggested that abstract lexical representations are not involved in persistent repetition effects.

More recently, Jacoby, Levy, and Steinbach (1992) argue that, despite the null results in the literature, the positive findings of Jacoby and Hayman (1987) and others underscore the potential importance of specific perceptual episodes in representation. Jacoby et al. (1992) demonstrate that repetition effects are affected by specific visual details even in a high-level reading task. Subjects read interrogative sentences silently and answered their questions aloud. When presented with the same questions later, the typeface was either the same, was changed to another, or the sentence was spoken. Large and reliable effects of presentation format were observed in three experiments under a variety of procedural manipulations, including a 24-hour delay. From these findings, Jacoby et al. argue that null findings of perceptual specificity are due to the nature of processing demanded in particular tasks, not due to different levels of lexical abstraction. Moreover, effects of perceptual specificity are not dependent upon difficult encoding operations, as Kolers (1976a) would suggest; one merely needs a task that maximizes the involvement of episodic memory and perceptual fluency in performance.

Theories of Visual Word Recognition

The growing body of data discussed above may be summarized as follows: Ample evidence suggests that specific perceptual events leave lasting episodic traces in memory. These episodic traces may be revealed in an explicit assessment of memory, but are more strongly evident when assessed via implicit, perceptual tests. Given these results, the next logical step for a theory of word recognition is clear: If specific episodic traces of words reside in memory and affect perceptual tests, why not assume they constitute the representational substrate of the mental lexicon?

Several theorists have observed these phenomena and have proposed theories of word recognition based on episodic representations. Jacoby (1983a; 1983b; Jacoby & Witherspoon, 1982; Jacoby & Brooks, 1984; Jacoby & Hayman, 1987; Jacoby, Marriott, & Collins, 1990) has argued in favor for such a model for years. Treisman (1978) described a formal model of identification applied to a collection of "word images." Feustel, Shiffrin, and Salasoo (1983; see also Salasoo, Shiffrin, and Feustel, 1985) describe a hybrid model of word recognition based on the interaction of episodic and abstract representations. Monsell (1991) makes similar suggestions. Kirsner and Dunn (1985; see also Kirsner, Dunn, and Standen, 1987; Dunn & Kirsner, 1989) have argued for a theory of word recognition based on procedural records, in the sense of Kolers (1976a).

Similar proposals are found in several current theories of memory and categorization. Although these theories do not address word recognition per se, their basic assumptions about representation and identification make them applicable to the problem. McClelland and Rumelhart (1985) discuss a memory system based on distributed representations and the possible effects of multiple episodic traces (superimposed on each other) in stimulus recognition. Hintzman (1986; 1988; see also Nosofsky, 1984; 1986) proposes a multiple-trace memory model that applies to categorization and recognition memory. Since word recognition may be a special case of these related processes, Hintzman's model may be considered a viable model of word recognition.

Considering the recent empirical and theoretical interest in an episodic approach to word recognition, one might conclude that the complexion of the field is changing, that new ideas about representation are cropping up in numerous discussions of lexical access. Unfortunately, the conclusion would be wrong.

For numerous reasons, the majority subject only to speculation, the mainstream literature on word recognition has not changed to accommodate new ideas about episodic representations. For lack of a better term, the dogma of abstractionism does not seem to be changing, despite the efforts of Jacoby and others. Most models of word recognition, especially the models that are widely cited, rely on abstract representations, such as "nodes," and template-matching processes to achieve access. Rather than address these assumptions or data from the memory and categorization literature, most models are applied to resolving the nature of word frequency and context effects, semantic priming, and the like. These basic attributes are found in Morton's (1969; 1979) logogen model, Forster's (1976) search model, Paap, Newsome, McDonald, and Schvaneveldt's (1982) activation-verification model, McClelland and Rumelhart's (1981; Rumelhart & McClelland, 1982) connectionist model, and most others as well. Certain models, such as the activation-verification model assume that lexical nodes may be prototypes, derived from countless exemplars, but this is a minor concession toward a theory like Jacoby's.

The persistence of the abstractionist dogma in word recognition is apparently due to a variety of factors. I consider only a few possibilities: First, many studies that reveal episodic effects on perception are couched as investigations of memory, often comparing implicit and explicit memory. Few authors, with the exceptions of Jacoby, Feustel et al., and Kirsner have emphasized the relation of implicit memory measures and generic word recognition. Second, models of word recognition traditionally address the identification of stimuli in the *mental lexicon,* following Oldfield's (1966) initial suggestion that words reside in a special area of memory. Third, the ideas of modularity (Fodor, 1983) have been widely applied to language, implying a virtual independence of word recognition and general perception. Fourth, and perhaps most importantly, contemporary psycholinguistics originates from traditional linguistic theory. As such, Chomsky's early assumptions of deep, abstract structures and the competence/performance distinction are reflected in theories today.

Spoken Language Processing

Studdert-Kennedy (1976) begins his well-known review chapter on speech perception with the standard assumptions of all theories:

> This paper is concerned solely with phonetic perception, the transformation of a more-or-less continuous acoustic signal into what may be transcribed as a sequence of discrete, phonetic symbols. The study of speech perception, in this sense, has in recent years begun to adopt the aims, and often the methods, of the information-processing models of cognitive psychology which have proved fruitful in the study of vision... The underlying assumption is that perception has a time-course, during which information in the sensory array is "transformed, reduced, elaborated" (Neisser, 1967) and brought into contact with long-term memory (recognized).

Studdert-Kennedy's quote suggests that, with respect to abstract representation, if dogma is evident in theories of visual word recognition, it is rampant in theories of speech perception and spoken word recognition. Most theories assume that the speech waveform is converted to a sequence of discrete features and segments; this sequence is then compared to abstract representations of words in the mental lexicon (e.g., Forster, 1976; Marslen-Wilson & Welsh, 1978; Oldfield, 1966; Paap, Newsome, McDonald, & Schvaneveldt, 1982). As in theories of visual word recognition, although the particular processing assumptions vary, the general notion of an abstract, canonical lexicon is accepted without question.

The basic assumptions of theories of speech perception and spoken word recognition stem from the extremely variable nature of the speech signal (for reviews, see Klatt, 1979; 1989; Goldinger, Pisoni, and Luce, in press). The acoustic realizations of phonetic contrasts are sensitive to changes across phonetic contexts, prosodic contours, speaking rates, and especially speakers. Four decades of speech research have failed to determine invariant properties of the speech signal that a generic pattern-matching system could identify as phonemes. Nevertheless, perception of acoustically variable patterns is categorical (Liberman, Harris, Hoffman, & Griffith, 1957); perception follows categories of phonemes and words rather than the analog signal. Apparently, some set of processes encodes the variable signal into its corresponding mental representations.

Almost every review of research and theory in speech perception refers to the "problem of variability--" the speech processing system somehow compensates for the unpredictable variance in the signal (e.g., Pisoni, 1978). The oldest and most influential theory of speech perception, the motor theory, invokes a specialized module that relates the sounds of speech to articulatory gestures (see Liberman, 1970; Liberman & Mattingly 1985; 1989). Although other theories of speech perception promote different views, virtually all are designed, first and foremost, to address the problem of variability. Most models of spoken word recognition assume that some undefined, sublexical speech pre-processor has already converted the signal into a string of discrete phonemes for comparison with canonical entries in the mental lexicon. Indeed, many models of visual word recognition (e.g., Morton, 1969; Forster, 1976; McClelland & Rumelhart, 1981) have been presented as models of spoken word recognition, with the minor substitution of phonemes for letters as the input.

Speaker Normalization

Sources of variability in speech, whether from phonetic context, stress, speaking rates, or speaker differences are typically considered "perceptual problems" to be solved by listeners, just as they must be solved in the design of any robust speech recognition systems (Gerstman, 1968; Shankweiler, Strange, & Verbrugge, 1976). A long-standing assumption in speech perception concerns the processes of *speaker normalization* in understanding speech from different speakers. Individuals differ in terms of the sizes and shapes of their vocal tracts (Fant, 1973; Joos, 1948; Peterson & Barney, 1952), glottal characteristics (Carr & Trill, 1964; Carrell, 1984; Monsen & Engebretson, 1977), their idiosyncratic articulatory strategies for producing phonemes (Ladefoged, 1980), as well as the dialects of their native regions. As such, there is widespread variability in the production of the same words and phrases across individuals. Nevertheless, human listeners accurately recognize speech across most speakers without any apparent difficulty. This paradox has been an important theoretical problem since the beginning of modern speech research.

Joos (1948) suggested that, given physical and dialectic differences across speakers, the perceptual system must somehow "calibrate" itself to each speaker. After calibration, the perceptual system can recognize the phonetic segments encoded in the speech waveform. Speaker normalization is generally characterized as a process that allows the listener's attention to follow the phonetic and semantic content of the spoken message, while the superficial characteristics of the signal can be exploited by the perceptual machinery and then discarded (e.g, Krulee, Tondo, & Wightman, 1983; Syrdal & Gopal, 1986). Taken together, the assumptions of speaker normalization and an abstract lexicon suggest that long-term memory for spoken words should consist of the elements of meaning; the elements of perception, such as voice-specific details, should be lost.

As the following review will show, several perceptual studies demonstrate that speech perception and spoken word recognition is more difficult in situations of high speaker variability. Several investigators have interpreted such effects as evidence that a capacity-demanding normalization process is invoked by each new voice. Further evidence from memory experiments, however, suggests that, if normalization occurs, it certainly does not result in loss of voice information from memory. Rather, voice-specific information appears to last alongside lexical information, at least for brief periods of time.

Speaker Variability and Speech Perception

One of the first empirical demonstrations of the effects of speaker variability was provided by Ladefoged and Broadbent (1957; although see Peters, 1955a, 1955b). They presented listeners with the synthesized sentence, "Please say what this word is:" followed by either *bit*, *bet*, *bat*, or *but*. The carrier phrase was altered in different conditions by raising (by 30%) or lowering (by 25%) either the first or second formant, or both. This manipulation changed the perceived dimensions of the speaker's vocal tract. Ladefoged and Broadbent observed reliable changes in identification of the target syllables depending on the perceived speaker. The authors concluded that the carrier phrase allowed the listener to "calibrate" the vowel space for each speaker and thereby adjust their perceptions of the target vowels accordingly.

In later research, however, Verbrugge, Strange, Shankweiler, and Edman (1976) questioned the premise of this approach. They noted that, despite speaker variability, error rates in vowel identification were rather low (e.g., only 4% in Peterson and Barney's, 1952, experiments). Verbrugge et al. re-examined vowel identification across speakers and found that accuracy of vowel identification was generally quite high, despite speaker variability. They also found that providing examples of a speaker's point vowels did not improve performance, in contrast to the notion of calibration. Finally, they found that listeners adjust their perceptual criteria according to perceived rate of articulation more than for perceived length of vocal tract. From all these data, Verbrugge et al. (1976) concluded that speaker normalization is either a process that requires very little prior information or is not a process that occurs in speech perception at all (see also Strange, Verbrugge, Shankweiler, & Edman, 1976). In later research, Verbrugge and Rakerd (1986) investigated the perception of "silent-center" bVb syllables. The authors presented listeners with /bVd/ syllables spoken by males and females, that had the middle 60% removed, leaving only the beginning and ending transitions with silence in between. Their results showed that considerable vowel identity information is contained in the formant transitions, and that this information is speaker-independent.

The results summarized above show that speaker variability does not strongly impair perception, suggesting that speaker normalization may be an unnecessary theoretical construct. This suggestion is supported by studies of development as well. Experiments conducted by Kuhl (1979) and by Kuhl and Miller (1982) showed that 6-month-old infants could accurately discriminate vowels produced by three different speakers. The infants did not attend to the voices, but focused on vowel identity instead. Similar results were reported by Holmberg, Morgan, and Kuhl (1977), who found that infants' discrimination of fricatives was not affected by speaker variability (however, see Carrell, Smith, & Pisoni, 1981). Finally, in a recent study, Jusczyk, Pisoni, and Mullennix (in press) examined the effects of speaker variability on infants' discrimination of CVC syllables. Jusczyk et al. found that infants could discriminate syllables, such as /bug/ and /dug/, as well in multiple-speaker conditions as in single-speaker conditions.

The data on the effects of speaker variability, however, are not completely unequivocal. The non-effects of speaker variability reviewed above came from tasks of perceptual identification or discrimination with few attentional or time constraints. Several more intensive experiments have shown reliable effects of speaker variability on speech perception and word recognition. In a paper entitled "The Case of the Unknown Talker," Creelman (1957) investigated the recognition of words that were presented in lists spoken by either 1, 2, 4, 8, or 16 speakers. Perceptual identification of these words in noise showed that lists produced by two or more speakers were recognized slightly less accurately (differences of 7% -- 10%) than single-speaker lists.

Later experiments with larger sets of stimulus materials have shown more substantial effects of speaker variability. Summerfield and Haggard (1973; see also Summerfield, 1975) observed slower reaction times to identify words in multiple-speaker lists than in single-speaker lists. More recently, Mullennix, Pisoni, and Martin (1989) investigated the effects of speaker variability on word recognition using a large sample of CVC monosyllables. Words were presented in lists spoken by either one speaker or by fifteen speakers; subjects performed either perceptual identification of words in noise or naming of non-degraded words. The authors found large and reliable effects of speaker variability; word recognition was slower and less accurate in multiple-speaker conditions than in single-speaker conditions. Moreover, speaker variability effects were more robust than word frequency and similarity neighborhood density effects (see Luce, 1986; Luce, Pisoni, & Goldinger, 1990). Finally, Mullennix et al. also found that speaker variability interacted with signal degradation, implying that noise and speaker variability affect a common, early stage of speech perception.

In a series of experiments investigating speeded classification of vowels, consonants, and words, Nusbaum and Morin (1992) also found that performance in multiple-speaker conditions was impaired, relative to single-speaker conditions. Also, the effects of speaker variability interacted with memory load; the difference in performance between single- and multiple-speaker conditions was larger under conditions of heavier load. They concluded that each token presented in a new voice invokes a capacity-demanding normalization process.

In a different type of experiment, Green, Kuhl, Meltzoff, and Stevens (1991) crossed female faces and male voices in a McGurk paradigm (McGurk & McDonald, 1976). In this procedure, a syllable is played over headphones while a videotape of a face is shown articulating a syllable. If the two modalities present incongruous information, an illusion is perceived. For example, if the spoken syllable /ba/ is presented with the visual syllable /ga/, the intermediate syllable /da/ is perceived. Green et al. found that, even when the genders of the voice and face are mismatched, the illusion persists. Green et al. suggested that the illusion remains because the speech signal is normalized early in phonetic processing. By their use of "normalization," Green et al. imply that the speech signal is reduced to an abstract label, reminiscent of Joos' early suggestions.

Speaker Variability and Attention

Further insight into speaker variability effects is provided by recent experiments in memory and attention. Martin, Mullennix, Pisoni, and Summers (1989) investigated serial recall of ten-word lists spoken by either 1 or 10 speakers. They found that recall for 10-speaker lists was impaired. Specifically, recall of multiple-speaker lists was less accurate than recall of single-speaker lists, but only for words in early list positions. Also, recall of visually-presented digits shown before the spoken lists was less accurate if the subsequent lists were multiple-speaker lists than if they were single-speaker lists. Martin et al. suggested that word lists produced by multiple speakers require more attention for rehearsal in working memory than lists produced by a single speaker.

Additional evidence of the attention-demanding nature of speaker variability was provided by Goldinger, Pisoni, and Logan (1991), who presented single- and multiple-speaker word lists at varying presentation rates for serial recall. Goldinger et al. found that speaker variability interacted with presentation rate whereas other stimulus variables, such as word frequency, did not. Figure 1.1 displays the serial recall data. At relatively fast presentation rates (upper panels), recall of single-speaker lists was superior to recall of multiple-speaker lists, as in the Martin et al. (1989) experiments. At slower presentation rates (lower panels), however, recall of multiple-speaker lists was actually more accurate than recall of single-speaker lists. The presentation rate is assumed to affect the rehearsal processes of the task (Murdock, 1962; Rundus, 1971); these results suggest that speaker variability taxes these attention-demanding stages of processing. A related finding was also reported by Lightfoot (1989). She found that subjects' familiarity with the speakers' voices further modified the differences in recall of single- and multiple-speaker lists. When subjects were trained to recognize the voices of the speakers and associate them with fictional names (i.e., Brad, Mary, Jane, Sam, etc.), the difference in recall between single- and multiple-speaker lists was reduced, in a manner similar to the Goldinger et al. (1991) results.

Insert Figure 1.1 about here

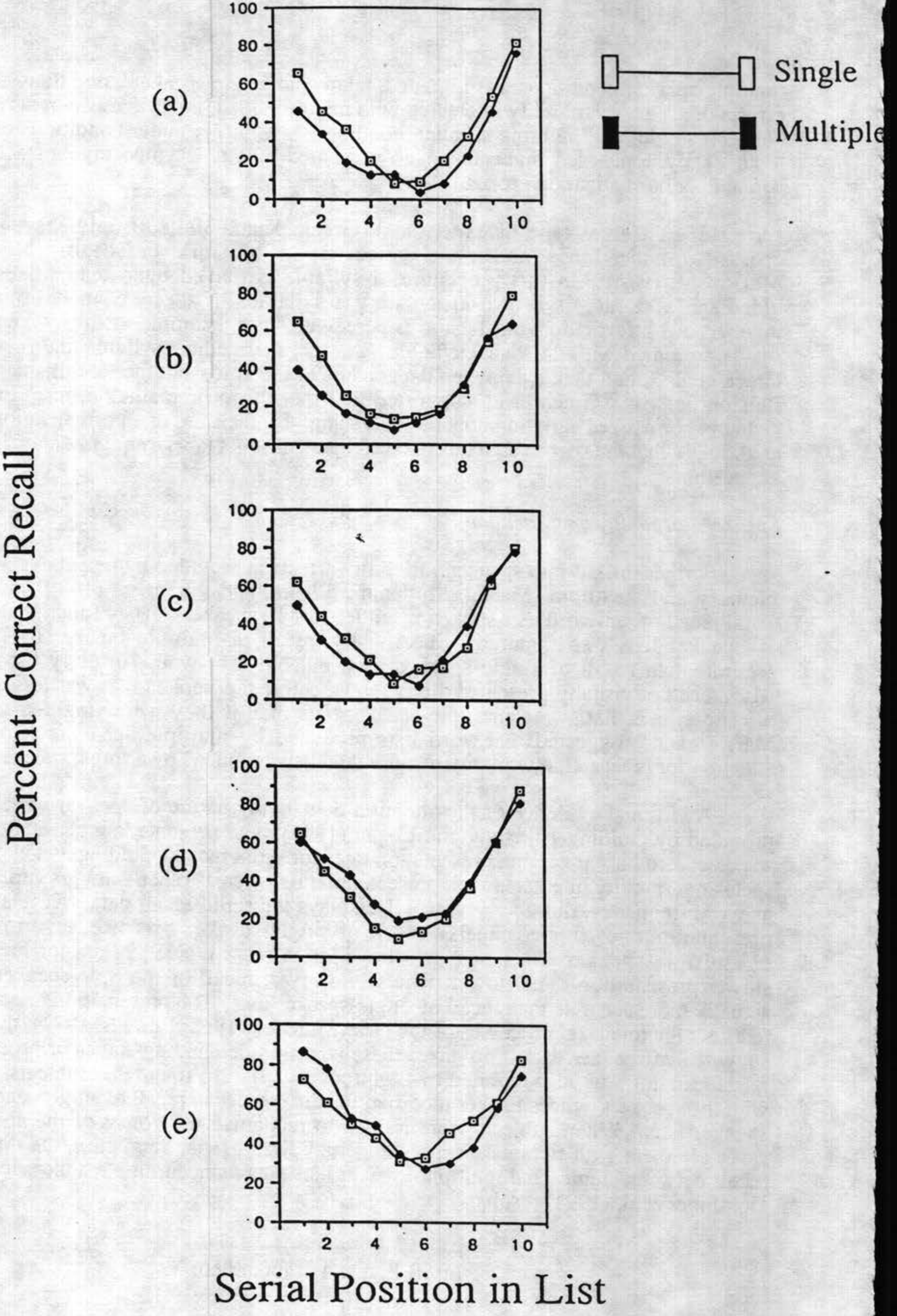


Figure 1.1 Serial-ordered recall data from Goldinger et al. (1991). Each panel displays recall of single- and multiple-speaker lists, with different inter-stimulus intervals. One word was presented every: a) 250 msec, b) 500 msec, c) 1000 msec, d) 2000 msec, e) 4000 msec.

The effects of speaker variability on memory and attention were investigated further by Jusczyk, Pisoni, and Mullennix (in press). Jusczyk et al. observed that infants recognize phonemic constancy very well despite variation of stimulus voices. However, Jusczyk et al. also employed a variation of the high-amplitude sucking (HAS) procedure (Eimas, Siqueland, Jusczyk, & Vigorito 1971) that included a two-minute delay period between the habituation to one syllable and the presentation of a new syllable. By this manipulation, Jusczyk et al. assessed the effects of speaker variability on infants' ability to encode and remember phonetic structure. Infants who heard a single speaker were able to detect a phonetic change across the two-minute delay, but infants who heard multiple speakers were not. Taken together with the adult data, these findings suggest that maintaining perceptual constancy across speakers requires attentional resources.

The consistent finding that voice information is retained alongside lexical information suggests that lexical information and voice information may be processed as integral dimensions of wholistic representations. If this were the case, then attending to the message level of an utterance would imply attention to the voice dimension as well. A demonstration of the integral nature of voice and phonetic information was provided by Mullennix and Pisoni (1990; see also Wood, 1974), who used the Garner (1974) speeded classification procedure to investigate processing dependencies between phonetic classification and speaker variability. Subjects classified monosyllabic words according to either initial phoneme (/b/ versus /p/) or voice (male versus female). They found that irrelevant variation could not be ignored; variation along either the phoneme or speaker dimension slowed classification along the other dimension. An asymmetry, however, showed that variability along the voice dimension impaired classification along the phonetic dimension more than vice versa. These data suggest that the processing of voice information and phonetic information are dependent on one another and that both processes share a limited-capacity cognitive system (see Cutting & Pisoni, 1978; Nusbaum & Morin, 1992). Mullennix and Pisoni suggest that voice and phonetic information are integral dimensions; attention to the meaning of spoken words implies attention to voice as well.

Memory for Voices

The majority findings summarized above suggest that speaker variability affects early stages of speech perception and spoken word recognition. It is not clear, however, that a normalization process is responsible for the effects. One could argue that speaker variability is salient and simply reduces performance by distracting subjects from their task. Whether normalization occurs in speech perception is not the issue at hand in the present investigation. Rather, the issue is the related claim that voice-specific information is not represented in long-term memory. In more recent literature, theories of normalization do not always imply that voice information is ignored or forgotten. For example, although Nusbaum and Morin (1992) argue that normalization is an essential process in speech understanding, they never claim that voice is ignored in further processing. Indeed, as in the research on memory for specific visual patterns and words, there is substantial evidence reviewed below that voices, as well as specific spoken exemplars, are stored in long-term memory and are accessible for long periods of time.

The human voice conveys information about a speaker's age and gender, as well as cultural information, such as the speaker's regional origin, temperament, and social group membership. These aspects of speech, known as *indexical information* (Abercrombie, 1967), do not relate directly to processes of phonetic perception, but are still relied upon in speech communication. Most of us are able to discriminate New York and Japanese accents, or the speech of children and adults. Also, indexical information alerts the listener to speaker identity and to important changes in the physical or emotional state of the speaker (Ladefoged & Broadbent, 1957, refer to these aspects of the voice as "personal information"). For example, the entire realm of vocal changes we call "tone of voice" are pervasive in conversation and are readily perceived as anger, depression, or joy. Tone of voice can even modify the semantic content of an utterance, such as in a sarcastic comment. These linguistic intuitions were confirmed by Geiselman and his colleagues, who demonstrated that listeners incidentally store detailed information about speakers' voices and implied connotative states (Geiselman & Bellezza, 1977; Geiselman & Crawley, 1983).

As reviewed earlier, humans display excellent long-term memory for faces. The natural analogy to this phenomenon is memory for voices. Common experience suggests that voice memory is also quite accurate; we recognize familiar voices over the telephone, etc. More formally, several experiments have examined the reliability of long-term memory for voices. In an early experiment, McGehee (1937) presented a spoken passage to 740 listeners. Voice recognition was tested at varying delays, ranging from 1 day to 5 months. In the recognition tests, listeners heard the original speaker and 4 new speakers reading the original passage. The listeners were instructed to select the original voice. Recognition accuracy was 83% after 1 day, 80% after 1 week, 57% after 1 month, 35% after 3 months, and 13% after 5 months. McGehee concluded that voice recognition is too inaccurate to be relied upon, especially in important situations such as court testimony.

Numerous researchers since McGehee's study have investigated the reliability of "earwitness testimony" (see Clifford, 1983; Tosi, 1979). Saslove and Yarney (1980) had subjects work on a cover task and overhear a taped conversation. The subjects heard a woman in the next room speak in an angry tone of voice for 11 seconds, then hang up. In a later recognition test, subjects heard the original speaker and four foils, all repeating the original message. Unlike McGehee's study, recognition was unaffected by a 24-hour delay. However, accuracy decreased when the recognition test was a surprise. Also, recognition performance decreased if the test sample of the original speaker was recorded in a normal tone of voice. Finally, subjects' accuracy of recognition and confidence in their recognition were only mildly correlated.

In another experiment on recognition of unfamiliar voices, Legge, Grossman, and Pieper (1984) played subjects recorded voice samples and tested recognition in a two-alternative forced-choice procedure. Accuracy was found to depend on the number of voices in the target set, the duration of the speech samples, and the presence or absence of faces corresponding to voices. Of particular interest, however, was the effect of delay between study and test. Like Saslove and Yarney (1980), Legge et al. (1984) found no significant effect of delay; recognition was equally accurate after either 15 minutes or 10 days.

Most of the research summarized above indicates that long-term memory for voices is rather poor. McGehee (1937) observed relatively inaccurate recognition memory for unfamiliar voices. Pollack, Pickett, and Sumby (1954) examined the role of several factors in voice recognition, including the number of possible voices, the duration of speech samples, the frequency range of the voices, and the presentation of simultaneous voices. Of all the factors examined, duration of the speech sample was found to be most important, suggesting that voice recognition depends on adequate sampling. Stevens, Williams, Carbonell, and Woods (1968) later replicated this finding, and also found that accuracy of speaker identification varies widely across speakers.

Other experiments, however, suggest that recognition memory for voices may be more robust than McGehee's data suggest. Carterette and Barneby (1975) observed voice recognition at levels well above chance in an immediate recognition test. More recently, Papcun, Kreiman, and Davis (1989) conducted a very interesting investigation of long-term memory for unfamiliar voices. Papcun et al. had subjects rank 10 voices (all young male Californians) in terms of how difficult the subjects believed the voices would be to remember. After collecting these ratings, Papcun et al. selected three voices to serve as target in a recognition memory test. One voice was ranked easy to remember, one was ranked hard, and one was ranked intermediate. The remaining seven voices were used as foils. Subjects heard the voices in a study session and were later tested after either 1, 2, or 4 weeks. Hit rates declined slightly across delays and were unaffected by anticipated ease of recognition. False alarms, however, were strongly affected by anticipated ease of recognition: Subjects committed fewer false alarms to voices judged easier to recognize. These data suggest that the underlying perceptual similarities among the stimulus voices may be a key element in recognition performance.

Language familiarity is another important factor in voice recognition. Recently, Goggin, Thompson, Strube, and Simental (1991) found that monolingual English listeners identified bilingual English-German speakers voices more accurately when they spoke English than when they spoke German. Monolingual German listeners showed the opposite pattern for the same stimulus materials. These results were replicated with English and Spanish. When listeners were also bilingual, however, no effect of language on speaker recognition was observed. In a similar investigation, Goldstein, Knight, Bailis, and Conover (1981) found that recognition memory was lower for accented voices than for unaccented voices, although only when relatively brief speech samples were used.

It is not surprising that the familiarity of voices is a powerful determinant of voice recognition. Bricker and Pruzansky (1966; see also Clifford, 1983) found that listeners were 98% accurate in recognizing voices of their co-workers, a considerable improvement over the accuracy in McGehee's experiment. Hollien, Majewski, and Doherty (1982) found a similar percentage. Ladefoged and Ladefoged (1980) tested Peter Ladefoged's ability to recognize voices of his friends and relatives from brief recordings; foils were interspersed among the target voices. With a single word, Ladefoged recognized familiar speakers at 31% accuracy; accuracy improved to 66% with a sentence, and to 83% with a 30-second passage.

Van Lancker and her colleagues (Van Lancker, Kreiman, & Emmorey, 1985; Van Lancker, Kreiman, & Wickens 1985) recently investigated recognition of familiar voices under several conditions. Van Lancker, Kreiman, and Emmorey (1985) examined recognition of celebrities' voices. Famous voices were recognized quite well (from a response set of 6 possibilities), whether played forwards or backwards. In another experiment, Van Lancker, Kreiman, and Wickens (1985) again presented famous voices to listeners; this time voices were normal, rate-compressed, or rate expanded. As with the backward voices, recognition of rate-altered voices was quite accurate. In both experiments, however, it was found that the recognizability of voices under different manipulations is extremely speaker-dependent. Some speaker's voices are robustly recognized when played backwards, but not when rate-altered. The opposite pattern was found for other speakers' voices. From these findings, Van Lancker and her colleagues suggested that familiarity improves voice recognition by alerting the listener to the idiosyncratic cues that distinguish individuals by voice.

Clearly, listeners have reliable memory for voices, especially for familiar voices. Moreover, memory for voices affects performance in other domains: Lightfoot (1989) found improved serial recall of spoken word lists after training subjects to associate names with voices. In an unusual experiment, Kosslyn and Matt (1977) played speech samples of two speakers to subjects. One of the speakers spoke quickly and the other spoke slowly. Later, subjects were asked to read passages that were allegedly written by one of the two speakers. In several conditions, Kosslyn and Matt found that the speaking rate of the assumed writer affected the reading rate of the subjects.

Memory for Specific Spoken Words

The evidence summarized above demonstrates that, during speech perception, listeners attend to and remember information about voices. One goal of the present investigation was to assess the persistence of specific exemplars of spoken words in memory and to evaluate their role in lexical processing. The diagnostic dimension examined in the present experiments was the voice associated with each word in the stimulus set. Words were presented in study sessions and were repeated in test sessions; the voices of half of the repetitions were changed between study and test. The effects of voice change on implicit and explicit memory for words were assessed.

For the present investigation, therefore, the findings that voices are represented in long-term memory are reassuring: If listeners could not remember voices, there would be little reason to assume voices are encoded into episodic or lexical representations. A related line of research, however, is more directly relevant to the present research. The present investigation closely resemble several experiments described above in the discussion of memory for specific visual words. The studies reviewed above tested repetition effects for visual words in conditions of exact format repetition and format change (e.g., new typography). This strategy was applied here with respect to voices, rather than typographies. Several previous reports of memory for specific spoken exemplars provide an important background:

Several studies in the literature have assessed the effects of voice change on memory for spoken words, in much the same manner as experiments have assessed the effects of typography change on memory for printed words. Indeed several articles report analogous experiments in both modalities (e.g., Hintzman, Block, & Inskeep, 1972; Light, Stansbury, Rubin, & Linde, 1973). Hintzman et al. (1972) presented words spoken by a male and a female speaker during a study session; during a surprise recognition test, half of the words were repeated in the same voice as their original presentation and half were repeated in a different voice. As in their data on visual formats, Hintzman et al. found that subjects could accurately discriminate old from new voices on repetitions. In a similar experiment, Light et al. (1973) essentially replicated the procedure used by Hintzman et al., but they presented entire sentences rather than words. Like Hintzman et al., Light et al. found that subjects could easily discriminate old-voice from new-voice repetitions.

Other early investigations on the persistence of voice memory were conducted by Cole, Coltheart, and Allard (1974) and by Allard and Henderson (1976). Cole et al. (1974) investigated "same-different" (AX) response latencies to spoken letter names. Subjects heard two successive letter names and indicated as quickly and as accurately as possible if the same name was spoken twice or if two different names were spoken. A male and a female speaker recorded the stimulus materials. Half of the trials presented the same voice twice and half presented both voices, one for the "A" stimulus and the other for the "X" stimulus. Regardless of the letter names to be compared, the results were consistent: Responses to both "same" and "different" trials were faster when both words were spoken in the same voice than when each word was spoken in a different voice. These results were replicated even when an 8-second interval separated the "A" and the

"X" words. Cole et al. concluded that voice information is encoded into memory and lasts for several seconds.

In another AX experiment, Allard and Henderson (1976) compared AX response times in trials of voice constancy, trials of intonation change within a single voice, and trials of voice change. A consistent pattern of results was observed across inter-stimulus delays of 0.5 seconds, 2 seconds, and 6 seconds: Responses to same-voice trials were faster than responses to same-voice/different-intonation trials, which were faster than different-voice trials. After six sessions of testing, however, the effects of voice disappeared. Again, Allard and Henderson concluded that voice information persists in memory, at least for brief periods.

A great deal of recent research on memory for words and voices comes from Geiselman and his colleagues. In several experiments, Geiselman investigated the incidental retention of voice information when subjects were instructed to remember words. Geiselman and Bellezza (1976) presented 20 sentences to subjects for free recall. The sentences were spoken by a male and a female speaker (half each); also, the sentences emanated from a speaker on either side of the room (half each). Subjects were either instructed to attend to voice and location information, or these factors were tested by surprise. Location information was only retained under explicit instructions, and attention to location decreased recall of sentences. Voice information, however, was retained under intentional or incidental conditions, and attending to voice did not affect sentence recall. In another study, Geiselman and Belleza (1977) found that the gender of the speaker was retained for sentences that had a neutral agent; when sentences already had a male or female agent, voice was not recalled. In later research, Geiselman (1979) found that voice information may not be encoded or retained when it interferes with other cognitive operations in sentence processing, such as encoding the gender of the agent.

In a slightly different experiment, Geiselman and Glenny (1977) investigated the role of voice imagery in recognition memory for spoken words. Subjects listened to recordings of a male and a female speaker. Afterward, subjects were shown pairs of words visually and were instructed to silently rehearse the words either in their own voice, the voice of the male speaker, or the voice of the female speaker. A surprise recognition test presented half of the words in each of the stimulus voices: Words were more accurately recognized when the test voice matched the rehearsal voice. No effect of test voice was observed when the subjects rehearsed in their own voice. This finding is related to the result reported by Kosslyn and Matt (1977), who found that speaking rates of imagined writers affected the reading rates of subjects.

From all their data, Geiselman and his colleagues (Geiselman, 1979; Geiselman & Belleza, 1976; 1977; Geiselman & Crawley, 1983) proposed a *voice connotation hypothesis,* wherein the voice of a speaker is assumed to modify the connotation of a sentence. Voice retention is expected when the connotation is subject to change; when a sentence already contains a strong gender-specific connotation, no retention of voice is observed. Unlike a more analog or episodic theory, the voice connotation hypothesis assumes voice is retained via semantic interpretation, rather than via some perceptual representation system (Schacter, 1990). This assumption was tested in a recent study by Palmeri, Goldinger, and Pisoni (1992), and was further evaluated in the present research.

Palmeri, Goldinger, and Pisoni (1992) recently investigated memory for specific words and voices using a continuous recognition procedure (cf. Shepard & Teghtsoonian, 1964). The study was a replication and extension of an earlier experiment by Craik and Kirsner (1974). Craik and Kirsner presented spoken words to subjects for continuous recognition (see Kirsner, 1973). Words were first spoken either by a male or female speaker; when repeated later, half of the words retained their original voice and half were switched. In one experiment, subjects only responded "old" or "new" to words, irrespective of voice. In another experiment, subjects were also asked to judge whether "old" words were repeated in their original voice or in a new voice. Across experiments, Craik and Kirsner found that same-voice repetitions were more accurately recognized than different-voice repetitions. Moreover, this same-voice benefit was observed from lags of 1 or 2 intervening items to lags of 32 intervening items. Voice recognition for correctly recognized words was also accurate at levels well above chance. Craik and Kirsner concluded that specific perceptual details of spoken words persist in memory for at least several minutes.

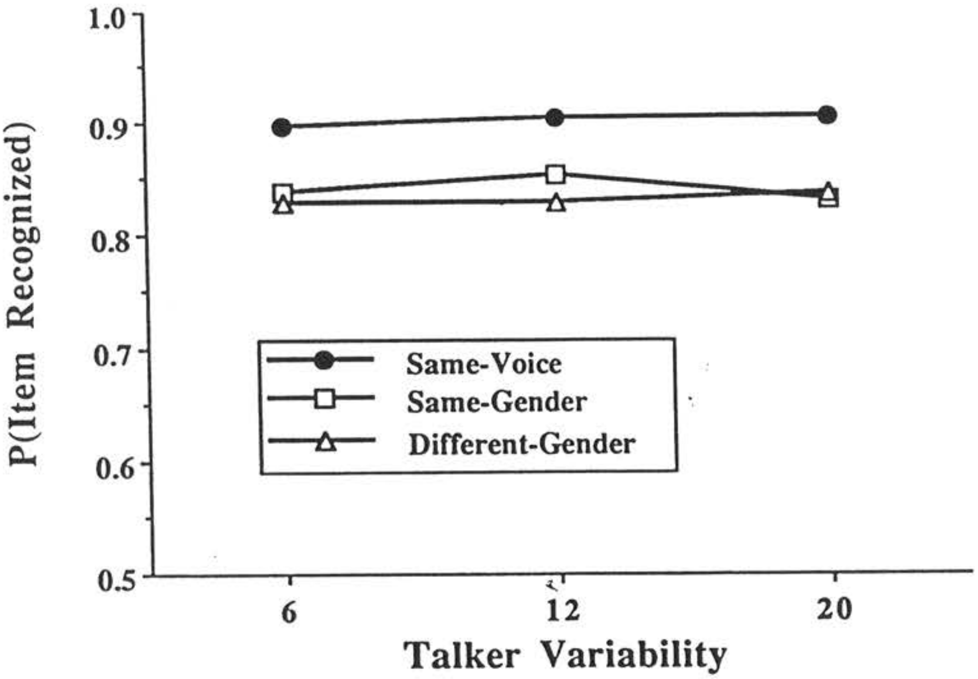
The Palmeri et al. (1992) study extended the experiments of Craik and Kirsner (1974) in several ways. Like Craik and Kirsner, we first examined word recognition alone, followed by joint recognition of words and voices. Unlike Craik and Kirsner, however, we introduced several levels of speaker variability in our stimulus lists; subjects heard lists containing 2, 6, 12, or 20 voices (a single-voice control condition was examined as well). In all multiple-voice lists, half of the speakers were male and half were female. Also, whereas Craik and Kirsner examined retention up to a lag of 32 intervening trials, we included lags up to 64 intervening trials.

The use of more than two voices permitted us to evaluate several aspects of this procedure that a two-voice, male/female list would not allow. First, we were able to evaluate the effects of increasing variability on the retention of voice information. If voices were encoded strategically in Craik and Kirsner's experiments, increasing the number of speakers from 2 to 20 should impair subjects' ability to process and retain voice information. Moreover, the use of multiple speakers of both genders allowed us to evaluate Geiselman's voice connotation hypothesis. By this view, female and male voices invoke different connotations of the same stimulus. Therefore, recognition should be gender-dependent rather than voice-dependent: Same-voice repetitions and different-voice/same-gender repetitions should yield equivalent recognition, and both should be superior to different-voice/different-gender repetitions.

Figure 1.2 displays results, in terms of recognition memory for words, from Palmeri et al., Experiment 1. The upper panel shows results as a function of test voice and number of talkers, averaged across lags. The lower panel shows results as a function of test voice and lags, averaged across number of talkers. In both panels, recognition is shown for same-voice, different-voice/same-gender, and different-voice/different-gender repetitions. (Only data from the 6-, 12-, and 20-voice lists are shown, because 2-voice lists do not include different-voice/different-gender repetitions. The results, in terms of same- versus different-voice repetitions, in the two-voice condition replicated the conditions shown in the figure.)

Insert Figure 1.2 about here

The upper panel of Figure 1.2 shows that recognition accuracy, and effects of test voice, were equivalent across levels of variability. Increasing the number of speakers had no appreciable effect on either word or voice encoding. This suggests that voice encoding is automatic, rather than strategic or capacity-demanding; there is no cost for increasing speaker variability.



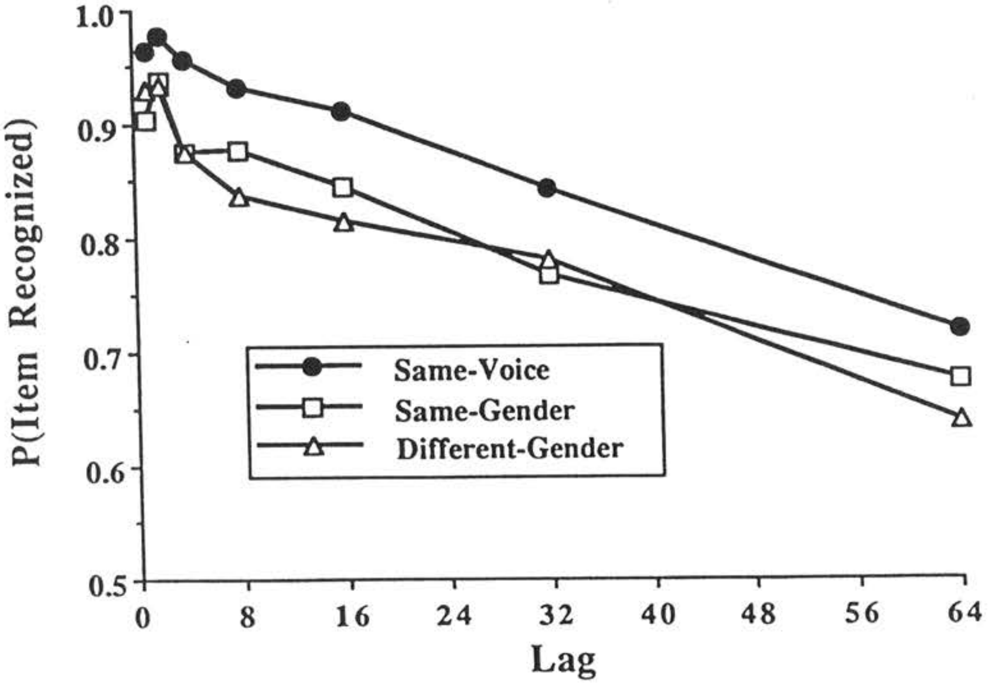


Figure 1.2 Continuous recognition memory data from Palmeri et al. (1992). Both panels display recognition of same-voice, different-voice/same-gender, and different-voice/different-gender repetitions. The upper panel displays results of 6-, 12- and 20-voice conditions, collapsed across lag; the lower panel displays results at all lags, collapsed across number of voices.

Both panels of Figure 1.2 show that recognition of same-voice repetitions was superior to recognition of different-voice repetitions, regardless of gender. This pattern was observed at all levels of speaker variability and across all values of lag, up to 64 intervening trials. The fact that recognition of different-voice trials was equivalent across gender conditions implies that an exact word/voice representation is encoded during perception; only repetition of this exact stimulus facilitates recognition. Thus, Geiselman's voice connotation hypothesis was not supported. Finally, the recognition advantage for same-voice repetitions was equivalent across lags, even lags twice as long as those used by Craik and Kirsner (1974). This suggests that word and voice information may form a coherent episodic trace in memory that may last indefinitely. Palmeri at al. (1992) concluded that spoken words create highly specific and long-lasting episodic traces. Moreover, these episodic traces seem to contain analog information, rather than propositional, gender-specific information.

Schacter and Church (in press) recently reported another study that examined memory for specific spoken exemplars. This study is of particular interest to the present investigation because, although both studies were designed independently, several of the same issues were examined. Schacter and Church examined the role of voice in implicit and explicit memory for spoken words across several levels of processing.

Schacter and Church examined implicit and explicit memory for words as a function of encoding task used at study, and as a function of repetition voice presented at test. The major goal of their investigation was to evaluate the role of a Perceptual Representational System (PRS) in auditory repetition effects (Schacter, 1990; Tulving & Schacter, 1990). By the PRS hypothesis, repetition effects occur in a presemantic representational system that is sensitive to surface features of stimuli, irrespective of their meanings or associations to other stimuli. Accordingly, data that support the PRS hypothesis are repetition effects observed without semantic or elaborative processing, and sensitivity to changes of surface features of stimuli between study and test. Schacter and Church sought to evaluate the PRS hypothesis with respect to memory for spoken words.

In one experiment, subjects performed either a semantic (pleasantness rating) or a nonsemantic task (pitch judgment of voice) during a study phase. During test, subjects performed an auditory stem completion task to syllables presented in the clear. After the implicit, stem completion task, subjects performed a cued recall task. The voice of the speaker was changed between study and test for half of the stimuli. Significant effects of voice were observed in stem completion, but not in cued recall. Reliable repetition effects were observed in stem completion for both same- and different voice repetitions, but the effect was larger for same-voice repetitions.

In another experiment, the encoding tasks were changed to ensure that the effects of voice were not an artifact of study conditions. During the study phase, some subjects performed a meaning rating task, in which they estimated the number of meanings for words. Other subjects performed a clarity rating task, in which they assessed how clearly each word was spoken. Results showed no difference in magnitude of repetition effects between the semantic and nonsemantic encoding tasks. Also, significant effects of voice were again observed to an equivalent degree in both encoding conditions. Schacter and Church concluded that implicit repetition effects of spoken words can be dissociated from explicit memory, as with visual words. Moreover, they argue that the data are consistent with the PRS hypothesis: Repetition effects were insensitive to level of processing at study. Repetition effects also displayed high sensitivity to voice differences between study and test.

Naturally, as was true for studies of memory for specific visual details, several researchers have reported null findings for effects of specific voice details of spoken words. For example, Jackson and Morton (1984) had subjects listen to spoken nouns and classify each along a living/non-living dimension. Afterward, subjects heard the same words presented in white noise for a perceptual identification test; half of the voices during test were preserved from the study session and half were changed. Jackson and Morton observed equivalent repetition effects for same-voice and different-voice trials and they concluded that abstract logogens mediate perception and repetition effects, and that these logogens are indifferent to the specific forms of spoken words.

Jackson and Morton's null findings of voice influence on repetition effects must be considered with caution. Null results can, of course, occur for a variety of reasons; abstract logogens are only one possible explanation. Taken together, the majority of evidence suggests that voice information is encoded as an integral aspect of the memory for a spoken word. As Pisoni (1990) points out, consideration of these data is a serious challenge for contemporary theories of speech perception, lexical representation, and lexical access.

Overview of the Present Investigation

A majority of the findings reviewed above suggest that specific perceptual episodes may persist in long-term memory and subserve later perception. One goal of the present investigation was to assess the persistence and specificity of voice information in long-term lexical memory. The present set of experiments assessed long-term, incidental memory for specific words spoken by specific voices. The objectives of the research were twofold: The primary goal was to evaluate the hypothesis that spoken word recognition entails access to lexical representations consisting of multiple perceptual episodes (cf. Kolers, 1976a; Kolers and Ostry, 1974; Jacoby, 1983a; 1983b; Jacoby & Brooks, 1984). The secondary goal was to assess the strong assumptions of speaker normalization that have characterized many theories of speech perception and spoken word recognition for over 40 years.

Memory for specific spoken exemplars was assessed via explicit and implicit memory measures. In both experiments, as in the studies summarized above, the major independent variable was the change or constancy of voices for each word presented between study and test. The present investigation consisted of three stages:

Perceptual Scaling

Preparation and scaling of stimulus materials was the first step in this investigation. After ten speakers were selected and all stimulus words were prepared, multidimensional scaling techniques were applied to determine the perceptual similarity relations among the voices. Deriving the similarity space for the voices created an important tool for later data analysis. In nearly every study conducted to assess the persistence of specific perceptual information in recognition memory, only two categories of stimuli are presented during the test phase. Same-format versus different-format words are typically presented; "format" denoting typography, voice, or some other physical attribute. In the present investigation, different-voice repetitions were analyzed with more precision by application of the scaling solution: Results were evaluated at a level of detail beyond simple "same voice" versus "different voice" (or "same gender" versus "different gender") comparisons.

Experiment I

After collecting the scaling data, two experiments were conducted. Experiment I examined implicit and explicit memory as a function of voice, number of voices, and delay between study and test. The implicit memory conditions consisted of successive sessions of perceptual identification of spoken words in white noise. Subjects identified the same set of words in two sessions separated by either a 5-minute, 1-day, or 1-week delay period. Some subjects heard only 2 voices in the stimulus set, others heard 6 or 10 voices. As in the research reviewed above, although the words were held constant across sessions, the voices of half the words were changed. In all cases, the new voices were evenly distributed across different-voice repetition trials. Repetition effects, in terms of identification accuracy between sessions, were examined.

The explicit memory conditions of Experiment I employed all the same procedural manipulations; only the study and test measures varied. During study, subjects merely identified spoken words presented in the clear. During test, subjects received a surprise recognition memory test. The predictions for Experiment I were derived directly from previous research: Following Schacter and Church (in press), effects of voice were expected in both implicit and explicit memory. However, voice information was expected to affect implicit memory over longer delays than explicit memory (Tulving, Schacter, & Stark, 1982). Following Geiselman and Belleza (1976), voice was expected to be encoded automatically. Therefore, the effects of voice were not expected to interact with the total number of speakers in the set. Finally, following previous research on recognition and categorization, the perceptual similarity between the study and test exemplars of each word was expected to affect the magnitude of repetition effects.

Experiment II

Experiment II examined implicit and explicit memory as a function of voice and levels of processing. All subjects in Experiment II heard 6 voices in the stimulus sets presented at study and test. Also, all subjects participated in study and test sessions separated by a 5-minute delay. During the study phase, subjects performed one of three speeded classification tasks to words presented in the clear. The difficulty of the classification, and therefore the level of processing, was varied across groups of subjects. Some subjects classified words according to the gender of the speakers, others classified words according to initial phonemes, and others classified words according to syntactic classes.

As in Experiment I, all the same words were presented in each session of Experiment II, but the voices of half the words were changed across sessions. In the implicit memory conditions, subjects performed their assigned classification task twice in a row; the difference in performance between sessions for the same- and different-voice trials were the measures of interest. In the explicit memory conditions, subjects performed their assigned classification task during study, then received a surprise recognition test. Again, accuracy of recognition for same- and different-voice repetitions were the measures of interest. As in Experiment I, the predictions for Experiment II were derived directly from previous research: Again, following Schacter and Church (in press), effects of voice were expected in both implicit and explicit memory. However, voice was expected to affect implicit memory equivalently across all levels of processing. In contrast, voice was expected to affect explicit memory to different degrees across levels of processing. Finally, as in Experiment I, the perceptual similarity between the study and test exemplars of each word was expected to affect the magnitude of their respective repetition effects.

CHAPTER II: Stimulus Materials and Perceptual Scaling

The overall purpose of the present investigation was to assess the role of specific physical details, in this case voice information, in the memory representation of spoken words. The representation of voice information in memory of spoken words was assessed via indirect, perceptual measures, and also via direct recognition memory measures. As the previous literature has shown (e.g., Jacoby & Hayman, 1987, Musen & Treisman, 1990), the perceptual similarity between study and test items is of critical importance, especially in studies of implicit memory. However, in almost all previous research using changes of typeface, picture orientation, or voice, stimuli presented in study and test sessions are denoted as either "same format" or "different format." Any change of the physical details between study and test is denoted "different format," with little care given to the magnitude of the perceptual differences involved.

In categorization research, Nosofsky (1984, 1988a, 1988b) has combined measures of the precise similarity relations among stimulus materials with models of choice behavior. Nosofsky's Generalized Context Model (GCM) combines the context model of classification (Medin & Schaffer, 1978) with measures of perceptual similarity among stimulus items. Using these perceptual measures to estimate the similarity parameters of the model, Nosofsky has repeatedly observed remarkably tight fits of the model to actual classification data. More recently, Nosofsky (1991) has recently discussed the relation of the classification model to recognition memory models. Among other aspects, the importance of the perceptual similarity of a probe to exemplars in memory is emphasized. Other researchers have stressed the importance of perceptual similarity between probes and exemplars in recognition memory (indeed, the accommodation of this important relation is a staple of all current models of recognition memory). For example, the SAM model of recognition (Gillund & Shiffrin, 1984) and Hintzman's MINERVA II model (Hintzman, 1986; 1988) both assume the importance of specific episodic traces in matching probes to memory.

Given the theoretical and empirical importance of perceptual similarity among study and test items in recognition memory, the present research sought to examine the effects of voice similarity in close detail. In the first experiment (Chapter III), the total number of voices in the stimulus set was varied. Subjects heard either two, six, or ten speakers in the stimulus sets presented at study and test. In the two-voice conditions, the effects of voice preservation on recognition could only be assessed in terms of "same" versus "different" voice, as in previous studies (e.g., Craik & Kirsner, 1974; Jackson & Morton, 1984). In the six-voice and ten-voice conditions, however, more precise comparisons were possible: "Different-voice" trials were sorted into separate categories defined by the speakers of the study and test tokens. In this manner, the role of perceptual similarity on recognition of different-voice tokens could be assessed with greater precision than previous studies allowed.

The first step toward this research objective was to select the stimulus materials to be used in the experiments and discover the similarity relations among the voices. Ten speakers were recruited, five males and five females. All speakers produced tokens of the same 300 words. Recordings of eight of the speakers were obtained from an existing database; two new speakers were recorded specifically for use in the present research. Details concerning the speakers, the recorded vocabulary, and the stimulus preparation procedures are presented below. The materials described in this chapter were used in all subsequent experiments presented in later chapters.

Once all the tokens from all the speakers were recorded, digitized, and edited, a "same-different" (AX) experiment was conducted to measure perceptual similarities among the voices. In this procedure, subjects are presented pairs of spoken words and they must indicate as quickly and accurately as possible whether the same word was spoken twice ("same") or if two different words were presented ("different"). In the present experiment, half of the trials were "same-word" trials and half were "different-word" trials. Every trial presented two different speakers; subjects were instructed to ignore speaker differences and respond only on the basis of word identity. Latency of correct responses was the primary dependent measure.

The critical data from the AX experiment came from the "same" trials, of which there were 225 per group. Across all groups of subjects, the "same" trials comprised inter-speaker comparisons to create an entire similarity matrix. Since ten speakers were included, there were 45 possible combinations of speakers, which composed the 45 cells of the eventual similarity matrix. Every subject received each of these 45 possible pairings of speaker 5 times in the "same" trials of the experiment. Across all groups of subjects, all of 225 words were presented in each possible cell of the matrix. In this manner, all cells juxtaposing different speakers contained identical words. These precautions ensured that the differences between cells of the matrix were due only to speaker differences, not due to differences in the words themselves. Because this kind of control could never be maintained in the "different" trials (the number of possible combinations of 225 "different" words multiplied by 45 cells of speakers reaches the high millions), only data from the "same" trials were included in the construction of the similarity matrix.

After the entire experiment was conducted, the complete similarity matrix was submitted to KYST, a multidimensional scaling program (see Kruskal, Young, & Sherry, 1973). The multidimensional scaling solution provided estimated "perceptual distances" between speakers. These estimated distances were used in the remainder of the experiments in this investigation. (For a thorough review of the theory and method of multidimensional scaling, see Kruskal & Wish, 1978; Shepard, 1980; Carroll & Arabie, 1980; Young, 1984).

The idea to use an AX procedure to collect similarity or scaling data does not originate from the present research; the method has been used quite extensively. Nickerson (1967; 1972; see also Bamber, 1969; Felfoldy, 1974) studied the AX task in detail to discover the nature of sequential comparisons; the number of dimensions on which stimuli differed was found to strongly affect response time. In research more directly relevant to the present investigation, Chananie and Tikofsky (1969) demonstrated that AX response time in a speech discrimination task closely reflect the number of contrastive features between pairs of words (following Miller & Nicely, 1955). And, in a similar investigation, McInish and Tikofsky (1969) found that the number of contrastive features between consonants in sequential CV syllables strongly predicted AX response latencies.

With respect to the use of an AX procedure to collect multidimensional scaling data, Weiner and Singh (1974) conducted a scaling study based on the earlier research by Chananie and Tikofsky (1969). Chananie and Tikofsky related AX performance to the distinctive phonetic features described by Miller and Nicely (1955). In contrast, Weiner and Singh collected AX data to phonetic contrasts, derived a multidimensional scaling solution from their data, and discovered a match between the resultant dimensions and the Miller and Nicely classification scheme. Weiner and Singh suggested that the AX task provides a good measure of similarity between spoken items. In visual research, Podgorny and Garner (1979) studied the relations between two measures of similarity.

They used only the letters of the alphabet as their stimulus materials. They found that the similarity relations among the letters derived by scaling direct similarity ratings closely match the similarity relations found by scaling AX reaction times.

In more recent research, Sergent and Takane (1987) also found close relations between scaling solutions derived from a speeded AX procedure and scaling solutions derived from an unspeeded variant of the task. Moreover, they observed that scaling solutions derived from AX latency data resemble scaling solutions derived from unspeeded data for integral- as well as separable-dimension stimuli. This latter finding is particularly germane to the present research: Mullennix and Pisoni (1990) used the Garner (1974) speeded-classification paradigm to study the perception of spoken words that varied in speaker's voice and initial phoneme. The patterns of dimensional interference suggested that the voice and the phonetic content of a spoken utterance are perceived in an integral fashion, such that variations along either dimension interferes with classification along the other dimension. Given the integral relation of voice to word, Sergent and Takane's findings are pertinent to the present research, validating the AX task as an appropriate measure of similarity.

EXPERIMENT

Method

Subjects

One-hundred and eighty-three students enrolled in introductory psychology courses at Indiana University served as subjects. Subjects received partial course credit for their participation. All subjects were native speakers of English and reported no history of any speech or hearing disorders at the time of testing.

Stimulus Materials

The stimulus materials for the present investigation consisted of a vocabulary of 300 monosyllabic English words; all words were recorded by ten different speakers. Recordings of eight of the speakers were obtained from an existing database of spoken monosyllabic words recorded by twenty different talkers. These materials had been previously recorded, digitized, and tested for intelligibility. In addition to these materials, two other speakers recorded the same vocabulary as the other eight speakers. These two speakers recorded the entire vocabulary twice, for purposes described below. Two different vocabulary sources were used to compile a list of 300 unique monosyllabic words: The majority of the words came from the vocabulary of the Modified Rhyme Test (House, Williams, Hecker, & Kryter, 1965). However, because the MRT vocabulary only contains 272 unique words (28 words are repeated), an additional 28 words were selected from phonetically-balanced word lists (Egan, 1948).

Once the words had been selected, digitized files containing tokens of each word were obtained from the database. Tokens from eight speakers, four males and four females, were obtained. In addition, another male and female speaker were recorded. All of the stimuli in the existing database were originally recorded on audio tape in a sound-attenuated booth using an Ampex AG500 tape deck and an Electro-Voice D054 microphone. All words were spoken in isolation. The stimuli were then low-pass filtered at 4.8 kHz and digitized at a sampling rate of 10 kHz using a 12-bit analog-to-digital converter. All words were excised from the lists on audio tape using a digitally controlled speech waveform editor (WAVES) on a PDP 11/34 computer (Luce and Carrell, 1981). The mean RMS amplitude of all tokens was equated using a digital signal processing package. The words spoken by the two new speakers were recorded and prepared following the same procedures. Also, just as the existing tokens were previously tested for intelligibility, all tokens from the two new speakers were identified in the clear by ten subjects in a pilot test. Tokens that were not correctly identified by nine of the ten subjects were replaced by more intelligible tokens.

Procedure

Subjects were tested in groups of six or fewer in a quiet testing room used for speech perception experiments. Stimuli were presented over matched and calibrated TDH-39 headphones at 75 dB (SPL). A PDP 11/34 computer was used to present the stimuli and to control the experimental procedure in real-time. The digitized stimuli were reproduced using a 12-bit digital-to-analog converter and were low-passed filtered at 4.8 kHz.

Each trial of the experiment began with illumination of a cuelight situated at the top of a two-button response box. The cuelight was illuminated for 500 msec to alert subjects that a trial was about to begin. After the cuelight turned off, a 500 msec silent interval elapsed. After the interval, two words were spoken in succession, separated by a 100 msec inter-stimulus interval. The subjects indicated as quickly and as accurately as possible whether the same word was spoken twice in a row (irrespective of voice changes) or if two different words were spoken. For all subjects, the right-hand button of the response box corresponded to the "same" response and the left-hand button corresponded to the "different" response. Responses and latencies were recorded from the onset of the second word (the X stimulus) in each pair. After presentation of each pair of words, the computer waited for all subjects to respond. If 5000 msec elapsed before all subjects entered responses, the computer entered an incorrect response for the remaining subjects and a new trial began.

The experiment consisted of 450 trials; half were "same" trials and half were "different" trials. Subjects were allowed to rest for several minutes halfway through the experiment.

Results

Each group of subjects in this experiment received a different set of stimulus items. All the words were identical across groups, and all ten speakers were represented equally in all lists. The exact pairings of words and speakers was varied between groups. Every list presented all 45 possible combinations of speakers 5 times each in the "same" trials; 45 groups of subjects participated. Across groups, all combinations of speakers were compared speaking all 225 words used in the "same" trials.

After the data was collected, the mean latencies of correct "same" responses were calculated for each of the 45 cells in the design. The resultant similarity matrix is shown in Figure 2.1. The labels along the upper and left-hand sides of the matrix refer to the speakers: F1 through F5 denote the five female speakers; M1 through M5 denote the five male speakers. The shaded area represents the diagonal entries, which are not necessary for multidimensional scaling and were not included in the design, and the lower half of the matrix, which is redundant with the upper half. Only the upper off-diagonal entries are required by the scaling procedures. Because there were 183 subjects in the experiment, each value in the matrix represents approximately 915 observations. Errorful trials were not included in the calculation of means. Errors only occurred on 1.35 percent of all trials in the experiment.

	F1	F2	F3	F4	F5	M1	M2	M3	M4	M5
F1		716	794	694	670	769	874	852	896	822
F2			777	728	709	756	867	746	746	712
F3				719	719	812	789	714	733	882
F4					702	665	695	696	743	783
F5						748	796	887	833	858
M1							706	759	738	756
M2								733	719	748
M3									707	714
M4										700
M5										

Figure 2.1 Same-different (AX) response latency matrix. Each entry represents mean response latency for correct "same word" responses. Only speakers varied across cells; all words were presented in all cells.

Insert Figure 2.1 about here

The values shown in the similarity matrix are the mean response times for subjects to correctly respond "same word" to each combination of speakers. Higher numbers mean that more time was taken to respond same; the longer the response time, the more dissimilar the voices. To convert the numbers from dissimilarities to similarities, the inverse of each value may be computed, although the scaling procedure does not require this transformation.

The similarity matrix shown in Figure 2.1 was used as the input to KYST, a well-known multidimensional scaling program (Kruskal, Young, & Sherry, 1973). Non-metric scaling was conducted on the similarity data. Kruskal and Wish (1978) describe several guidelines for selecting the proper number of dimensions for any given data set. One important rule of thumb is that the number of dimensions multiplied by four should not exceed the number of objects being scaled. Therefore, because the entire stimulus array consisted of only ten voices, only a two-dimensional solution was derived. The scaling program was run several times with different starting configurations and random seeds, and always produced a consistent outcome of stress values and spatial patterns.

Insert Figure 2.2 about here

The two-dimensional representation obtained from the scaling solution is shown in Figure 2.2. Each of the circles represents a speaker, denoted by the same labels as the matrix. As Figure 2.2 shows, the ten speakers were spread over a considerable range of the two-dimensional space. For the purposes of the present investigation, the exact nature of the emergent dimensions is not important. Nevertheless, the dimensions appear to conform to two easily identified properties of the stimuli: The horizontal axis corresponds to the speaker's gender; males tend to fall to the right of the midpoint and females to the left (with the exception of F5).

Identification of the vertical axis required more careful analysis. All of the words from each speaker were analyzed on a VAX 3500 microcomputer using ILS, a software package that performs acoustic analyses on digitized signals. Only gross analyses were performed, yielding estimates of each speaker's average pitch, vowel space, and formant variance. The measure that most closely corresponded to the vertical axis of Figure 2.2 was the pitch of the speaker's voice, conditionalized by speaker gender. Among both the male and female speakers, the increases along the vertical axis correspond monotonically with increases in vocal pitch. The vertical axis does not seem to correspond to *absolute* pitch, however, since almost all the male speakers should have values lower than the female speakers. Since speakers' positions along this axis are gender-dependent, the axis is denoted *relative pitch*. The emergence of relative pitch as a primary dimension, rather than absolute pitch, implies that listeners' perception of "high" and "low" voices is contingent upon gender. Such perceptual judgments are apparently made with reference to the proper class of comparison speakers.

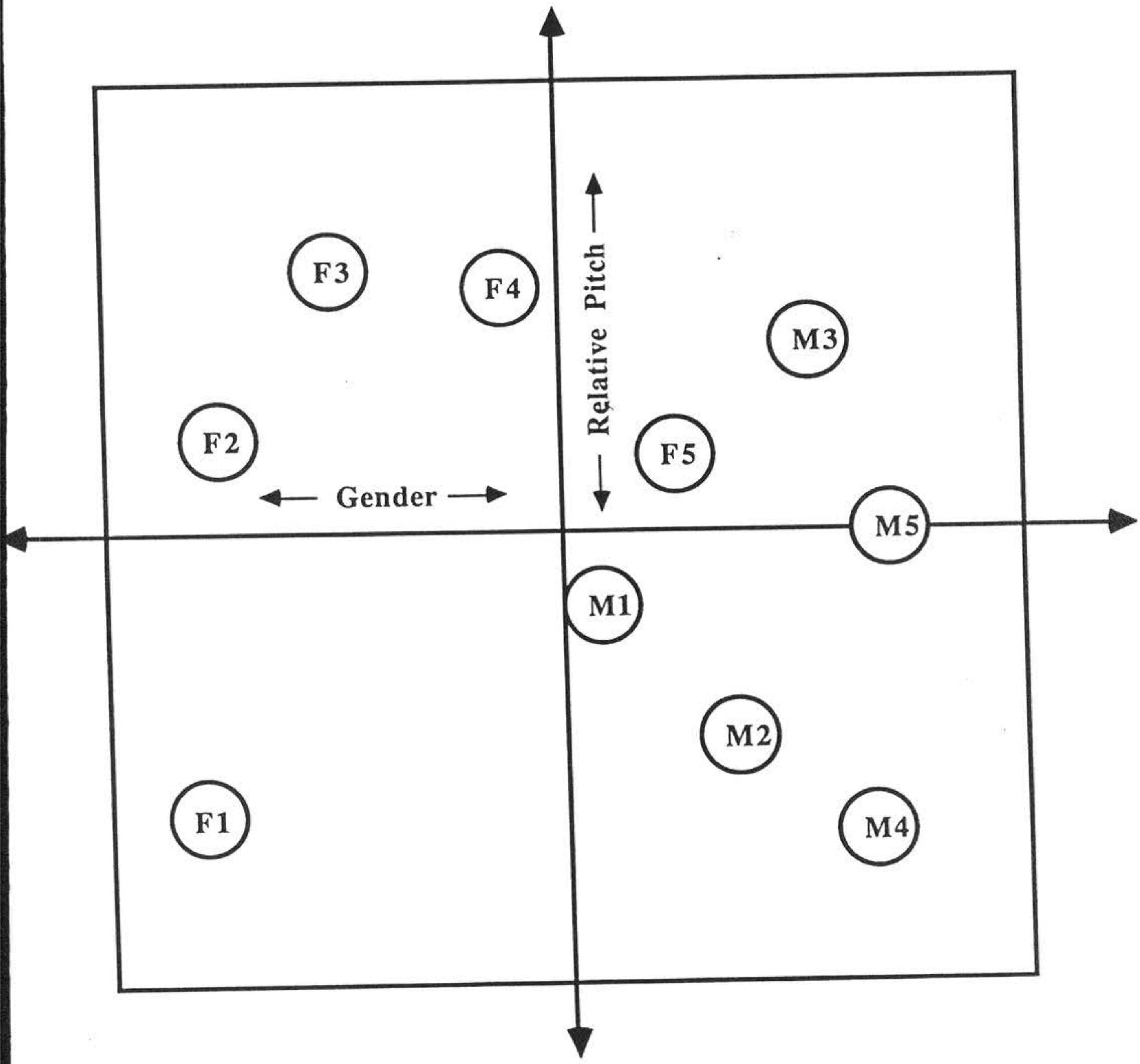


Figure 2.2 Two-dimensional perceptual space derived from response latency matrix. Male and female speakers are represented by symbols M1-M5 and F1-F5, respectively. The horizontal axis corresponds to gender; the vertical axis corresponds to relative pitch (see text).

Discussion

The experiment described in this chapter was conducted to develop a tool for later data analysis. After the scaling solution was derived from the AX data, the coordinates for all speakers in the two-dimensional space were used to estimate "perceptual distances" between all pairs of speakers. In the remaining experiments, the role of perceptual similarity among voices in word memory was assessed. The estimated distances from the scaling were used in the analyses of these remaining experiments.

CHAPTER III: Implicit and Explicit Memory: Effects of Retention Interval

The goal of the present investigation was to assess the role of specific perceptual details-- in the present case voice-- in memory for spoken words. Experiment I examined implicit and explicit memory for specific tokens across a variable number of speakers and across variable delays. The critical variable in all conditions of the experiment was voice: Words from a study session were repeated in a test session; each word was repeated either in the same voice as its original presentation, or in a different voice, using perceptual identification and recognition memory procedures.

The implicit memory conditions consisted of successive sessions of perceptual identification of spoken words in white noise. Subjects identified the same set of words in two sessions separated by either a 5-minute, 1-day, or 1-week delay period. Some subjects heard 2 voices in the stimulus set, others heard 6 or 10 voices. Although the same words were presented across sessions, the voices of half the words were changed. Repetition effects, in terms of the difference in identification accuracy between sessions, were examined.

The explicit memory conditions employed the same procedural manipulations as the implicit memory conditions; only the tasks during study and test varied. During study, subjects identified spoken words presented in the clear. During test, subjects received a surprise recognition memory test. The recognition test assessed memory for words, with no direct assessment of memory for voices.

The predictions for Experiment I follow from previous research: Following Schacter and Church (in press), effects of voice were expected in both implicit and explicit memory. However, voice was expected to affect implicit memory over longer delays than explicit memory (Parkin, Reid, & Russo, 1990; Tulving, Schacter, & Stark, 1982). Following Geiselman and Belleza (1976), voice was expected to be encoded automatically, so effects of voice were not expected to interact with the number of speakers in the set (see also Palmeri et al., 1992). Finally, following research on recognition and categorization, the perceptual similarity between the study and test exemplars of words was expected to affect the magnitude of their respective repetition effects.

EXPERIMENT

Method

Subjects

Three-hundred and sixty students served as subjects. All subjects participated in two sessions of the experiment. Two-hundred and twenty-six of the subjects were enrolled in introductory psychology courses at Indiana University and received partial course credit for their participation. The remaining 134 subjects were paid volunteers who received $5.00 for each session of participation. All subjects were native speakers of English and reported no history of speech or hearing disorders at the time of testing.

Half of the subjects participated in the implicit test (perceptual identification), and half participated in the explicit test (recognition memory). Within either condition, subjects were further divided into sub-conditions, as described below, yielding a total of 20 subjects per condition.

Stimulus Materials

The stimulus materials consisted of 300 words recorded by ten speakers, as described in the previous chapter. Two of the speakers recorded two tokens of each word. In the 2-speaker conditions, therefore, different tokens of all words (both same-voice and different-voice repetitions) were presented across sessions. In the 6- and 10-voice conditions, same-voice repetitions in the test sessions always included the exact tokens presented in the study sessions. Referring to the perceptual space shown in Figure 2.2, the speakers in the 2-voice conditions were F1 and M1; the speakers in the 6-voice conditions were F1, F2, F3, M1, M2, and M3.

Design

The experiment consisted of two major conditions: the perceptual identification task and the explicit recognition memory task. Each of these major conditions contained several sub-conditions as described below:

Both the perceptual identification and the explicit recognition tasks consisted of nine completely between-subjects sub-conditions, with 20 subjects per sub-condition. The nine sub-conditions were created by combining three levels of talker variability with three levels of delay interval. In both the study and test sessions, subjects received stimulus sets containing tokens recorded by either 2, 6, or 10 speakers. Between the study and test sessions, subjects waited for either a 5-minute, 1-day, or 1-week delay period to elapse.

Procedure

Perceptual Identification Task

In the perceptual identification task, the procedures were identical across the study and test sessions: Subjects were tested in groups of five or fewer in a quiet testing room used for speech perception experiments. Each subject was seated in an individual booth equipped with a pair of matched and calibrated TDH-39 headphones and an ADM computer terminal with keyboard. A PDP 11/34 computer was used to present the stimuli and to collect responses.

Each trial of the experiment began with a warning prompt on the ADM terminal screen. The phrase "GET READY FOR NEXT TRIAL" appeared in the center of the screen and remained for 750 msec. Five-hundred msec after the offset of the warning prompt, continuous, band-limited white noise was presented at 70 dB (SPL). Fifty msec after the onset of the noise, a stimulus word was presented at 75 dB (SPL), yielding a +5 dB signal-to-noise ratio. Fifty msec after the offset of the word, the noise was discontinued. The digitized stimulus words were reproduced using a 12-bit digital-to-analog converter and were low-passed filtered at 4.8 kHz.

After stimulus presentation, subjects had up to 20 seconds to identify the spoken word and type their response on the ADM keyboard. If all subjects responded before the end of the 20-second period, a 1000 msec inter-trial interval elapsed, and a new trial was initiated. Both the study and test sessions consisted of 300 randomly ordered trials.

Explicit Recognition Memory Task

In the explicit recognition memory task, different procedures were used in the study and test sessions. In both sessions, however, subjects were tested in groups of five or fewer using the same apparatus described above.

In the study session, subjects listened to words presented in the clear and they simply identified each word. As in perceptual identification, each trial of the experiment began with a warning prompt on the ADM terminal screen. The phrase "GET READY FOR NEXT TRIAL" appeared on the center of the screen and remained for 750 msec. Five-hundred msec after the offset of the warning prompt, a stimulus word was presented at 75 dB (SPL). The digitized stimulus words were reproduced using a 12-bit digital-to-analog converter and were low-passed filtered at 4.8 kHz.

After stimulus presentation, subjects had up to 20 seconds to identify the spoken word and type their response on the ADM keyboard. If all subjects responded before the end of the 20-second period, a 1000 msec inter-trial interval elapsed, and a new trial was initiated. The study sessions consisted of 150 randomly ordered trials.

In the test session, subjects again listed to words presented in the clear under the same listening conditions used in the study session. In the test session, however, subjects typed "old" or "new" in response to each word, rather than typing the word itself. After stimulus presentation, subjects had up to 20 seconds to render their recognition judgment and type "old" or "new" on the ADM keyboard. If all subjects responded before the end of the 20-second period, a 1000 msec inter-trial interval elapsed, and a new trial was initiated. The study sessions consisted of 300 randomly ordered trials, half of which were "new" words and half of which were "old" words.

Results

Results are presented in three sections. First, the overall results of implicit versus explicit memory are compared. Second, same-voice and different-voice trials are compared across memory tasks, levels of talker variability, and delay period. Third, the effects of perceptual similarity between study and test voices are assessed across memory tasks, levels of talker variability, and delay period.

Overall Performance

For perceptual identification, the percentage of words correctly identified was calculated for each subject. For a response to be considered correct, the entire response had to match the target item exactly or be a homophone (e.g., *there, their*). Responses were corrected for simple spelling or typing errors, such as letter transpositions, prior to analysis. For recognition memory, the percentages of hits, misses, false alarms, and correct rejections were calculated for each subject. These values were used to derive a d' and β for each subject. Finally, d' and β were used to derive a *P(C)max* for each subject, providing an estimate of percent correct performance from the recognition memory data (see Blosser, 1965).

Figure 3.1 displays overall performance as a function of memory task, level of talker variability, and delay period. The upper panel displays results, in terms of percent correct, of the test sessions of the perceptual identification task. The lower panel displays results, in terms of *P(C)max,* of the test sessions of the explicit recognition memory task.

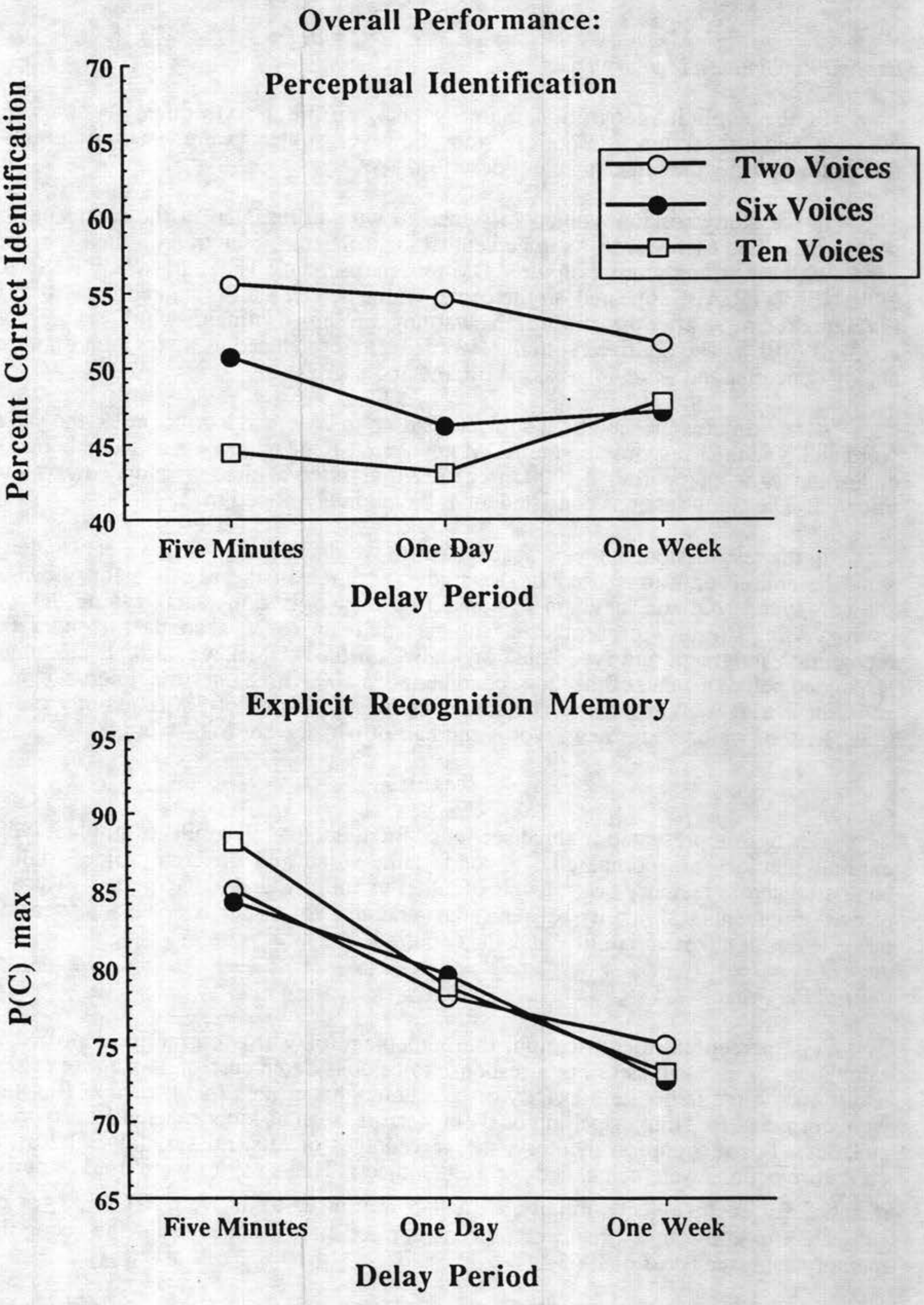


Figure 3.1 Overall performance in perceptual identification and recognition memory as a function of delay and number of voices. The upper panel displays perceptual identification results; the lower panel displays recognition memory results.

Both panels display results as a function of the total number of voices in the stimulus set and the delay period between study and test.

Insert Figure 3.1 about here

The basic pattern of results shown in Figure 3.1 is easily summarized: In the perceptual identification task (upper panel of Figure 3.1), accuracy was highest in the 2-voice condition and was progressively lower in the 6- and 10-voice conditions. No effects of delay were observed. In the explicit recognition memory task (lower panel of Figure 3.1), no effects of number of voices was observed, but there were strong effects of delay.

The data from the perceptual identification task were analyzed in a 3 X 3 (Number of Voices X Delay) ANOVA, treating all variables as completely between-subjects. In this analysis, a significant main effect of Number of Voices was observed [F (2,171) = 19.47, $MS_e = 0.85$, $p < .0001$], reflecting the generally decreased accuracy at increased levels of talker variability. The main effect of Delay did not approach significance [F (2,171) = 1.02, $MS_e = 0.85$, n.s.]. The mean accuracy did not change across delays. Despite the smaller effect of Number of Voices at the 1-week delay, the two-way interaction of Number of Voices and Delay was not significant either [F (2,171) = 1.60, $MS_e = 0.85$, n.s.]. The data from the explicit recognition task were also analyzed in a 3 X 3 (Number of Voices X Delay) ANOVA, treating all variables as completely between-subjects. The main effect of Number of Voices was not significant [F (2,171) = 0.91, $MS_e = 1.10$, n.s.], reflecting the almost identical levels of performance across levels of talker variability. A significant main effect of Delay was observed, however [F (2,171) = 44.78, $MS_e = 1.10$, $p < .0001$], reflecting the general decrease in accuracy at increasing delays. The two-way interaction of Number of Voices and Delay did not approach significance [F (2,171) = 0.04, $MS_e = 1.10$, n.s.].

Effects of Voice

The purpose of the present experiment was to examine the role of voice in memory for spoken words. In both the implicit and explicit memory conditions, words were presented at study and were later repeated at test. Half of the repeated words were presented in the same voices as their original presentations; half were repeated in different voices as their original presentations. This section presents analyses of repetition effects as a function of voice. For the moment, "different voice" trials are treated as a uniform category of data. The precise role of voice differences in memory is assessed in the following section on perceptual similarity effects.

The results from the perceptual identification task are presented first, followed by results from the explicit recognition memory task:

Perceptual Identification Task

The percentage of words correctly identified was calculated for each subject in each session. Figure 3.2 displays the percentages of correctly identified words in all conditions of perceptual identification. Four categories of data are shown in the figure: In each set of four bars, two bars represent results from the study session and two bars represent results from the test session. The dark striped bars represent words in the study

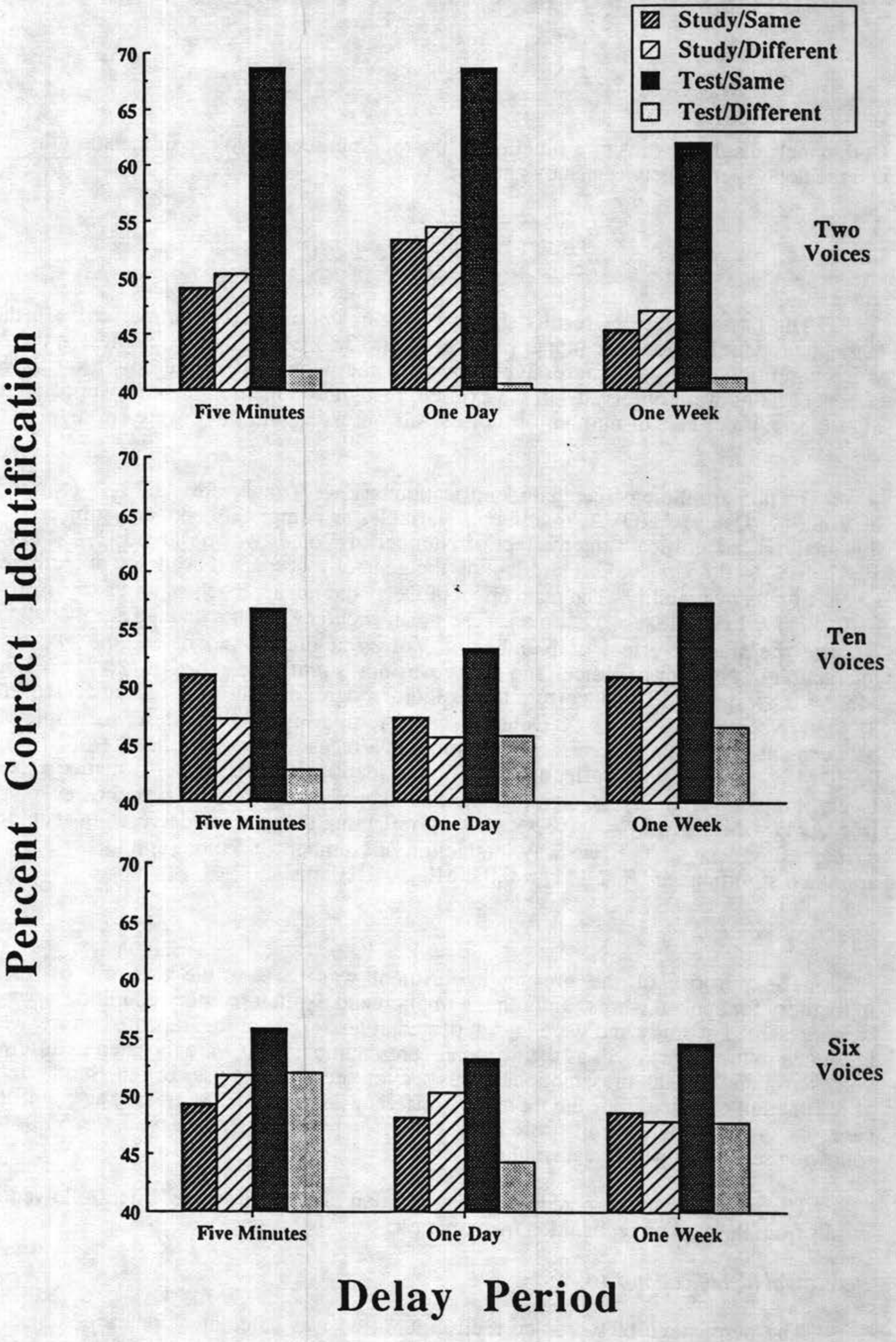


Figure 3.2 Percent correct perceptual identification as a function of number of voices, session, delay between sessions, and repetition voice.

session that were eventually repeated in the same voice during the test session; light striped bars represent words in the study session that were eventually repeated in a different voice during the test session. The solid black bars represent same-voice repetitions in the test session; the solid gray bars represent different-voice repetitions in the test session.

The upper panel of Figure 3.2 displays results from the 2-voice conditions; the center panel displays results from the 6-voice conditions and the lower panel displays results from the 10-voice conditions. All three panels display results as a function of delay period between study and test.

Insert Figure 3.2 about here

The consistent finding in all conditions shown in Figure 3.2 is that "same-voice" repetitions were recognized more accurately than "different-voice repetitions. Because the effect of interest in the experiment was the role of voice in the repetition effect, a *net repetition score* was calculated for each subject. Analyzing the data in terms of net repetition scores eliminates the irrelevant variance in the data due to the between-subjects design (see Schacter & Church, in press). The net repetition score for each subject was calculated by subtracting the percentage of correct responses in the study session from the percentage of correct responses in the test session. These scores were derived for both the same-voice and different-voice trials.

Figure 3.3 displays the results of the perceptual identification task as a function of voice, number of voices, and delay period. All the data are displayed in terms of net repetition effects, centered about zero. The net repetition scores were derived from the results shown in Figure 3.2. A positive net repetition score implies an improvement in performance in the test session; a negative net repetition score implies a decrement in performance in the test session. The upper panel displays results from the 2-voice conditions; the center panel displays results from the 6-voice conditions and the lower panel displays results from the 10-voice conditions. All three panels display results as a function of delay period between study and test.

Insert Figure 3.3 about here

As Figure 3.3 shows, "same-voice" repetitions uniformly yielded positive repetition effects; "different-voice" repetitions yielded either negative or negligible repetition effects. A 2 X 3 X 3 (Voice X Number of Voices X Delay) ANOVA was conducted on the means of the net repetition scores. A significant main effect of Voice was observed [F (1,171) = 72.25, MS_e = 2.67, $p <$ .0001]. Same-voice repetitions produced positive repetition effects at all levels of talker variability and across all delays. Different-voice repetitions produced either negative repetition effects or no repetition effects at all. Tukey's HSD analyses were conducted to compare same-voice and different-voice repetitions at all levels of talker variability and delay. In all nine comparisons shown in Figure 3.2, the same- and different-voice repetitions were significantly different ($p <$ .01).

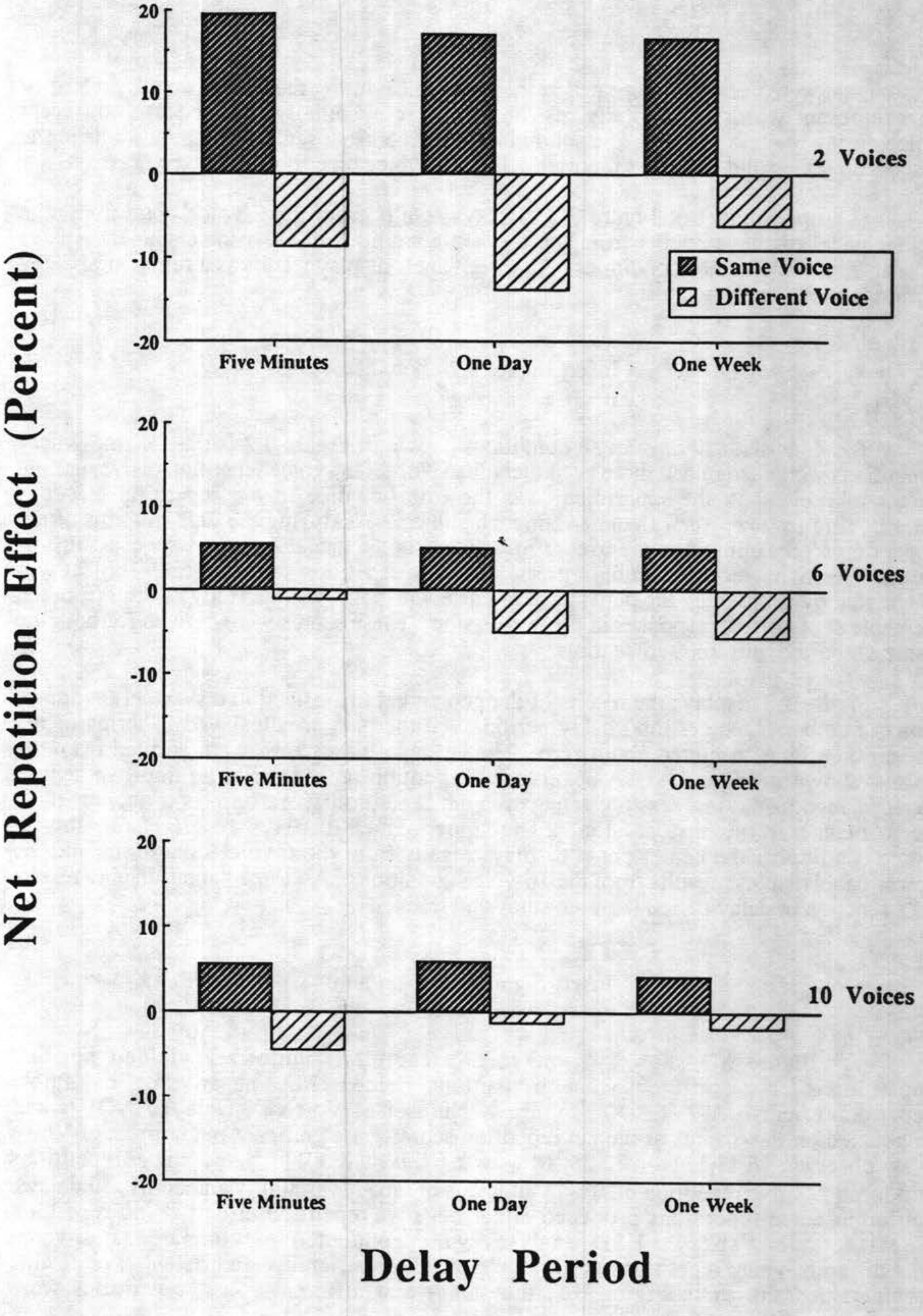


Figure 3.3 Net repetition effects in perceptual identification as a function of delay between sessions and repetition voice.

There were no significant main effects of Number of Voices [*F* (1,171) = 1.88, MS_e = 2.28, n.s.] or Delay [*F* (1,171) = 0.43, MS_e = 2.28, n.s.]. However, a significant interaction of Voice X Number of Voices was observed [*F* (1,171) = 14.87, MS_e = 2.67, $p < .0001$], reflecting the larger effects of Voice in the 2-voice condition than in the 6- or 10-voice condition. This was the only significant interaction observed in the analysis.

Explicit Recognition Memory Task

Hit rates for same- and different-voice repetitions were calculated for each subject in every condition. False alarms cannot be analyzed in terms of repetition voice (since they are responses to new words), so only hit rates were analyzed.

Figure 3.4 displays the hit rates in the explicit recognition memory task as a function of voice, number of voices, and delay period. The upper panel displays results from the 2-voice condition, the center panel displays results from the 6-voice condition, and the lower panel displays results from the 10-voice condition. All three panels present the results as a function of delay period between study and test.

Insert Figure 3.4 about here

As Figure 3.4 shows, at short delays between sessions, "same-voice" repetitions were recognized more accurately than "different-voice" repetitions. This difference diminished, however, across longer delays. A 2 X 3 X 3 (Voice X Number of Voices X Delay) ANOVA was conducted on the mean hit rates. A significant main effect of Voice was observed [*F* (1,171) = 18.59, MS_e = 1.12, $p < .0001$]. Same-voice repetitions produced generally higher hit rates at all levels of talker variability. Tukey's HSD analyses were conducted to compare same-voice and different-voice repetitions at all levels of talker variability and delay. Same- and different-voice repetitions were significantly different in all of the 5-minute delay conditions, the 2-voice/1-day condition, and the 10-voice/1-day condition ($p < .01$). The differences between same- and different-voice repetitions were not significant at the 1-week delay.

In addition to the main effect and specific effects of Voice, a significant main effect of Delay was observed [*F* (1,171) = 21.31, MS_e = 1.05, $p < .0001$], reflecting the consistent decrease in hit rates at increased delays. Moreover, a significant two-way interaction of Voice X Delay was observed [*F* (1,171) = 11.61, MS_e = 1.05, $p < .0001$], reflecting the decrease of the Voice effect at increased delays. The main effect of Number of Voices did not approach significance [*F* (1,171) = 1.65, MS_e = 1.05, n.s.], nor did Number of Voice interact significantly with any other factors.

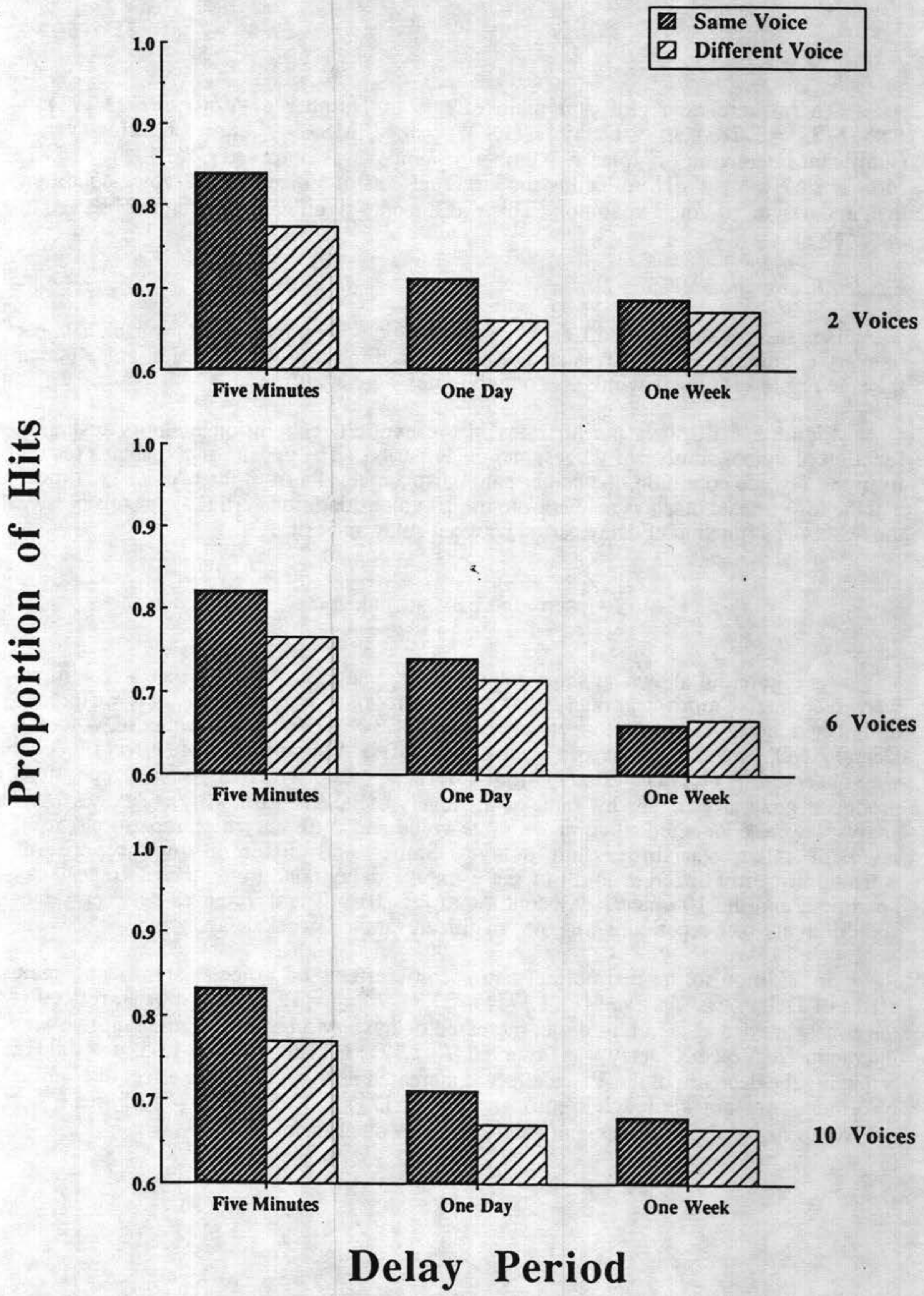


Figure 3.4 Hit rates in explicit recognition memory as a function of delay between sessions and repetition voice.

Effects of Perceptual Similarity

After examination of overall effects of voice, the implicit and explicit memory conditions were analyzed in terms of perceptual similarity between study and test exemplars. These analyses provided a more precise assessment of the role of voice in implicit and explicit memory.

The analyses were conducted as follows: All data derived from different-voice repetitions were divided into categories corresponding to all possible combinations of speakers. For the 6-voice conditions, there were 15 unique combinations of two speakers; for the 10-voice conditions, there were 45 unique combinations of two speakers. (Recall that perceptual distances for all possible combinations of speakers were derived from the earlier scaling solution.) In both the 6- and 10-voice conditions, all categories contained an equal number of trials. For the perceptual identification task, the net repetition score in each category for each subject was computed. For the explicit recognition task, the hit rate in each category for each subject was computed. Finally, the net repetition scores (or hit rates) in each category for each subject were correlated with the perceptual distances corresponding to each category.

The correlations of repetition effects and perceptual distance are presented first for the 6-voice conditions, then for the 10-voice conditions:

Figure 3.5 displays correlations between repetition effects and perceptual distance for the 6-voice conditions. The upper panel displays results from the implicit, perceptual identification task; the lower panel displays results from the explicit recognition memory task. Both panels display results as a function of delay period between study and test: Open circles represent the 5-minute delay; closed circles represent the 1-day delay; open squares represent the 1-week delay. The best-fitting lines through all data sets are drawn and labeled on the figure.

Insert Figure 3.5 about here

As Figure 3.5 shows, a consistent negative correlation was observed between repetition effects and perceptual distance in perceptual identification (upper panel). The strength of the correlations were equivalent across delays. In the explicit recognition memory task (lower panel), similar negative correlations were observed, but the strength of the correlations diminished at longer delays.

In the implicit test (upper panel of Figure 3.5), the slopes of the correlations between the net repetition scores and perceptual distances were very similar and were significant across all levels of delay. The correlations were as follows: The correlation for the 5-minute condition was $r = -.568$ [$F(1,298) = 34.60, p < .0001$]; the correlation for the 1-day condition was $r = -.597$ [$F(1,298) = 43.59, p < .0001$]; and the correlation for the 1-week condition was $r = -.564$ [$F(1,298) = 33.63, p < .0001$].

In the explicit test (lower panel of Figure 3.5), the slopes of the correlations between hit rates and perceptual distance decreased across levels of delay. The correlations were as follows: The correlation for the 5-minute condition was $r = -.540$ [$F(1,298) = 30.07, p < .0001$]. The correlation for the 1-day condition was $r = -.396$ [$F(1,298) = 8.99, p < .0001$]. Finally, the correlation for the 1-week condition was $r = -.089$ [$F(1,298) = 2.04$, n.s.].

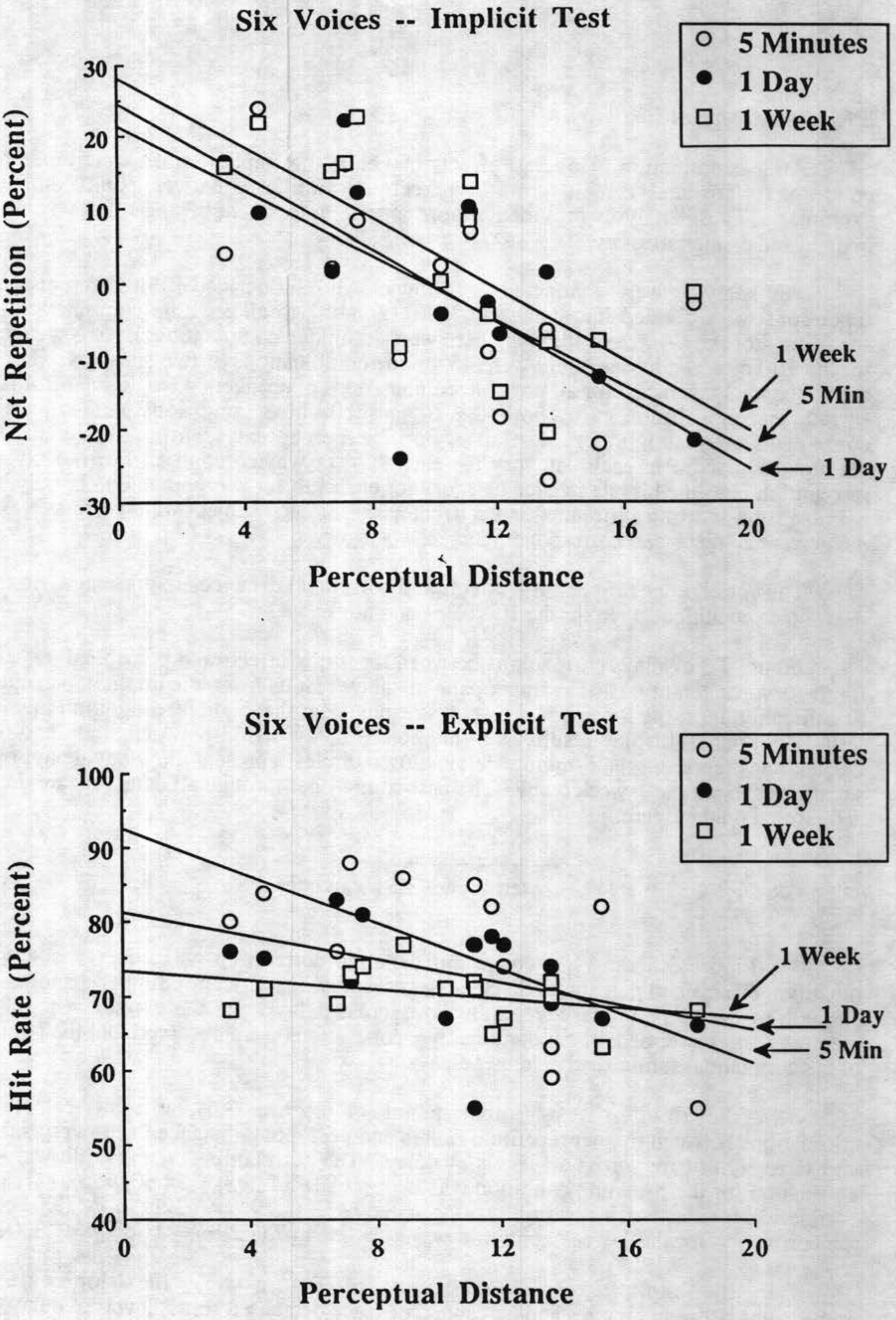


Figure 3.5 Correlations of perceptual similarity and "different-voice" repetition effects for the 6-voice conditions. The upper panel displays net repetition effects from perceptual identification; the lower panel displays hit rates from recognition memory. Both panels display results as a function of perceptual distance and delay between sessions.

Comparison of the upper and lower panels of Figure 3.5 reveals that the length of the delay period had no effect on implicit memory, but a strong effect on explicit memory. To assess this statistically, the slopes for subjects in each condition were entered into separate 1-way (Delay) ANOVAs. In the implicit test (upper panel of Figure 3.5), the main effect of Delay was not significant [$F\ (1,59) = 1.71$, $MS_e = 0.43$, n.s.]. In the explicit test (lower panel of Figure 3.5), however, a significant main effect of Delay was observed [$F\ (1,59) = 121.21$, $MS_e = 0.11$, $p < .0001$]. These results confirm that the delay period modulated the effects of perceptual similarity only in the explicit memory conditions.

Figure 3.6 displays the correlations between the repetition effect and perceptual distance for the 10-voice conditions. The upper panel shows the results from the implicit, perceptual identification task; the lower panel displays results from the explicit recognition memory task. Both panels display results as a function of delay period between study and test: Open circles represent the 5-minute delay; closed circles represent the 1-day delay; open squares represent the 1-week delay. The best-fitting lines through all data sets are drawn and labeled on the figure.

Insert Figure 3.6 about here

The results shown in Figure 3.6 closely resemble the results shown in Figure 3.5. A consistent negative correlation was again observed between repetition effects and perceptual distance in perceptual identification (upper panel). The strength of the correlations were equivalent across delays. In the explicit recognition memory task (lower panel), similar negative correlations were observed, but the strength of the correlations diminished at longer delays.

In the implicit test (upper panel of Figure 3.6), the slopes of the correlations between net repetition effect and perceptual distance were again similar across levels of delay. The correlations were as follows: The correlation for the 5-minute condition was $r = -.448$ [$F\ (1,898) = 37.85$, $p < .0001$]; the correlation for the 1-day condition was $r = -.462$ [$F\ (1,898) = 42.79$, $p < .0001$]; and the correlation for the 1-week condition was $r = -.485$ [$F\ (1,898) = 52.30$, $p < .0001$].

In the explicit recognition memory task (lower panel of Figure 3.6), the slopes of the correlations between hit rates and perceptual distance decreased across levels of delay. This pattern closely resembles the pattern found in the 6-voice conditions. The correlations were as follows: The correlation for the 5-minute condition was $r = -.576$ [$F\ (1,898) = 79.31$, $p < .0001$]; the correlation for the 1-day condition was $r = -.437$ [$F\ (1,898) = 16.01$, $p < .0001$]; and the correlation for the 1-week condition was $r = -.063$ [$F\ (1,898) = 1.10$, n.s.].

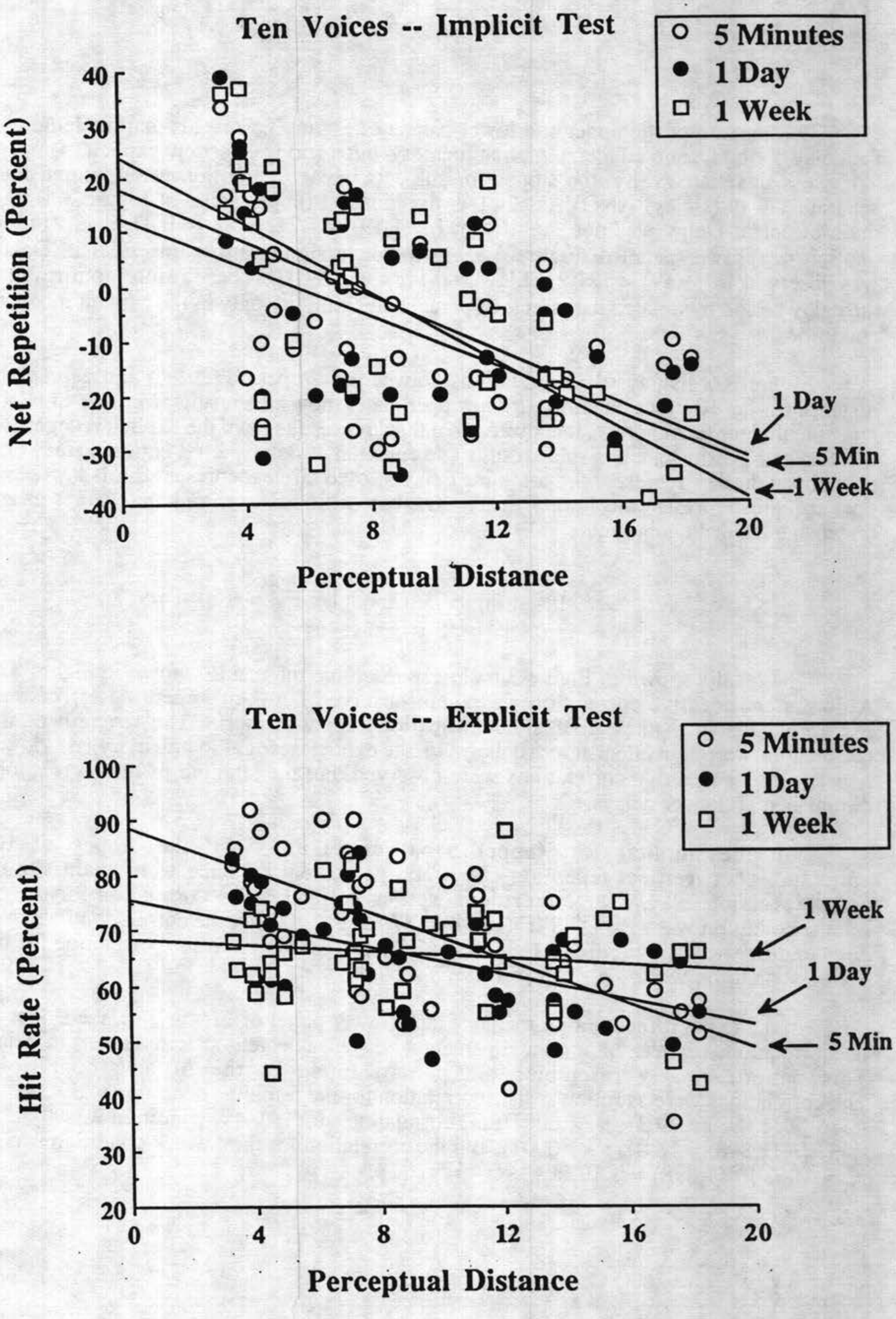


Figure 3.6 Correlations of perceptual similarity and "different-voice" repetition effects for the 10-voice conditions. The upper panel displays net repetition effects from perceptual identification; the lower panel displays hit rates from recognition memory. Both panels display results as a function of perceptual distance and delay between sessions.

As in the 6-voice conditions, comparison of the upper and lower panels of Figure 3.6 shows that the length of the delay period had no effect on implicit memory, but a strong effect on explicit memory. As before, the slopes for each subject were entered into separate 1-way (Delay) ANOVAs. In the implicit test (upper panel of Figure 3.6), the main effect of Delay only approached significance [F (1,59) = 2.89, $\mathrm{MS_e}$ = 1.13, n.s.]. In the explicit test (lower panel of Figure 3.6), however, a significant main effect of Delay was again observed [F (1,59) = 44.01, $\mathrm{MS_e}$ = 0.88, $p <$ .0001]. Thus, as in the 6-voice conditions, delay only interacted with perceptual similarity effects in the explicit recognition memory task.

Discussion

Results are discussed in three sections. First, the overall results of implicit versus explicit memory are discussed. Second, the effects of voice are discussed. Third, the effects of perceptual similarity between study and test voices are discussed.

Overall Performance

The overall results of the implicit and explicit memory conditions can be summarized easily. In the implicit memory (perceptual identification) condition, a strong effect of talker variability was observed. Increasing the number of talkers decreased accuracy of perceptual identification. This result replicates previous research on talker variability in speech perception and spoken word recognition: Mullennix et al. (1989) found that perceptual identification of spoken words was more accurate in a single-talker condition than a multiple-talker condition. Although Mullennix et al. did not compare levels of talker variability, the present findings are clearly comparable to their findings. Moreover, Mullennix and Pisoni (1990) found that Garner interference increased as the number of speakers was increased. Again, their finding of impaired performance with increased speaker variability is analogous to the present findings.

In the explicit recognition memory task, increasing the number of talkers had no effect on performance. This result replicates the recently reported findings of Palmeri, Goldinger, and Pisoni (1992). As Palmeri et al. discussed, a signal-detection theory of recognition memory predicts this non-effect (see Banks, 1970; Parks, 1966). Signal detection theory assumes that subjects judge the previous occurrence of words by assessing their overall familiarity. More specifically, subjects must judge the familiarity of targets, relative to the familiarity of distractors. Assume that voice cues contribute to familiarity: In a stimulus set with two voices, any presented word is similar (along the voice dimension) to half of the total set. This means that targets will seem quite familiar, but so will distractors. In a stimulus set with ten voices, any presented word is similar to only one-tenth of the total set. This means that targets will seem rather unfamiliar, but, again, so will distractors. Thus, by signal detection theory, the predicted net effect of increasing talker variability may be no effect at all.

Effects of Voice

Examination of the effects of voice revealed an asymmetry between the implicit and explicit memory conditions. Repetition effects for same- and different-voice trials were compared across delays and across memory tasks. In the implicit, perceptual identification task, a benefit for same-voice repetitions was observed at all delays. Also, a negative repetition effect, or a "cost," was observed for different-voice repetitions across delays, although these negative repetition effects were not as robust as the positive repetition effects. The magnitudes of the repetition effects, whether positive or negative, were independent of the length of the delay period between the study and test sessions.

In the explicit recognition memory task, however, the effects of voice and delay were not independent. At the 5-minute delay, same-voice repetitions increased hit rates more than different-voice repetitions by an average of 7.5 percent. At the 1-day delay, however, same-voice repetitions increased hit rates more than different-voice repetitions by an average of only 4.1 percent. Finally, at the 1-week delay, same-voice repetitions increased hit rates more than different-voice repetitions by an average of only 1.6 percent. Indeed, at the 1-week delay, no reliable effects of voice were observed at all.

The asymmetry between implicit and explicit voice memory over delays is similar to other findings in the recent memory literature. Numerous researchers have recently explored the similarities and differences between implicit and explicit expressions of memory. In these investigations, a common finding is that information appears to be lost when assessed by explicit measures, but is clearly present when assessed by implicit measures (e.g., Jacoby & Dallas, 1981; Musen & Treisman, 1990; Tulving, Schacter, & Stark, 1982).

Another important aspect of the present findings involves the effects of *similar* versus *identical* tokens. In the 2-voice conditions, no tokens were presented twice; all new recorded tokens were presented in the test session. Nevertheless, the repetition effects for same-voice trials were robust (indeed, larger than the repetition effects in the 6- or 10-voice conditions). This finding implies that perceptual similarity between episodes, not just an exact match of tokens, is sufficient to access an episodic trace of the study word.

The total number of voices in the stimulus set did not strongly interact with voice effects in either the implicit or explicit memory conditions. In the implicit memory condition, a reliable two-way interaction of voice and number of voices was observed. This was due primarily to the larger voice effects in the 2-voice conditions. The benefit of same-voice repetitions and the cost of different-voice repetitions were both greatest in the 2-voice conditions. The underlying source of this interaction is not readily apparent. One possibility is that the reduced attentional demands of the 2-voice condition allowed for more complete voice encoding of the study trials (see Mullennix & Pisoni, 1990).

Effects of Perceptual Similarity

The asymmetry between implicit and explicit voice memory also emerged in assessing the effects of perceptual similarity on different-voice repetitions. In the 6-voice and 10-voice conditions, perceptual distances between study and test exemplars of spoken words were correlated with the magnitude of their respective repetition effects. In the implicit memory conditions, with either 6 or 10 voices, a strong relationship between perceptual distance and the magnitude of the repetition effect was observed. In all cases, as the perceptual distance between the study and test exemplars of spoken words increased, the size of the repetition effect decreased. Moreover, this relation was remarkably consistent across all three delay periods.

In contrast to the implicit memory condition, the effects of perceptual similarity in the explicit memory condition were not independent of delay. A strong relation between perceptual distances and hit rates was observed in the condition with a 5-minute delay; a weaker, but significant, relation was observed in the condition with a 1-day delay; and no relation was observed in the condition with a 1-week delay.

The negative correlations observed in the present data indicate the following: The absolute perceptual similarity between study and test exemplars of words strongly affects the influence of the study exemplar on perceptual performance. If the study and test exemplars of a particular word are similar to each other, the test exemplar accesses a memory of the study exemplar. This probability of access decreases steadily with decreased similarity. In the implicit test, this strong and consistent relation between perceptual similarity and memory lasted for a full week. In the explicit test, all evidence of this relation was eliminated by the end of the week.

These findings have several important implications for the role of specific episodes in perception and memory. Detailed representations of spoken words appear to persist in memory, and to affect later perception of similar words. Direct or intentional, conscious access to these representations, however, appears to be limited. Researchers such as Kolers (1975a; 1976a), Tulving and Schacter (1990), Jacoby and Brooks (1984), Feustel, Shiffrin, and Salasoo (1983) and others have discussed the role of specific episodes in later perception. As discussed in detail in Chapter V, the present results complement these earlier findings and provide additional support for certain aspects of their theoretical views.

Summary

The present experiment explored implicit and explicit memory across several delay periods, and revealed several important properties of memory for speech. The first major finding was that stimulus-specific details, in the present case attributes of speakers' voices, persist in the episodic traces of spoken words for at least one full week. This finding extends the earlier findings of Craik and Kirsner (1974), Cole, Coltheart, and Allard (1974), Allard and Henderson (1976), and Palmeri, Goldinger, and Pisoni (1992). All of these studies revealed memory for specific pairings of words and voices that lasted up to a maximum of about 5 minutes. The present findings attest to the relative permanence of specific perceptual episodes in memory for spoken words.

The second major finding was that the implicit test revealed stimulus-specific memory at longer delays than the explicit test. As reviewed above, this dissociation of memory measures replicates numerous findings in the literature. This asymmetry between implicit and explicit memory suggests that stimulus-specific memories may participate in the perception of new stimuli. This interpretation of the present data closely follows the conclusions of other researchers. Tulving and Schacter (1990), Jacoby (1978; 1983a; 1983b), Brooks (1987), (Roediger (1990), and Biederman and Cooper (1991) have all discussed the role of specific perceptual episodes in perception and recognition memory. These episodic theories are discussed in more depth in the General Discussion (Chapter V).

The third major finding was that perceptual distance was strongly related to implicit memory for specific perceptual episodes. A similar relation was observed in explicit memory, but did not appear at the longer delays. These findings also reveal the importance of long-lasting, specific episodes in perception and memory. Moreover, these findings suggest that the savings found for same-voice repetitions are due to perceptual similarity, not just to the repetition of exact tokens. The data from the different-voice repetitions were not random: The similarity relations among study and test tokens also influenced memory. Same-voice repetitions, which involved identical tokens, were simply trials with exceptionally high similarity between study and test. As such, any exemplar-based model of memory or categorization (e.g., Nosofsky, 1988a; Hintzman, 1986) should be able to account for the present findings.

CHAPTER IV: Implicit and Explicit Memory: Effects of Levels of Processing

The first experiment provided strong evidence for highly detailed, episodic traces of spoken words. In implicit memory, same-voice repetitions led to higher accuracy of perceptual identification than different-voice repetitions. This same-voice advantage was obtained to equivalent degrees at all delays. Moreover, for different-voice repetitions, the degree of perceptual similarity between the original voice and the test voice for each word had a significant effect on the magnitude of repetition benefit. Again, the observed relation between repetition effects and perceptual similarity was constant across delay periods.

In explicit memory, although same-voice repetitions led to higher accuracy of explicit recognition than different-voice repetitions, the same-voice advantage diminished at longer delays. Also, for the different-voice repetitions, a significant effect of perceptual similarity was again observed. However, unlike the implicit memory conditions, this relation was strong and significant at the 5-minute delay, it was weaker but still significant at the 1-day delay, and it was absent at the 1-week delay.

Taken together, the results of Experiment I suggest that specific perceptual details, in this case voice details, are encoded and retained in memory for appreciable periods of time. Moreover, these specific episodic memories affect performance on perceptual tests long after they are absent in a conscious assessment of memory. As Jacoby (1983a; 1983b) argues, this pattern of results suggests that specific perceptual episodes are collected in memory and influence future perception.

The delay period between study and test has been shown in previous research, and in the present investigation, to differentially affect performance in implicit and explicit memory tests. This asymmetry was used in Experiment I to investigate the nature of long-term memory for attributes of spoken words. Experiment II included a levels of processing manipulation that has also been shown to differentially affect performance in implicit and explicit memory tests.

"Levels of processing" refers to both a theory of memory encoding and a set of experimental manipulations. Craik and Lockhart (1972) proposed that stimuli subjected to deep or elaborate processing form more durable memories than stimuli subjected to shallow processing. For example, Hyde and Jenkins (1973) found that rating the semantic features of sentences led to greater recall than rating the surface features of sentences. Similar findings have since been reported many times in the literature (e.g., Craik & Tulving, 1975; Eysenck & Eysenck, 1980; Rogers, Kuiper, & Kirker, 1977; Graf & Mandler, 1984).

It has also become well-established in recent investigations, however, that manipulating the level of stimulus processing during study has differential effects on performance in implicit and explicit memory tests. Numerous studies have shown that levels of processing strongly affects explicit memory, but not implicit memory. The implicit tasks that have shown non-effects with respect to levels of processing include fragment completion (Graf & Mandler, 1984), lexical decision (Kirsner, Milech, & Standen, 1983), perceptual identification of words (Jacoby & Dallas, 1981), perceptual identification of pictures, and picture naming (Carroll, Byrne, & Kirsner, 1985).

All of these studies examined stimulus presentation in the visual modality; Schacter and Church (in press) recently compared implicit and explicit memory across levels of processing with spoken words. As in the studies cited above, Schacter and Church found a strong levels of processing effect on explicit recognition, but only a small effect in auditory fragment completion.

Experiment II examined implicit and explicit memory as a function of voice and levels of processing, as determined by three classification tasks performed at study. All subjects in Experiment II heard 6 voices in the stimulus lists, and all subjects participated in study and test sessions separated by a 5-minute delay. During study, subjects performed one of three speeded classification tasks, with all words presented in the clear. The difficulty of classification, and thereby the level of processing, was varied across groups of subjects. Some subjects classified words according to the gender of the speakers, others according to initial phonemes, and others according to syntactic classes.

As in Experiment I, the same set of words was presented in each session of Experiment II, but the voices of half the words were changed across sessions. In the implicit memory conditions, subjects performed the same classification task twice in a row; the difference in performance between sessions for the same- and different-voice trials was measured. In the explicit memory conditions, subjects performed the classification task during study, then received a surprise recognition test. Accuracy of recognition for same- and different-voice repetitions was measured. The predictions for Experiment II follow from previous research: As in Experiment I, effects of voice were expected in both implicit and explicit memory conditions. Also, in implicit memory, voice was expected to affect performance equivalently across all levels of processing. In contrast, in explicit memory, voice was expected to affect performance to different degrees across levels of processing. Finally, based on the findings of Experiment I, perceptual similarity between the study and test exemplars of words was expected to affect the magnitude of the observed repetition effects, but not to equivalent degrees in all conditions. Experiment I showed that the relation between perceptual similarity and net repetition effect was constant across delays in implicit memory; this relation was not observed in explicit memory. Accordingly, in Experiment II, the effects of perceptual similarity were expected to be constant across levels of processing in the implicit memory condition, but not in the explicit memory condition.

EXPERIMENT

Method

Subjects

Two-hundred and ten students served as subjects. The subjects were enrolled in introductory psychology courses at Indiana University and received partial course credit for their participation. All subjects were native speakers of English and reported no history of speech or hearing disorders at the time of testing.

Half of the subjects participated in the implicit task (speeded classification), and half participated in the explicit task (recognition memory). Within each test condition, subjects were further divided into three sub-conditions determined by level of processing, as described below, yielding a total of 35 subjects assigned to each condition.

Stimulus Materials

The stimulus materials consisted of the same words described in Chapter II. In this experiment, however, only the 6-voice lists were presented. Referring to the perceptual space shown in Figure 2.2, the six speakers were: F1, F2, F3, M1, M2, and M3.

Design

The experiment consisted of two major conditions corresponding to the implicit and the explicit memory tasks. In the implicit memory condition, subjects performed a speeded-classification task in a study session and then again in a test session. The two sessions were separated by a 5-minute delay. In the explicit memory condition, subjects performed a speeded classification task in the study session, and then received a surprise recognition memory task in the test phase. Again, the sessions were separated by a 5-minute delay. Each of these major conditions contained three sub-conditions determined by the level of processing of the classification task.

Both the speeded classification and the explicit recognition memory tasks contained three completely between-subjects sub-conditions, with 35 subjects per sub-condition. The three sub-conditions were created by using three levels of processing during speeded classification. Subjects classified spoken words according to the gender of the speaker, the initial phonemes of the words, or the syntactic categories of the words.

Procedure

Speeded Classification Task

In the implicit memory condition, the procedures were identical across the study and test sessions: Subjects were tested in groups of six or fewer in a quiet testing room used for speech perception experiments. Each subject was seated in an individual booth equipped with a pair of matched and calibrated TDH-39 headphones, a CRT screen, and a two-button response box. A PDP 11/34 computer was used to present the stimuli and collect responses.

Each trial of the experiment began with a warning prompt on the CRT screen. The phrase "GET READY" appeared on the center of the screen and remained for 500 msec. Immediately following the offset of the warning prompt, two response labels appeared in the lower left-hand and lower right-hand corners of the screen. The responses shown on each side of the screen corresponded to the buttons on each side of the response box (e.g., if "male" appeared on the left-hand side of the screen and a word produced by a male speaker was presented, subjects would indicate their gender classification decision by pressing the left-hand button as quickly as possible). The correct response was mapped to each button an equal number of times. Five-hundred msec after the onset of the response labels, a stimulus word was presented at 75 dB (SPL) over headphones. The digitized stimulus words were reproduced using a 12-bit digital-to-analog converter and were low-passed filtered at 4.8 kHz.

After stimulus presentation, subjects had up to 5 seconds to classify the stimulus item. If all subjects responded before the end of the 5-second period, a 750 msec inter-trial interval elapsed, and a new trial was initiated. Both the study and test sessions consisted of 300 randomly ordered trials. Instructions to subjects stressed both speed and accuracy of responding.

Explicit Recognition Memory Task

In the explicit recognition task, different procedures were used in the study and test sessions. In both sessions, however, subjects were tested in groups of six or fewer using the apparatus described above.

In the study session, subjects performed the speeded classification task described above. However, in the explicit memory conditions, subjects only performed 150 trials of speeded classification during study. In the test session, subjects received a surprise recognition memory test. The recognition test materials included the 150 words from the study session (with half of the voices changed), and 150 new words as well. As in speeded classification, each trial of the recognition memory test began with a warning prompt on the CRT screen. The phrase "GET READY" appeared on the center of the screen and remained for 500 msec. Immediately following the offset of the warning prompt, two response labels ("old" and "new") appeared in the lower left-hand and lower right-hand corners of the screen. As in speeded classification, the responses shown on each side of the screen corresponded to the buttons on each side of the response box. The correct response was mapped to each button an equal number of times. Five-hundred msec after the onset of the response labels, a stimulus word was presented at 75 dB (SPL).

After stimulus presentation, subjects had up to 5 seconds to classify the spoken word. If all subjects responded before the end of the 5-second period, a 750 msec inter-trial interval elapsed, and a new trial was initiated. Both the study and test sessions consisted of 300 randomly ordered trials. Instructions to subjects stressed both the speed and accuracy of responding.

In both the implicit and explicit memory conditions, levels of processing were manipulated as follows: In the gender classification conditions, the response labels "male" and "female" appeared on every trial and subjects classified words according to the speakers' genders. In the phoneme classification conditions, response labels were minimal pairs of words and subjects classified words according to initial phonemes. For example, if the spoken stimulus was the word "bid," the response labels "bid" and "did" were shown, and subjects selected the proper response as quickly as possible. In the syntax classification conditions, response labels represented syntactic categories and subjects classified words according to the categories. For example, for the stimulus word "dog," the response labels "noun" and "adjective" were shown, and subjects selected the proper response as quickly as possible. On every syntax classification trial, one of the responses was always incorrect. So, if the stimulus word was "ride," which is both a noun and a verb, the choices provided were "verb" and "preposition," or some other incorrect option.

Results

Results are presented in three sections. First, the overall results of implicit versus explicit memory are compared. Second, same-voice and different-voice trials are compared across memory tasks and levels of processing. Third, the effects of perceptual similarity between study and test exemplars are assessed across memory tasks and levels of processing.

Overall Performance

For the speeded classification task, the mean percentages of correct responses and mean latencies of correct responses were calculated for each subject. For recognition memory, the percentages of hits, misses, false alarms, and correct rejections were calculated for each subject. These values were converted to *d'* and β, which were then used to estimate *P(C)max*, as in Experiment I. Also, the mean latencies of correct responses, including hits and correct rejections, were calculated for each subject.

Figure 4.1 displays overall performance as a function of memory task and level of processing. The upper panel displays the results, in terms of percent correct and response latency, of the test sessions for the implicit memory conditions. The lower panel displays results, in terms of *P(C)max* and response latency, of the test sessions of the explicit recognition conditions. In both panels, the left-hand axis and the line with open circles corresponds to the accuracy data; the right-hand axis and the line with closed circles corresponds to the latency data. Also, both panels display results as a function of level of processing.

Insert Figure 4.1 about here

Both panels of Figure 4.1 show effects of the levels of processing manipulation. In the speeded classification task (upper panel of Figure 4.1), accuracy decreased and response latency increased at deeper levels of classification. In the recognition memory task (lower panel of Figure 4.1), accuracy increased and response latency decreased at deeper levels of classification.

The data from the speeded classification task were analyzed in two between-subjects one-way (Levels of Processing) ANOVAs. Separate ANOVAs were conducted on the accuracy and latency data. In the accuracy data, a significant main effect of Levels of Processing was observed [$F(1,102) = 120.70$, $MS_e = 9.36$, $p < .0001$], reflecting the decreased accuracy at deeper levels of processing. In the latency data, a significant main effect of Levels of Processing was also observed [$F (1,102) = 332.75$, $MS_e = 120.64$, $p < .0001$], reflecting the increased response latency at deeper levels of processing.

The data from the explicit recognition memory task were also analyzed in two between-subjects one-way ANOVAs. Separate ANOVAs were conducted on the accuracy and latency data. In the accuracy data (analyzed in terms of *P(C)max*), a significant main effect of Levels of Processing was observed [$F (1,102) = 69.22$, $MS_e = 11.91$, $p < .0001$], reflecting the increased accuracy at deeper levels of processing. In the latency data, a significant main effect of Levels of Processing was also observed [$F (1,102) = 83.60$, $MS_e = 93.82$, $p < .0001$], reflecting the decreased response latency at deeper levels of processing.

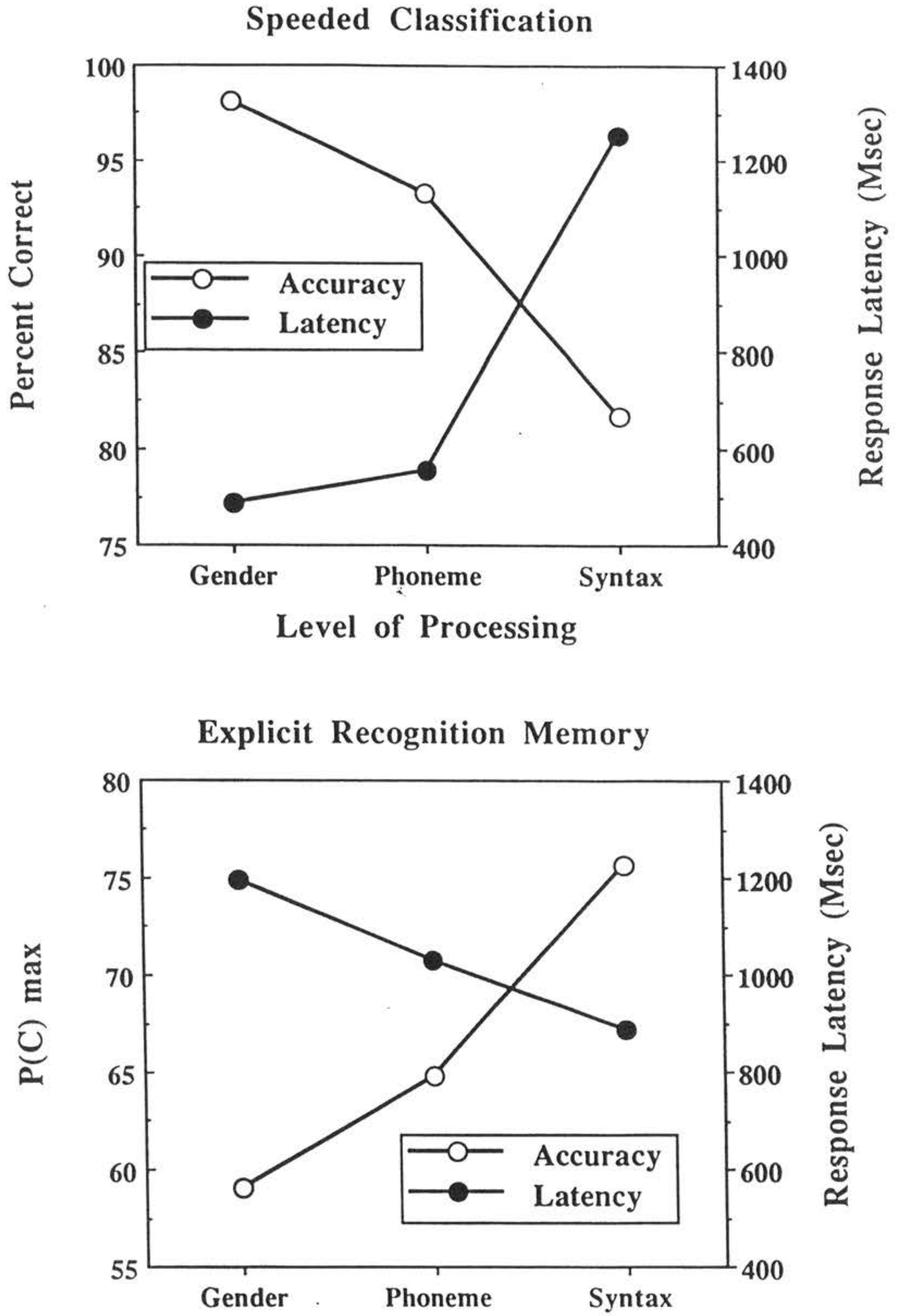


Figure 4.1 Overall effects of the levels of processing manipulation. The upper panel displays percent correct classification and response latency for speeded classification; the lower panel displays percent correct recognition and response latency for recognition memory.

Effects of Voice

Speeded Classification Task

The major purpose of this experiment was to examine the role of voice in memory for spoken words. In both the implicit and explicit memory conditions, words were presented at study and were later repeated at test. Half of the repeated words were spoken in the same voices as their study presentations; half were repeated in different voices as their study presentations. This section presents analyses of repetition effects as a function of voice. For the moment, "different voice" trials are treated as a uniform category of data. The role of precise voice differences on memory is assessed in another section on perceptual similarity effects.

Because the interest in this experiment was the role of voice in the repetition effect, a *net repetition score* was calculated for each subject. Analyzing the data in terms of net repetition scores eliminates the irrelevant variance in the data due to the between-subjects design (see Schacter & Church, in press). For the implicit memory condition, the net repetition score in the accuracy data for each subject was calculated by subtracting the percentage of correct responses in the study session from the percentage of correct responses in the test session. In the latency data, the net repetition score was calculated by subtracting the mean response latency in the study session from the mean response latency in the test session. Accordingly, a positive net repetition score implies an improvement in performance in the test session; a negative net repetition score implies a decrement in performance in the test session.

Figure 4.2 displays the results of the implicit memory conditions, in terms of repetition scores, as a function of voice and level of processing. The upper panel displays response latency data; the lower panel displays accuracy data. Both panels display repetition effects for "same-voice" trials with dark bars and "different-voice" trials with light bars. Also, both panels display repetition effects as a function of level of processing.

Insert Figure 4.2 about here

As Figure 4.2 shows, classification of "same-voice" repetitions was faster and more accurate than classification of "different-voice" repetitions. Moreover, the "same-voice" advantage was consistent across all levels of processing.

Separate 2 X 3 (Voice X Levels of Processing) ANOVAs were conducted on the mean repetition scores for both the accuracy and latency data. In the latency data (upper panel of Figure 4.2), a significant main effect of Voice was observed [$F\ (1,102) = 12.35$, $MS_e = 75.74, p < .0001$]. Same-voice repetitions were correctly classified faster that different-voice repetitions at all levels of processing. After the ANOVA, Tukey's HSD analyses were conducted to compare same-voice and different-voice repetitions at all levels of processing. In all three conditions shown in the upper panel Figure 4.2, the response latencies to classify same- and different-voice repetitions were significantly different ($p < .01$).

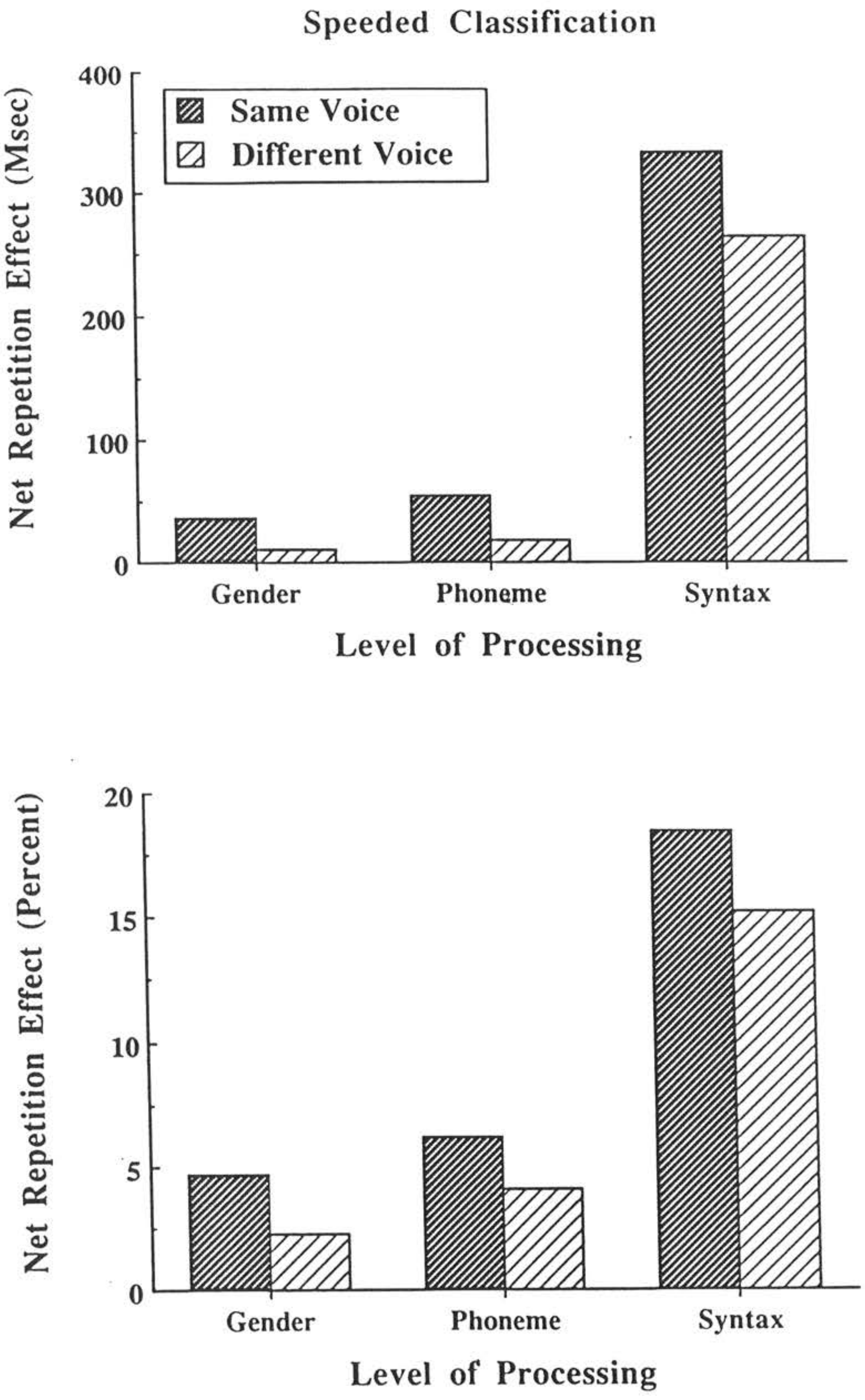


Figure 4.2 Repetition effects in speeded classification as a function of level of processing and repetition voice. The upper panel displays response latency; the lower panel displays accuracy.

As reported above, a significant main effect of Levels of Processing was also observed [$F(1,102) = 332.75$, $MS_e = 120.64$, $p < .0001$]. This reflected the increased response latency at deeper levels of processing. The two-way interaction of Voice X Levels of Processing, however, was not significant [$F(2,102) = 2.07$, $MS_e = 75.74$, n.s.]. The difference in response latency between same- and different-voice repetitions was consistent in direction and magnitude at all levels of processing.

In the accuracy data (lower panel of Figure 4.2), a significant main effect of Voice was observed [$F(1,102) = 23.72$, $MS_e = 9.56$, $p < .0001$]. Same-voice repetitions produced greater repetition effects than different-voice repetitions at all levels of processing. After the ANOVA, conservative Tukey's HSD analyses were conducted to compare same-voice and different-voice repetitions at all levels of processing. In all three conditions shown in Figure 4.2, the accuracy for same- and different-voice repetition trials were significantly different ($p < .01$).

As reported above, a significant main effect of Levels of Processing was also observed [$F(1,102) = 120.70$, $MS_e = 9.36$, $p < .0001$], reflecting the decreased accuracy at deeper levels of processing. The two-way interaction of Voice X Levels of Processing, however, was not significant [$F(2,102) = 1.81$, $MS_e = 5.97$, n.s.]. The difference in repetition effects between same- and different-voice trials was consistent in direction and magnitude at all levels of processing.

Explicit Recognition Memory Task

For the explicit recognition memory task, the mean hit rates for the same- and different-voice repetitions were calculated for each subject. The effects of voice were only assessed via hit rates, rather than sensitivity measures such as *d'* or *P(C)max*. Because false alarms cannot be analyzed in terms of voice (since false alarms are responses to new words), only the hit rates and response latencies to generate hits were analyzed.

Figure 4.3 displays the hit rates and response latencies in the explicit recognition task as a function of voice and level of processing. The upper panel displays hit rates; the lower panel displays response latencies. Both panels display responses to "same-voice" repetitions with dark bars and "different-voice" repetitions with light bars. Also, both panels display repetition effects as a function of level of processing.

Insert Figure 4.3 about here

As Figure 4.3 shows, recognition of "same-voice" repetitions was generally faster and more accurate than recognition of "different-voice" repetitions. However, the "same-voice" advantage was not consistent across levels of processing; the effects of voice were larger in the conditions of shallow processing (gender or phoneme classification) than in the deeper, syntax classification condition.

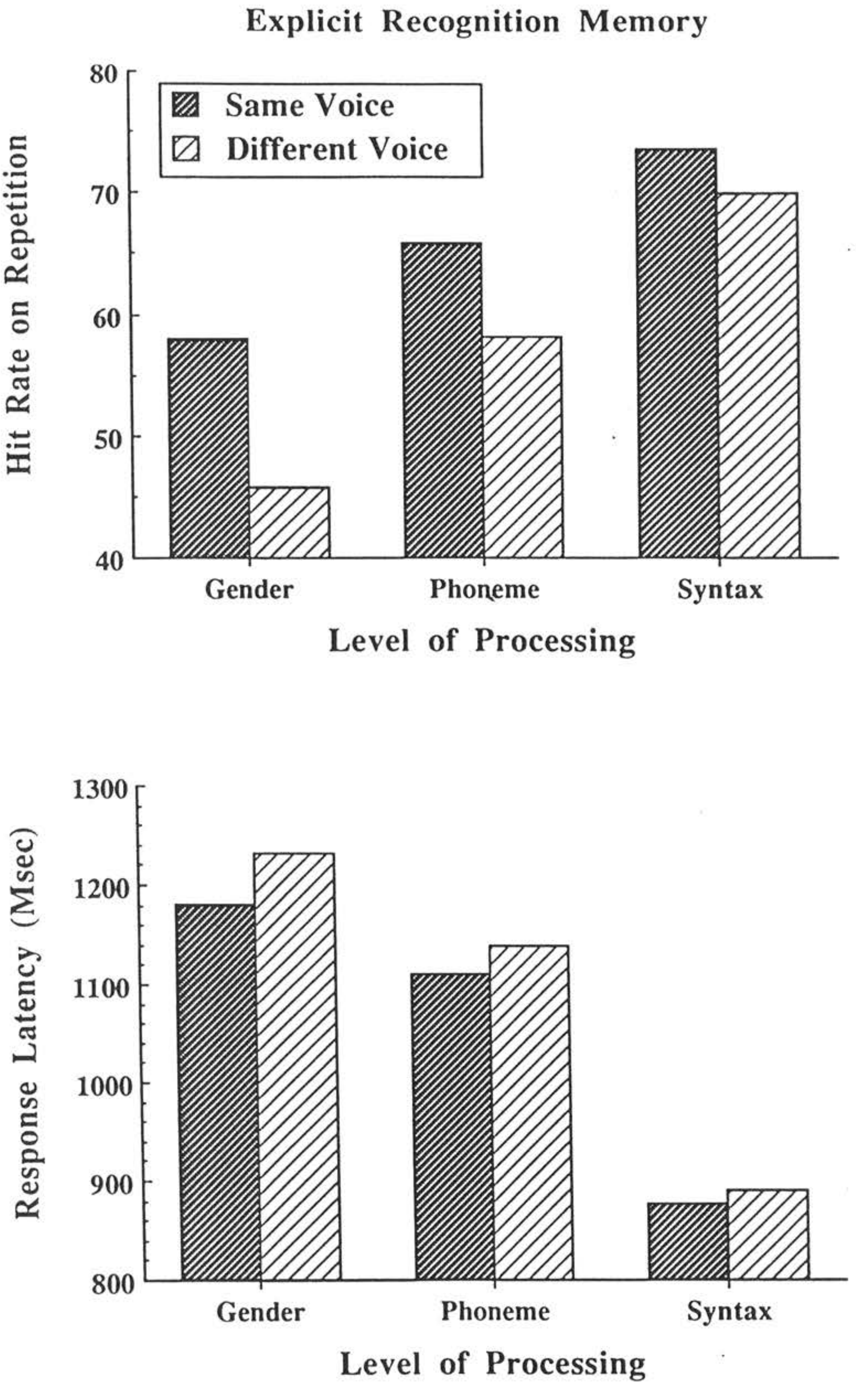


Figure 4.3 Hit rates and response latencies in recognition memory as a function of level of processing and repetition voice. The upper panel displays hit rates; the lower panel displays response latencies.

Separate 2 X 3 (Voice X Levels of Processing) ANOVAs were conducted on the mean hit rates and response latencies. In analyses of the hit rate data (upper panel of Figure 4.3), a significant main effect of Voice was observed [F (1,102) = 85.11, MS_e = 33.95, $p < .0001$]. Same-voice repetitions were correctly recognized more frequently than different-voice repetitions at all levels of processing. After the ANOVA, Tukey's HSD analyses were conducted to compare hit rates to same-voice and different-voice repetitions at all levels of processing. In all three conditions shown in the upper panel Figure 4.3, the hit rates to recognize same- and different-voice repetitions were significantly different ($p < .01$).

A significant main effect of Levels of Processing was also observed [F (1,102) = 26.77, MS_e = 145.62, $p < .0001$]. This reflected the increased hit rates at deeper levels of processing. Moreover, the two-way interaction of Voice X Levels of Processing was also significant [F (2,102) = 10.07, MS_e = 33.95, $p < .0001$]. The difference in hit rates between same- and different-voice repetitions decreased at deeper levels of processing.

In the latency data (lower panel of Figure 4.3, a significant main effect of Voice was also observed [F (1,102) = 13.60, MS_e = 97.42, $p < .0001$]. Same-voice repetitions were correctly recognized faster than different-voice repetitions at all levels of processing. After the ANOVA, Tukey's HSD analyses were conducted to compare latencies to recognize same-voice and different-voice repetitions at all levels of processing. The latencies of same- and different-voice repetition trials were significantly different in the gender and phoneme classification conditions ($p < .01$), but did not differ in the syntax classification condition.

A significant main effect of Levels of Processing was also observed [F (1,102) = 64.09, MS_e = 93.82, $p < .0001$]. This reflected the decreased response latency at deeper levels of processing. Moreover, the two-way interaction of Voice X Levels of Processing was also significant [F (2,102) = 12.38, MS_e = 97.42, $p < .0001$]. The difference in response latency between same- and different-voice repetitions decreased at deeper levels of processing.

Effects of Perceptual Similarity

The overall effect of voice, in terms of same- versus different-voice repetitions, was shown above. Next, the results of the implicit and explicit memory tasks were analyzed in terms of perceptual similarity effects.

These analyses were conducted as follows: All data from different-voice repetitions were divided into categories corresponding to all possible combinations of speakers. Since both conditions included 6 speakers, there were 15 unique combinations of two speakers. The perceptual distances for all 15 combinations of speakers were derived from the earlier scaling solution. All categories contained an equal number of trials. Finally, the dependent measure in each category for each subject was correlated with the perceptual distances corresponding to each category.

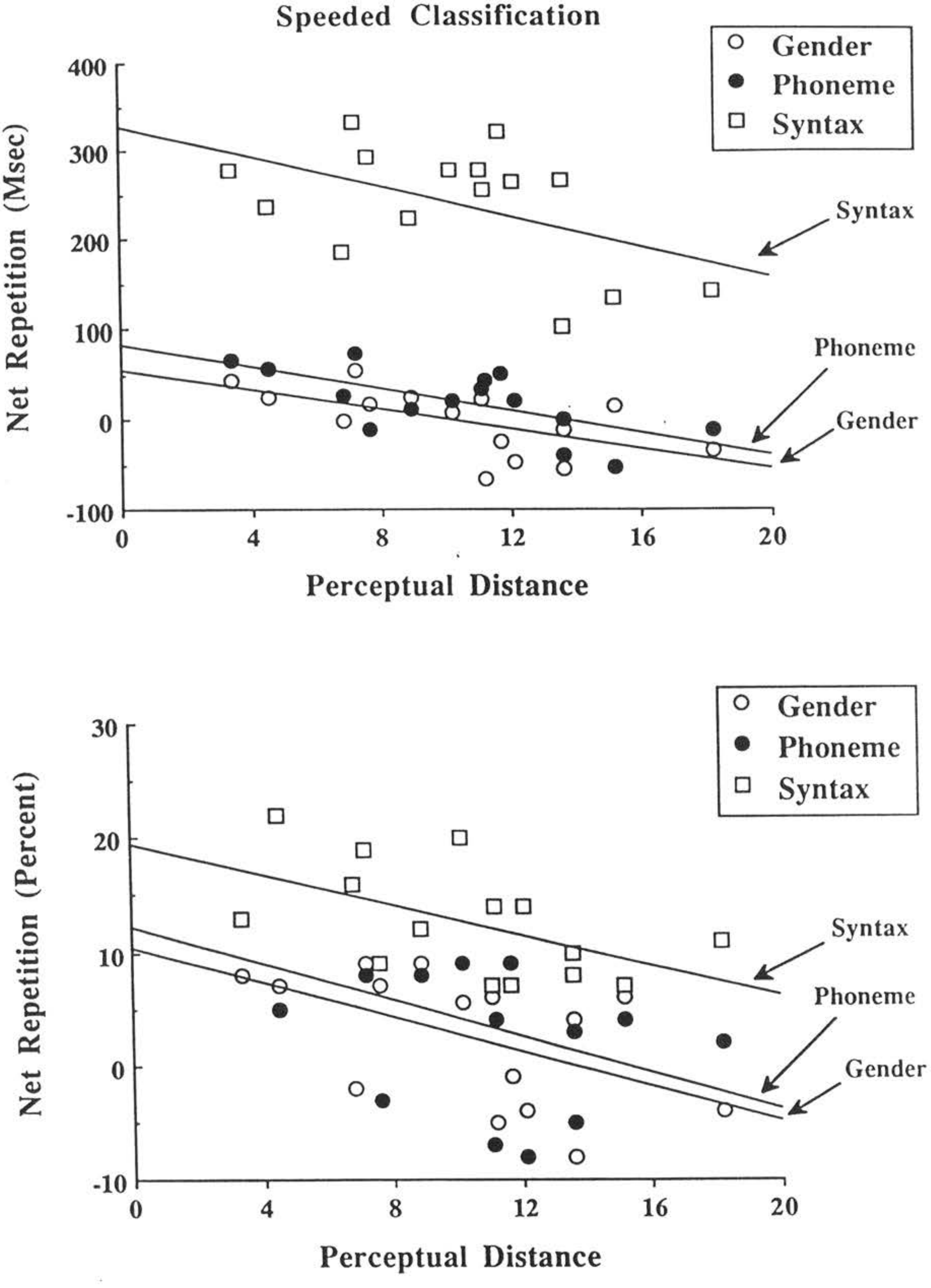


Figure 4.4 Correlations of perceptual similarity and "different-voice" repetition effects in speeded classification. The upper panel displays net repetition effects, in terms of response latency; the lower panel displays net repetition effects, in terms of percent correct classification. Both panels display results as a function of perceptual distance and level of processing.

Figure 4.4 displays correlations between repetition effects and perceptual distance for the implicit memory conditions. The upper panel displays the correlation between perceptual distance and net repetition effects, measured in terms of response latency. The lower panel displays the correlation between perceptual distance and net repetition effects, measured in terms of percent correct classification. Both panels display results as a function of level of processing: Open circles represent the gender classification condition; closed circles represent the phoneme classification condition; open squares represent the syntax classification condition. The best-fitting lines through all data sets are drawn and labeled on the figure.

Insert Figure 4.4 about here

The results in Figure 4.4 are easily summarized: In both latency and accuracy measures, repetition effects in speeded classification were negatively correlated with perceptual distance between voices. Moreover, the strength of the correlations was very consistent across levels of processing, as shown by the parallel slopes in both panels.

In the latency data (upper panel of Figure 4.4), the slopes of the correlations between net repetition and perceptual distance were very similar across all levels of processing. The correlation for the gender classification condition was $r = -.439$ [$F(1,523) = 15.55$, $p < .0001$]; the correlation for the phoneme classification condition was $r = -.448$ [$F(1,523) = 20.71$, $p < .0001$]; and the correlation for the syntax classification condition was $r = -.506$ [$F(1,523) = 36.69$, $p < .0001$].

In the accuracy data (lower panel of Figure 4.4), the slopes of the correlations between net repetition and perceptual distance were again very similar across levels of processing. The correlation for the gender classification condition was $r = -.531$ [$F(1,523) = 77.01$, $p < .0001$]; the correlation for the phoneme classification condition was $r = -.511$ [$F(1,523) = 41.03$, $p < .0001$]; and the correlation for the syntax classification condition was $r = -.549$ [$F(1,523) = 66.20$, $p < .0001$].

As examination of Figure 4.4 shows, the influence of perceptual similarity on repetition effects was constant across levels of processing in implicit memory. This was not the case in explicit recognition memory: Figure 4.5 displays correlations between repetition effects and perceptual distance for the explicit memory conditions. The upper panel displays the correlation between perceptual distance and net repetition effects, measured in terms of hit rates. The lower panel displays the correlation between perceptual distance and net repetition effects, measured in terms of response latencies. Both panels display results as a function of level of processing: Open circles represent the gender classification condition; closed circles represent the phoneme classification condition; open squares represent the syntax classification condition. The best-fitting lines through all data sets are drawn and labeled on the figure.

Insert Figure 4.5 about here

The results in Figure 4.5 are easily summarized: In the accuracy measure (upper panel), recognition memory was negatively correlated with perceptual distance between voices. In the latency measure (lower panel), recognition memory was positively correlated with perceptual distance between voices. In both measures, the strength of the correlations varied across levels of processing; the relationship was weaker at deeper levels of processing.

In the accuracy data (upper panel of Figure 4.5), the correlations between hit rates and perceptual distance varied across levels of processing. The correlation for the gender classification condition was $r = -.550$ [$F(1,523) = 52.33$, $p < .0001$]; the correlation for the phoneme classification condition was $r = -.444$ [$F(1,523) = 20.41$, $p < .0001$]; and the correlation for the syntax classification condition was $r = -.335$ [$F(1,523) = 6.65$, $p < .01$].

In the latency data (lower panel of Figure 4.5), the correlations between response latencies and perceptual distance again varied across levels of processing. The correlation for the gender classification condition was $r = +.635$ [$F(1,523) = 31.91$, $p < .0001$]; the correlation for the phoneme classification condition was $r = +.486$ [$F(1,523) = 16.51$, $p < .0001$]; and the correlation for the syntax classification condition was $r = +.145$, but was not significant [$F(1,523) = 1.28$, n.s.].

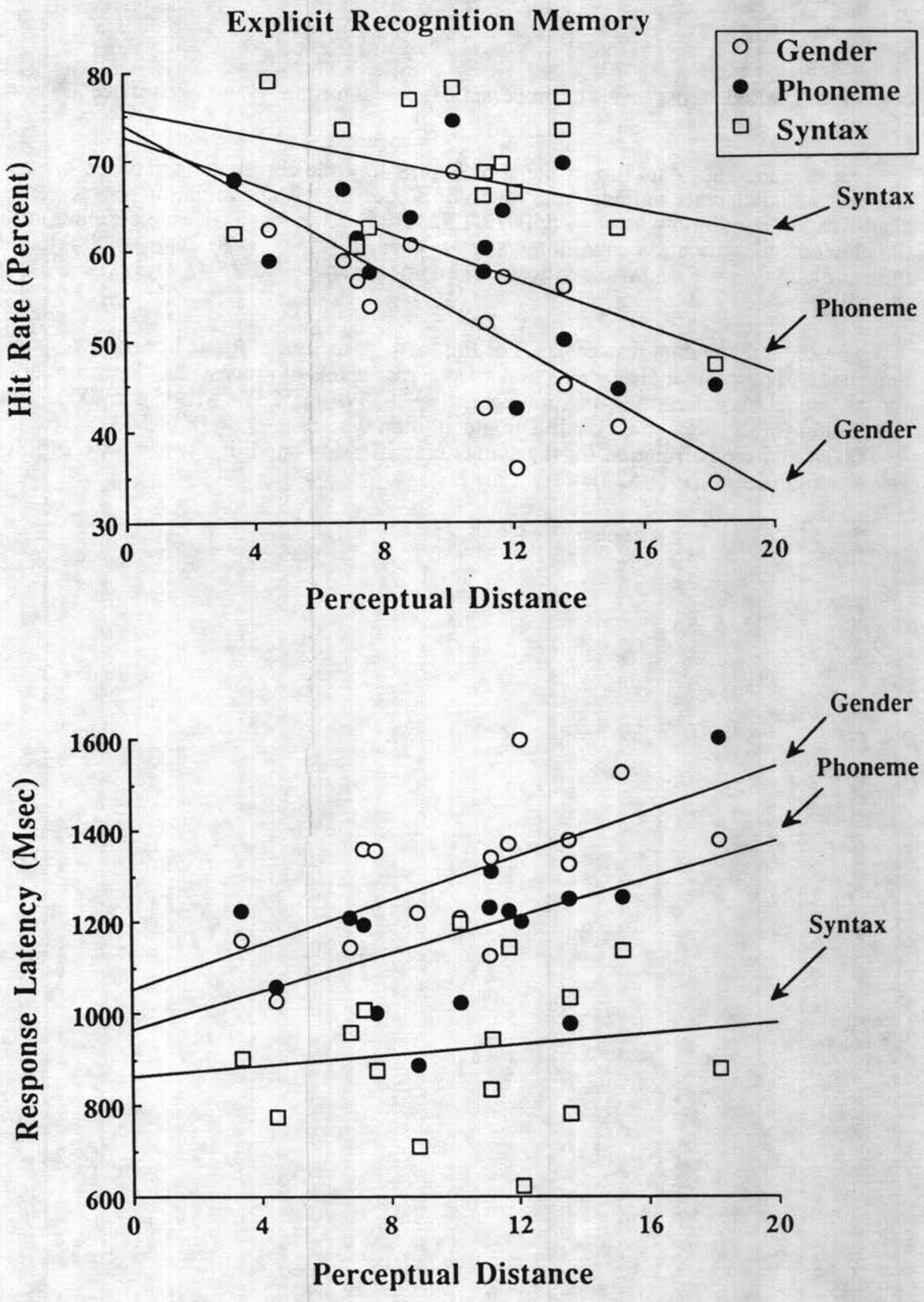


Figure 4.5 Correlations of perceptual similarity and recognition memory for "different-voice" repetitions. The upper panel displays hit rates; the lower panel displays response latencies. Both panels display results as a function of perceptual distance and level of processing.

Discussion

Results are discussed in three sections below. First, the overall results of implicit and explicit memory and levels of processing are discussed. Second, the effects of voice are discussed. Third, the effects of perceptual similarity between study and test exemplars are discussed.

Overall Performance

The overall results of the implicit and explicit memory conditions can be easily summarized. In both conditions, changing the level of processing affected performance. In the implicit, speeded classification task, deeper levels of processing led to slower and less accurate classification. In the explicit recognition memory task, deeper levels of processing led to faster and more accurate recognition. These differences clearly reflect the different task demands.

Effects of Voice

Examination of the data with respect to voice revealed an asymmetry between the implicit and explicit memory conditions. Repetition effects for same- and different-voice repetitions were compared across memory tasks and levels of processing. In the implicit, speeded classification tests, a larger benefit was observed for same-voice repetitions than for different-voice repetitions. This difference was observed to an equivalent degree at all levels of processing, and was present in both the latency and accuracy measures.

In the explicit recognition task, however, the effects of voice and levels of processing were not independent: In the gender classification condition, the same-voice repetition benefit was much larger than the different-voice repetition benefit. In the phoneme classification condition, the difference between same- and different-voice repetition effects was reduced. And, in the syntax classification condition, the difference was reduced still further. In all conditions, the latency and accuracy data displayed the same pattern. The asymmetry between implicit and explicit memory across levels of processing replicates numerous findings in the recent memory literature. As in other investigations, levels of processing affected explicit memory for perceptual details, whereas implicit memory was insensitive to the manipulation (Jacoby & Dallas, 1981; Schacter & Church, in press).

Taken together, the comparison of same- and different-voice repetitions across all conditions shows that voice information is encoded into long-term memory and is accessible via both direct and indirect means. The asymmetry between the two memory measures suggests that repeating specific stimulus tokens benefits perception more than explicit recognition. These results are comparable to the results of Experiment I: In conditions that diminish the effects of voice in explicit memory (long delays or deeper levels of processing), the effects of voice remain robust in implicit memory.

Effects of Perceptual Similarity

The asymmetry between implicit and explicit memory also emerged in an examination of the effects of perceptual similarity on different-voice repetitions. Perceptual distances between study and test exemplars of spoken words were correlated with the magnitude of their respective repetition effects. In the implicit memory conditions, a strong relationship was observed. In all cases, as the perceptual distance between study and test exemplars increased, the magnitude of the repetition effect decreased. Moreover, this relation was remarkably consistent across all levels of processing.

In contrast, the effects of perceptual similarity in the explicit memory conditions were not independent of level of processing. A strong relation between perceptual distance and repetition benefit was observed in the gender classification condition; a weaker relation was observed in the phoneme classification condition; and then a still weaker relation was observed in the syntax classification condition.

As in Experiment I, the absolute perceptual similarity between study and test exemplars of spoken words strongly affected performance. If the study and test exemplars of a particular word were perceptually similar to each other, the test item accessed memory of the study item. This probability of access decreased steadily with decreased similarity. In the implicit tests, this relation was consistent regardless of level of processing. In the explicit tests, the importance of this relation diminished at deeper levels of processing.

These findings, taken together with the findings of overall voice memory, imply a central role of specific episodes in perception and memory. Detailed representations of spoken words persist in memory, and affect later processing of perceptually similar words. Direct or intentional access to these representations, however, was partially determined by the focus of attention at study. When attention was directed to physical stimulus aspects during study, effects of perceptual similarity were observed. When attention was directed at more meaningful aspects of words during study, the effects of perceptual similarity were small or absent. These findings again suggest an asymmetry between implicit and explicit measures of memory across levels of processing. As reviewed in Chapter V, these findings complement findings and theories of Jacoby, Feustel et al. (1983), Schacter, and others.

Summary

Experiment II used implicit and explicit memory measures combined with a levels of processing manipulation to assess the specificity of memory for spoken words. The first finding was that voice information persists in episodic traces of spoken words. This finding complements earlier findings of Schacter and Church (in press); these researchers revealed memory for specific spoken exemplars in explicit recognition memory and fragment completion tasks.

The second finding was that the implicit, speeded classification tests revealed stimulus-specific memory at all levels of processing. As reviewed above, this dissociation of memory measures replicates numerous findings from the recent literature. More importantly, this asymmetry between implicit and explicit memory suggests that perceptual episodes participate in the perception of new stimuli. The findings of Experiment II suggest that specific details of perceptual episodes are encoded automatically; this information is available to later perceptual tests regardless of focus of attention during study. This interpretation follows the suggestions of other researchers: Schacter (1990), Jacoby (1978; 1983a; 1983b), Roediger (1990), and others have all discussed the role of specific perceptual episodes in both perception and recognition memory.

The third finding was that, as in Experiment I, perceptual distance was strongly related to implicit memory for specific perceptual episodes. A similar relation was observed in explicit memory, but it became weaker at deeper levels of processing. These findings again reveal the importance of specific episodes in perception and memory. The data from the different-voice repetitions were not random: The similarity of study and test exemplars of any given word influenced memory for that word. These findings suggest that the savings found for same-voice repetitions were due to perceptual similarity between study and test exemplars, not just to repetition of exact tokens. Same-voice repetitions, which involved identical tokens, simply had exceptionally high similarity between study and test. As such, an exemplar-based model of memory or categorization (e.g., Nosofsky, 1988a; Hintzman, 1986) can be used to account for the findings.

Taken together, the results of Experiment II closely resemble and complement the results of Experiment I: Specific perceptual episodes are encoded into memory and play an important role in later perception and recognition.

CHAPTER V: General Discussion

The present investigation revealed several important characteristics about memory for spoken words. In both the implicit and explicit memory conditions, same-voice repetitions led to greater facilitation of memory or perception than different-voice repetitions of the same nominal stimulus. Moreover, among different-voice repetitions, greater perceptual similarity between study and test exemplars also led to greater facilitation.

The effects of specific perceptual details, however, interacted with the type of memory measure examined. In Experiment I, the magnitude of voice effects diminished at longer delays in explicit memory, but remained constant across delays in implicit memory. In Experiment II, the magnitude of voice effects decreased with deeper levels of processing in explicit memory, but remained constant across levels of processing in implicit memory.

Taken together, the findings of the present investigation suggest that the recognition of spoken words leads to creation and storage of detailed, perceptual episodic traces. If the listener primarily attends to physical details during study, these perceptual episodes influence conscious recognition memory for a brief period. Regardless of the initial focus of attention, perceptual episodes influence later perception for at least a week. Moreover, the effects of perceptual similarity suggest that these findings truly reflect detailed perceptual memory, rather than something more abstract, such as a gender-specific connotative code (e.g., Geiselman & Crawley, 1983).

Relation to Previous Findings

The present findings bear close relation to many previous findings, although several aspects of the present data are unique. Regarding the comparison of implicit and explicit memory measures, the present findings replicate several typical patterns reported in the previous literature. The following sections discuss the similarities and differences between the present set of results and previously published findings:

Consistent Findings

In the present study, although explicit measures suggest that perceptual details are lost from memory with the passage of time, implicit measures reveal their persistent effects on perception. These findings replicate many experiments comparing implicit and explicit measures across delays. For example, Tulving, Schacter, and Stark (1982) reported that explicit recognition memory for words greatly diminished after a 1-week delay; implicit recognition, as measured by fragment completion, was unchanged across the delay. Masson (1984) found that, whereas explicit recognition memory for surface details of transformed sentences diminished over the course of a week, implicit memory was unaffected by delay. Similarly, Musen and Treisman (1990) found that explicit recognition memory for novel visual patterns diminished over a 1-week delay, but implicit memory was again unaffected by delay.

Also, in the present study, although explicit measures suggest that perceptual details are not encoded into memory when attention is focused on deeper levels of processing, implicit measures again reveal their persistent effects on perception. Like the effects of delay on implicit and explicit memory, this finding replicates numerous past investigations. As examples, Jacoby and Dallas (1981) found that a single presentation of visually presented words affect their later perceptual identification. The level of processing during original presentation had no effect on later perceptual identification, but had a large effect on explicit recognition memory. Masson (1984) reported that explicit recognition memory for surface details of transformed sentences was sensitive to level of processing during study, but re-reading times were not. Schacter and Church (in press) reported that repetition effects in auditory fragment completion were independent of level of processing, but explicit recognition was not.

The results of the present investigation also replicate previous findings with respect to memory for specific perceptual episodes. As reviewed in the Introduction, numerous studies have reported either implicit or explicit memory for the specific perceptual details of faces, pictures, songs, sentences, transformed text, and other stimuli. The largest body of data concerns memory for the physical details of visual and spoken words (Craik & Kirsner, 1974; Kirsner, 1973; Jacoby & Hayman, 1987; Schacter & Church, in press). The net result of all these studies is that information for specific, linguistically unimportant details does enter and persist in memory.

The present experiments demonstrated long-memory for the specific details of spoken words. Even with 6 or 10 speakers, the episodic representations of spoken words preserved voice information. Moreover, this voice information affected perceptual performance across delays and across levels of processing. Like many previous studies, then, the present investigation suggests that perceptual details persist in memory and subserve later perception and recognition (Jacoby & Brooks, 1984).

Unlike previous studies, however, the present investigation also assessed the role of perceptual similarity in episodically-mediated perception. The use of perceptual scaling for the stimulus voices led to several findings unique to the present data. Most investigations of memory for specific perceptual details have used two stimulus formats at study and test. Words are presented in one typeface or the other, or one voice or the other, at study and the formats are either changed or retained at test. Although these studies have demonstrated the importance of exact stimulus repetition on memory and perception, they have shed little light on the role of psychological similarity in mediating these effects.

The present data revealed a monotonic relation between the perceptual similarity of study and test exemplars of spoken words and the repetition benefits derived from those words. This finding is consistent with a view of perception based on prior episodes. Jacoby and Brooks (1984) discuss repetition effects in terms of *perceptual fluency*, or the ease of perception for any given stimulus. Perceptual fluency for a stimulus is partly determined by the similarity of the stimulus to previous perceptual episodes in memory. Episodes are assumed to preserve detailed information about perceptual events, including environmental context, affect during experience, etc (see also Gillund & Shiffrin, 1984).

The assumption that episodes retain contextual information appears to be necessitated by the present results. Subjects were presented with relatively common spoken words during study. It is assumed that, during the period between study and test, the subjects were exposed to many words in many voices. In fact, it is almost a certainty that subjects heard most of the words in the stimulus set, and they probably heard many voices like the voices in the stimulus set. Nevertheless, strong effects of perceptual similarity between the exact tokens presented across sessions were observed. Without assuming that experimental context was encoded into study episodes, it is difficult to account for the observed repetition effects (see also Craik, 1981; Godden & Baddeley, 1975; Smith, Glenberg, & Bjork, 1978). Indeed, Jacoby and Witherspoon (1982) discuss a perceptual identification experiment that measured repetition effects across contexts: Words were presented in one room via slide projector during study, and in another room via tachistoscope during test. Repetition effects were substantially reduced, relative to a same-context condition.

Assuming that contextual and perceptual information is stored in episodes, however, the observed effects of perceptual similarity can be accounted for by Jacoby and Brooks' hypothesis: Greater perceptual overlap between study and test exemplars of spoken words led to greater perceptual fluency. According to Jacoby and Brooks (1984; see also Jacoby, 1991; Johnston, Dark, & Jacoby, 1983), greater fluency not only improves perceptual performance, but also increases feelings of familiarity during explicit recognition tests. The present data are consistent with this hypothesis as well. In short, measuring the effects of perceptual similarity on repetition effects provided valuable new insights into the specificity of information in perceptual episodes.

An Inconsistent Finding

One aspect of the present data is not entirely consistent with previous research. Schacter and Church (in press) reported significant effects of voice on auditory fragment completion and recognition memory. However, they only observed significant voice effects when words were presented clearly. If words were masked by white noise, either during study or test, no effects of voice were found. Schacter and Church discuss the intriguing possibility that the interaction of masking and voice effects is due to hemispheric asymmetries: The left hemisphere reportedly operates on categorical, abstract information, discarding perceptual information such as voice (Liberman, 1982; Mann & Liberman, 1983; Safer & Leventhal, 1977). In contrast, the right hemisphere operates on more veridical perceptual information, preserving information such as voice (Safer & Leventhal, 1977; Van Lancker & Kreiman, 1987; Van Lancker, Kreiman, & Emmorey, 1985). These asymmetries suggest that, if processing in the right hemisphere is impaired, voice effects will not be observed. As it happens, Zaidel (1978) reported that split-brain patients experience greater difficulty recognizing spoken words in noise presented to the right hemisphere than to the left hemisphere. Taken together, these findings suggest that voice effects should not be observed when words are masked by noise because processing in the voice-sensitive right hemisphere is selectively impaired. This prediction is consistent with Schacter and Church's results. The neuropsychological argument outlined by Schacter and Church (in press) is parsimonious and well-founded in the literature. Moreover, it may be valid to some extent; voice effects may be attenuated in noise via right hemisphere interference. In Experiment I of the present investigation, however, large and significant effects of voice were observed, despite the use of perceptual identification of words in noise at study and test. At first glance, this finding appears to contradict the hypothesis proposed by Schacter and Church.

It is important to note, however, that the present data do not invalidate the neuropsychologically-derived hypothesis of Schacter and Church (in press). Their hypothesis does not explicitly predict that voice effects should *never* be observed with words presented in noise; the prediction is only that the right hemisphere is primarily responsible for voice effects. As such, voice effects should be *attenuated* with degraded words, relative to non-degraded words, but not necessarily absent. This direct comparison was not carried out in the present investigation. As Schacter and Church note, the right-hemisphere/voice hypothesis could be tested in future research via comparisons of ear differences in voice sensitivity, or by examining patients with specific hemispheric disorders.

Several differences between the methods used by Schacter and Church (in press) and the methods used in the present investigation may explain the different findings with respect to noise and voice effects. First, the present investigation used many more trials and more subjects than the Schacter and Church experiments. Therefore, the present investigation simply had more power to detect a significant difference. Second, and more interesting, Experiment I of the present research differed from the Schacter and Church experiments with respect to preservation of stimulus format: In the Schacter and Church experiments, words were always presented in the clear during study and noise was used only during test. In Experiment I of the present research, subjects performed perceptual identification to words in noise in both the study and test sessions. By the present method, therefore, subjects received the exact same stimulus array (signal+noise) in both sessions.

In research, there is a tendency to equate variables with psychological reality. The procedural variables we manipulate, such as the presence of a mask, are assumed to affect only the procedures of mind. We do not typically entertain the idea that our manipulations may actually become part of the mind itself. If one believes that word recognition proceeds by matching inputs to abstract units, such as logogens (Morton, 1969), it is only natural to assume that the mask increases task difficulty and is then forgotten. By this model, the word recognition system extracts meaning from the signal, then discards the details of the episode. If one believes that every episode leaves a unique trace in memory, however, this off-hand treatment of a mask is not easily justified: By this model, the word recognition system extracts meaning from the signal, then combines that derived meaning with the perceptual details of the episode to create a new memory (Gillund & Shiffrin, 1984; Hintzman, 1986; Jacoby & Brooks, 1984). Accordingly, the masking stimulus should be encoded along with all other perceptual and conceptual details. Thus, repetition of the exact stimulus/mask combination during a later test might be expected to yield greater repetition effects. This greater benefit may arise because of exact stimulus encoding (e.g., Hintzman, 1986).

Alternatively, a compromise between the abstractionist and episodic models may be appropriate: The repetition of a stimulus in the same masking noise may invoke the same perceptual operations at test as in study. By this account, perceptual processes operate to identify the signal in noise, thereby converting the physical signal to an abstract interpretation. A detailed record of the perceptual operations is kept, however, and this record constitutes the episodic representation. These proposals follow from Kolers and his colleagues (Kolers, 1973; 1974a; 1974b; 1975a; 1975b; 1976a; 1976b; Kolers & Magee, 1978; Kolers & Ostry, 1974; Kolers, Palef, & Stelmach, 1980; Rudnicky & Kolers, 1984), and from Kirsner, Dunn, and Standen (1987; see also Kirsner & Dunn, 1985). The idea of record-based word recognition and representation closely approximates the idea of detailed perceptual episodes in terms of empirical predictions. Nevertheless, they are distinct claims in theory, and are considered more fully below.

Implications for Theories of Speaker Normalization

The present results have several implications for current theories of speech perception, especially with respect to the problems of perceptual normalization. As reviewed in the Introduction (Chapter I), most theories of speech perception are cast in terms of stages of information processing. The continuously varying speech signal is assumed to be converted to a series of discrete units, such as phonemes and words, via processes of information reduction and matching to memory (Oden & Massaro, 1978; Pisoni & Sawusch, 1975; Studdert-Kennedy, 1976). Inherent in this view of speech perception is the concept of perceptual normalization: Following Neisser (1967), this term implies that a speech pattern is pre-processed with respect to phonetically irrelevant information, such as voice-specific details. The pre-processing is assumed to reduce the stimulus to its essential abstract elements, such as phonemes, so it can be compared with canonical units in long-term memory (Posner, 1964).

Joos (1948) was the first theorist to directly address the problem of speaker variability in speech perception. Joos noted that speakers vary widely in terms of vocal tract dimensions, dialects, and other factors, all of which contribute to the acoustic realization of speech. Peterson and Barney (1952) later demonstrated empirically that vowel spaces vary dramatically across men, women, and children. Nevertheless, the listener apparently has little difficulty recognizing the messages that speakers intend to communicate. Joos (1948) suggested that the listener takes a small sample of speech from a speaker and then uses this sample to estimate the entire vowel space for the speaker's vocal tract dimensions. Vowel perception then makes use of this constructed representation.

Speaker normalization conducted in the manner described by Joos (1948; see also Gerstman, 1968) has been called *extrinsic normalization*. This view may be contrasted with *intrinsic normalization*, in which it is assumed that every token from a speaker contains enough information for identification, without the need for a representation of the speaker's entire vowel space. According to intrinsic theories, each token of speech contains relational information among formants that uniquely specifies the dimensions of the speaker's vocal tract (e.g., Johnson, 1990; Miller, 1989; Nearey, 1989; Syrdal & Gopal, 1986). For purely parsimonious reasons, the intrinsic theories are preferable to the extrinsic theories. Extrinsic normalization is an inherently circular concept: The listener must identify several vowels in order to create a representational background which is then used to recognize vowels. This circularity (which Nearey, 1989, calls the "bootstrapping problem") weakens the explanatory power of extrinsic normalization. It may also be a problem for intrinsic normalization, depending upon the specific model in question. The relative merits of these normalization theories is not particularly important here; the critical point is that any theory of speaker normalization indicates a theoretical disposition that may not be warranted.

Although Joos never implied that information is "lost" via normalization, the concept of information loss was certainly consistent with later information processing models of speech perception (see Pisoni, 1990). For example, although it is rarely stated in so many words, most theories of speech perception treat voice-specific details as "noise" that the speech processor must bypass en route to deriving a phonetic percept (e.g, Elman & McClelland, 1986). For example, Nusbaum and Morin (1992) write:

> In order to determine the intended phonetic category corresponding to a particular acoustic pattern, a listener must recognize an acoustic pattern by taking into account the vocal characteristics of the talker. The acoustic pattern of a segment must be interpreted in the context of the talker who produced the segment. In other words, the listener must normalize talker differences in order to recognize the phonetic structure of speech. (Page 114).

Clearly, although Nusbaum and Morin (1992) never assert that voice information is discarded after normalization, the quote implies that voice is the vehicle that carries abstract information. The perceptual system is designed to exploit the vehicle, but only attend to the passengers. I believe the view held by Nusbaum and Morin is representative of most theories of speech perception.

The results of the present investigation, as well as many other investigations, suggests that speaker normalization does not operate by information reduction. The present findings demonstrate specific long-term memory for voice. Similar data have been reported by Craik and Kirsner (1974), Light et al. (1973), Goldinger et al. (1991), Palmeri et al. (1992), and Schacter and Church (in press). Springer (1973) found that irrelevant vocal pitch is remembered for several minutes after subjects performed a phoneme-based AX task. Jusczyk, Pisoni, and Mullennix (in press) found that 2-month old infants are able to recognize phonetic constancy across voices, yet they remember voice information for several minutes afterward. Green, Kuhl, Meltzoff, and Stevens (1991) examined the McGurk illusion in conditions of incongruous information. Green et al. found that, even on trials that paired a female face and a male voice, the sources of information combined to yield a phonetic illusion. This result suggests that the perceptual system normalizes the signal at early stages of processing, leaving fusion to operate on "neutral" representations. Nevertheless, Green et al. note that perceptual information remains available to the subject:

> The results of the present study are compatible with this notion because they demonstrate that differences in the gender of the talker producing the auditory and visual signals had no impact on the integration of phonetic information. Thus, by the time the phonetic information was integrated from the auditory and visual modalities, it was sufficiently abstract as to be neutral with respect to the talker differences. Nonetheless, the results of Experiment 3 show that observers are very aware of an incompatibility between the cross-gender face-voice pairs. This suggests that the neutralization of talker differences for the purposes of phonetic categorization does not result in a loss of detailed information about the talker (cf. Pisoni, 1990). (Page 533).

In sum, there is no solid evidence to suggest that speaker normalization serves to reduce the perceptual complexity of the signal. If speaker normalization occurs at all, it apparently serves only to adjust perceptual criteria for phonetic classification; it does not "normalize for voice" in the manner the term would imply. Note that this does not invalidate the concept of normalization in any fundamental sense; it merely constrains theories of normalization. To remain faithful to all the relevant data, speaker normalization must be construed as perceptual compensation, rather than perceptual filtering. As Hyde (1972) writes:

> The human speech communication process is remarkably resistant to very severe corrupting influences. There is evidence that the listener can make use of information from a great many sources, employing both instinctive and acquired knowledge in different forms of information in the speech wave itself. It is most unlikely that, in developing the speech facility, the evolutionary process has omitted valuable information... (Page 423).

Following Hyde's speculations, the present data suggest that, because voice details remain in memory, perhaps they are used in later perception. As it turns out, many theories of vowel normalization posit mechanisms of perceptual compensation without necessary information loss. Models proposed by Johnson (1990), Miller (1989), Nearey (1989), Syrdal and Gopal (1986) and others address the computations necessary to recover vocal tract dimensions from formant frequency relations. These computations are assumed to provide a reference frame for subsequent interpretation of the speaker's vowels. None of these models proposes that voice-specific information is lost as a result of higher-level linguistic processing.

Computational models of vowel compensation demonstrate that the concepts of speaker normalization and voice memory can peacefully co-exist. However, it is still important to consider two questions: First, is normalization a sound theoretical construct? Second, is normalization a real phenomenon? With respect to the first question, theories of speech perception have typically assumed speaker normalization as a logical necessity. These theories, however, all begin with the assumption that speech perception and spoken word recognition functions by matching the signal to canonical, abstract units in memory. Thus, the highly variable acoustic signal must be converted to a "normal" form prior to its comparison with memory. Clearly, different representational assumptions in a theory of speech perception could lead to the removal of this "logical necessity." A lexicon containing detailed, perceptual episodes may be able to match spoken words directly to similar prior traces (Gillund & Shiffrin, 1984; Hintzman, 1986; Jacoby & Brooks, 1984; Nosofsky, 1986), without the intervening normalization processes.

With respect to the empirical question, the data on speaker variability do not unequivocally demonstrate the operation of normalization processes. First, consider vowel perception: It is well-established that listeners' criteria for vowel perception change as a function of the speaker (e.g., Ladefoged & Broadbent, 1957; Johnson, 1990). Moreover, vowel perception is modulated by speaking rate (Verbrugge, Strange, Shankweiler, & Edman, 1976) and even by changes of room reverberation (Watkins, 1991). However, these influences on vowel perceptions need not imply normalization; they merely demonstrate that listeners are able to recognize intended vowels across variable contexts.

Research by Verbrugge et al (1976) and by Verbrugge and Rakerd (1986) suggests that the speech signal contains enough rich dynamic information to specify vowel identity without reference to vocal tract dimensions. For example, Verbrugge and Rakerd presented "silent-center" syllables to listeners for identification. These stimuli were /bVb/ syllables with the central 60% removed, leaving only the initial and final consonants with partial vowel transitions and silence in-between. Verbrugge and Rakerd found that listeners could easily identify the missing vowels from the remaining time-varying transitional information. In another condition, the initial portions of syllables produced by male and female speakers were spliced together to create new silent-center stimuli. Although the vowel spaces of the speakers differed widely, missing vowels were still accurately identified. From these data, Verbrugge and Rakerd concluded that vowels need not be identified via their center frequencies, as most models of vowel perception assume. Instead, speaker-independent dynamic information is sufficient for accurate vowel perception. Verbrugge and Rakerd (see also Verbrugge et al., 1976; Shankweiler, Strange & Edman, 1976) suggest that speech may be accurately perceived without any processes of speaker normalization.

In addition to experiments on vowel perception, another line of evidence relevant to speaker normalization comes from recent studies of speaker variability effects in spoken word recognition, speeded classification, and memory. As reviewed in the Introduction, several recent experiments have demonstrated that speaker variability impairs phoneme classification, spoken word recognition, selective attention, and serial recall. Mullennix, Pisoni, and Martin (1989) found that speed and accuracy of spoken word recognition was impaired in multiple-speaker conditions, relative to single-speaker conditions. Nusbaum and Morin (1992) also observed slower phonemic and lexical classification in multiple-speaker conditions, relative to single-speaker conditions. Martin, Mullennix, Pisoni, and Summers (1989) found that serial recall of words in the primacy portion of the serial position curve was less accurate in multiple-talker lists than in single-talker lists. In all of these reports, speaker variability effects were interpreted in terms of speaker normalization. It was concluded that the presentation of each new voice invoked an obligatory normalization process. This process is assumed to take time and demand processing capacity, thereby reducing performance.

The problem with the speaker normalization explanation of these findings is the confounding presence of stimulus variability. In all of these studies, performance was compared across conditions containing no speaker variability to conditions of extreme speaker variability (10 voices). Using a Garner (1974) speeded classification procedure, Mullennix and Pisoni (1990) demonstrated that phonetic content and speaker's voice are integral dimensions of spoken words. They found that changes of speaker could not be ignored, despite subjects' efforts to selectively attend to another dimension. Indeed, Mullennix and Pisoni found that voice was more salient than phonetic content. Processing interference was asymmetric across dimensions; voice changes interfered more with phonetic classification than vise-versa.

The Mullennix and Pisoni (1990) findings cast some doubt on the normalization-based explanation of previous studies. The finding that speaker variability is salient and attention-demanding implies that, even if no normalization process occurs, speaker variability should impair performance of subjects operating under time pressure. Indeed, other results have shown that when time pressure is alleviated, the effects of speaker variability are reduced. Goldinger, Pisoni, and Logan (1991) found that, when subjects were given ample time for list rehearsal, the detrimental effects of speaker variability on serial recall were eliminated, or even reversed. Given enough rehearsal time, subjects recalled primacy items from multiple-voice lists better than the same items from single-voice lists. This finding suggests that speaker variability may disrupt serial recall by distracting attention, rather than requiring normalization.

Virtually all of the findings reviewed above could be due to mere distraction, rather than speaker normalization. The information processing assumptions of speech perception theories are so deeply ingrained in most researchers that this possibility has been largely ignored. Very recently, Sommers, Nygaard, and Pisoni (1992) have tried to address this problem by comparing conditions of speaker variability, speaking rate variability, and amplitude variability. Sommers et al. observed typical effects of speaker variability on serial recall, but no significant effects of amplitude variability. Although this finding is a positive step toward resolving the variability confound, considerably more research is needed. Most importantly, the perceptual scales of speaker variability and amplitude variability must be compared directly, perhaps via another Garner task, to ensure that both dimensions are equally salient. Preliminary research by Tomiak, Green, and Kuhl (1991) has already examined Garner interference between dimensions of phonetic identity, speaker identity, and speaking rate. If further research demonstrates that the dimensions are not equally salient, the non-effect of amplitude variability does not dismiss the confound. Regardless of the outcome, the theoretical point is that all variability may or may not be psychologically equivalent. Determining the relative importance of different sources of variability to the speech perception system may provide insight into its representations and processes.

Taken together, the research and theory do not create a compelling argument for a speaker normalization process. At present, there is no more reason to assume perceptual re-scaling or tuning than any other possible response to speaker variability. The dominant assumptions of normalization arise from a theory of speech perception based on abstract, template-like representations of symbolic entities. Once the representations in the theory are questioned, the processes used to access the representations must be questioned as well. Moreover, the data on speaker variability effects are also equivocal. Additional research and theoretical developments are clearly needed. Aspects of a theory of word recognition based on perceptual episodes are discussed below; the implications of such a theory for speaker normalization are described. The assumptions of an episodic, non-analytic theory may provide new directions for research on these difficult problems in speech perception.

Implications for Theories of Speech Perception

Recognizing spoken language is a very complex process; multiple knowledge sources and levels of representation interact in countless combinations. The complexity of language has, to date, precluded the formulation of theories that are both global and empirically testable. As such, the situation in language perception is similar to other areas of cognitive science; investigators have typically examined only the details of specific phenomena within narrow experimental paradigms, rather than more complex or integrative issues of their fields (see Newell, 1973).

The following sections illustrate this situation by their unfortunate dichotomy. The first section reviews several models of speech perception; the second reviews several models of spoken word recognition. This segregation is a reflection of the orientation of the models themselves. With a few notable exceptions (e.g., Klatt's LAFS model and TRACE), the majority of models in the literature were formulated to explain either the identification of phonemes in the speech signal or the mapping of strings of phonemes onto lexical representations in memory. Few models are specified in enough detail to address the integrated processes of speech perception and word recognition (see Pisoni & Luce, 1987). Although one trend, evident in the connectionist movement, has been to group these processes into unitary models, another trend has been to justify the segregation of speech processes by arguing that the processes are segregated in perception. For example, the concept of the phonetic module (Liberman & Mattingly, 1985, 1989) purportedly justifies narrow consideration of phonetic perception, without regard to the mapping of speech representations onto lexical representations. I believe the theoretical demarcation of speech perception and spoken word recognition is rather contrived; data relevant to one domain are almost certainly relevant to the other. This demarcation of theories appears especially problematic in light of well-known lexical effects on speech perception (e.g., Ganong, 1980; Samuel, 1986, Samuel & Ressler, 1986; Warren, 1970).

Aside from the issues surrounding speaker normalization, the present findings have several implications for theories of speech perception. The present experiments suggest that specific perceptual details of spoken words are encoded and retained in memory. Moreover, the results from the implicit memory conditions suggest that these detailed episodic traces are involved in later recognition of perceptually similar words. None of the current theories of speech perception or spoken word recognition consider episodic traces as part of the representational substrate or perceptual mechanism.

Before discussing any specific theories, a comment on parsimony: In the context of most theories of speech perception, the present data would typically be considered either anomalous or irrelevant. After all, despite the memory left behind by the recognition process, the fact remains that most phonemes and words were correctly recognized. Many theorists would claim that the recognition and the memorial encoding of words are separate processes, each depending upon separate modules (Fodor, 1983). Indeed, the observation that Jacoby's considerable body of data has been ignored by most theories of visual word recognition exemplifies this position: Data such as those reported here are considered relevant to theories of memory, rather than theories of perception. This is an unreasonable bias for several reasons: First, all current theories assume that speech perception depends on memory and categorization. There is no a priori reason to assume the memories in question are abstract phonemic or lexical units. Moreover, findings such as Jacoby and Hayman's (1987) and the present data indicate that specific memories are involved in perception. If episodes are automatically created, and if they affect later perception, it is most parsimonious to assume that they also constitute an important aspect of the human speech perception system.

Information-Processing Theories

How might contemporary theories of speech perception account for these findings? For the wide class of generic, information-processing models, the notion of non-analytic perception is antithetical to the most basic theoretical assumptions. As an example, consider Massaro's Fuzzy Logical Model of Perception (FLMP) (Derr & Massaro, 1980; Massaro, 1972, 1987, 1989; Massaro & Cohen, 1976, 1977; Massaro & Oden, 1980; Oden & Massaro, 1978):

> FLMP assumes three operations in phoneme identification. First, *feature evaluation* determines the degree to which any given acoustic-phonetic feature is present in a stretch of sound. Unlike more conventional feature-detector theories, FLMP assumes that features are evaluated along a continuous scale, rather than along an absolute binary scale. Features are assigned continuous, "fuzzy" values ranging from zero to one, indicating the degree of certainty that the feature is actually present in the signal (Zadeh, 1965). The second operation in FLMP is *prototype matching*, in which the feature profiles derived by evaluation are compared to phoneme prototypes stored in memory. Prototypes are stored as sets of propositions that describe idealized representations of the acoustic correlates of each phoneme. The prototype matching operation specifies the degree of correspondence between these idealized phonemes and the input features. The final operation, *pattern classification*, determines the best match between the candidate phonemes and the input, using "goodness of fit" algorithms. FLMP provides flexibility in pattern classification by using a variety of logical rules for feature integration, so that perfect matches between the input and the prototypes are not required for phoneme identification to succeed.

FLMP is an attractive and powerful model for many reasons (see Goldinger, Pisoni, & Luce, in press). The present data, however, suggest that most of the processes in FLMP may be unnecessary. Like virtually all other information processing models of speech perception, the operations of FLMP are designed to match a variable signal to an idealized, segmental representation. If speech perception actually occurs against a background of numerous, variable representations, such operations could be bypassed. Without belaboring the point with examples, it is sufficient to note here that this comment applies to other theories of speech perception such as LAFS (Klatt, 1979) and TRACE (McClelland & Elman, 1986). (However, as discussed below, both LAFS and TRACE contain assumptions that could be modified to accommodate the present data.)

Motor Theory of Speech Perception

The original motor theory described by Liberman et al. (1957) was based on the assumption that "speech is perceived by processes that are also involved in its production (page 452)." This view of speech perception was motivated by the fact that a listener is also a speaker, and that a direct link exists between the acoustics of speech sounds and their underlying articulation. As such, an effective and economical means to perceive speech is to perceive the gestures that produce sounds. The motor theory argues that a solution to the invariance problem lies in the reliable nature of articulatory gestures, relative to acoustic phonemes, as units of perception (Fowler, 1986).

The present data have little relevance to claims of gestural versus auditory speech perception. (As noted by Klatt, 1989, and by Goldinger, Pisoni, and Luce, in press, this claim has been difficult to resolve by any empirical means.) The present data demonstrate that the precise details of speech perception are, in fact, preserved in memory; the data cannot elucidate the underlying structure of these perceptual details. With respect to another claim of motor theory, however, the present data are clearly relevant.

Recently, the motor theory has been revised in several key regards (Liberman & Mattingly, 1985, 1989). For the purposes of the present discussion, the most important change entails an assumption of strict modularity in speech perception (see chapters in Mattingly & Studdert-Kennedy, 1991). By this modification of the motor theory, articulatory gestures are assumed to be perceived directly (following Gibson, 1966) by an innate phonetic module. The phonetic module is assumed to be specialized for speech perception, to operate independently of general auditory processes, and to be separate from other perceptual and cognitive systems (for a comprehensive review of modularity, see Fodor, 1983; 1985).

The concept of modularity was invoked in the revised motor theory to provide a more precise definition of speech specialization, an inherent aspect of the theory. For the concept of modularity to carry any theoretical merit, however, it is important that some clear delineations exist between cognitive domains. The careful reader of Fodor (1983) will notice that one module that Fodor discusses at length is the "language module." As Jusczyk and Cohen (1985) note, a module of this magnitude does little to improve our concept of cognition. The problem is exacerbated when the prevalence of memory is considered: Memory is involved in language perception and production, visual perception, skilled motor behavior, odor recognition, and countless other human activities. To what extent does the phonetic module rely on lexical memory? The most recent descriptions of the motor theory (Liberman & Mattingly, 1989) make no reference to any extra-modular sources of information. Nevertheless, the present data, and related data, suggest that prior perceptual episodes contribute to later speech perception. This result leaves the modular assumption in a difficult quandary: If countless episodes are assumed to reside in the module, the concept of the module loses all explanatory power. If episodes are assumed to reside elsewhere in "general memory," the tenet of *cognitive impenetrability* states that they should have no effect on phonetic perception.

Along these lines, it is important to note that Fowler (1986; Fowler & Rosenblum, 1990, 1991) has recently described a framework for a *direct-realist* approach to speech perception. The direct-realist approach assumes that, as in Gibson's (1966) view of visual perception, the perception of speech entails the recognition of natural "phonetic events." As in the motor theory, Fowler assumes that the relevant events perceived in speech are phonetically structured articulations (or gestures). Fowler's approach is similar to the motor theory in several respects. An important difference, however, is the conception of modularity: Fowler and Rosenblum (1991) argue that phonetic perception need not be a modular process, suggesting instead that general perceptual principles can be invoked to perceive distal speech events.

Unfortunately, although the direct-realist approach avoids the pitfalls of modularity, the approach has difficulty with the present results at a more fundamental level. The central hypothesis of Gibsonian perception is that perception is *direct,* meaning that perception occurs without any cognitive analysis or mediation. Without this basic assumption, the theory loses its distinctive appeal. This basic assumption is clearly incompatible with the finding that episodic traces affect speech perception: By definition, a theory that assumes that multiple memory traces affect perception is a theory of cognitive mediation of perception, not direct perception.

At this level of analysis, both the motor theory and the direct-realist theory are incompatible with the present results. Either of these theories, however, might be spared if the locus of the present effects is considered to be a stage of processing later than phonetic perception. Fowler might argue, for example, that the phonetic content of an utterance is directly perceived via gestural events, then episodic memory affects lexical matching after the phonetic content is specified. In other words, if one assumes a modularized demarcation of speech sound perception and matching inputs to the lexicon, perceptual episodes may affect spoken word recognition, but not speech perception. The present study provides no empirical refutation to this claim, although the strict demarcation between speech perception and spoken word recognition appears rather contrived (see McClelland & Elman, 1986).

Implications for Theories of Spoken Word Recognition

Since many speech perception theorists discriminate between speech sound perception and word recognition, the present data may be considered more relevant to theories of lexical representation and spoken word recognition than to theories of speech perception. As discussed above, the major finding of the present investigation is that spoken word recognition is influenced by specific perceptual episodes in memory. As Jacoby and Dallas (1981; see also Jacoby, 1983b) argues, these findings suggest that lexical representations may be based on multiple perceptual traces, rather than purely abstract, symbolic, units. Most models of word recognition assume that the signal is analyzed into segmental constituents, then this sequence is compared to canonical lexical representations in memory. If a sufficient match is found, the stimulus is "recognized." The recognition process is typically characterized in terms of a unit or node crossing a threshold, an entry being tagged, or a candidate emerging from its nearest competitors (see below). Several contemporary models of word recognition are reviewed below, all typical of the current literature, and all sharing several key constructs:

Logogen Theory

In Morton's (1969, 1979, 1982) logogen theory, passive sensing devices called "logogens" are associated with each word in the mental lexicon. Each logogen contains all information about a given word, such as its meaning, syntactic functions, and its phonetic and orthographic structure. Logogens monitor discourse for any information indicating that their respective words are present in the signal. Once such information is encountered, the activation levels of all consistent logogens are increased. Upon sufficient activation, the proper logogen crosses a threshold, at which time the word is said to be "recognized."

The specific details of logogen theory have changed somewhat over the years. For example, although Morton (1982) divided the logogen system into separate visual and auditory subsystems, the fundamental notion of passive threshold devices that monitor information various sources has remained. An aspect of the theory that has never changed is the emphasis on pure abstraction in lexical representation. Indeed, Jackson and Morton (1984) report repetition effects that were similar to the present data, but with no significant effects of voice change on recognition. They argue that word recognition is mediated by abstract logogens that are indifferent to the physical variations across input stimuli. The present data contradict these findings; repetition effects were strongly mediated by the physical variations across processing episodes. As Jacoby and Hayman (1987) argue, such episodic effects on perception suggest that perception relies on detailed representations, rather than abstract logogens.

Forster's Autonomous Search Theory

Forster's (1976, 1979) search theory of word recognition assumes autonomous, serial processing in the lexicon. Forster's model may be considered the word-recognition embodiment of Fodor's (1983) modularity principles (see Forster, 1989, 1990; Tanenhaus & Lucas, 1987), emphasizing algorithmic, non-interactive processing in separate, hierarchically related components. The language processor is composed of autonomous, non-penetrable modules. The lexical processor is independent of the syntactic and message processors, and the entire linguistic system is independent of the general cognitive system, as Fodor (1983) suggests.

Word recognition occurs in Forster's model in several stages: In the first stage, information from peripheral perceptual systems is submitted to the lexical processor. The lexical processor attempts to locate a matching entry in three peripheral access files: an orthographic file for visual input, a phonetic file for auditory input, and a syntactic-semantic file for either form of input. The peripheral access files are submitted to a frequency-ordered search, with higher-frequency words being compared to the input stimulus before lower-frequency words. Word recognition is accomplished at the level of the peripheral access files, where the input pattern is matched to a stored representation. Once the entry is thus located, lexical access is accomplished by locating the entry in the master lexicon, where complete lexical information is stored.

Forster's search model borrows its architecture from the design of computer databases. Like logogen theory, the search model assumes representations that are completely abstract and independent of episodic memory. Therefore, like logogen theory, the search model has no mechanism to account for the present findings. Even if it is assumed that every recognized word leaves a unique trace in memory, the model explicitly assumes that lexical processing is independent of other memory processes.

Activation-Verification Model

In the activation-verification model of word recognition (Paap, Newsome, McDonald, & Schvaneveldt, 1982; Becker, 1976; 1980; Becker & Killion, 1977), presentation of a stimulus word activates a pool of lexical candidates in memory. The activation of lexical candidates is based on coarse sensory analysis; activation is isomorphic to similarity with the input. The activated candidates are then subjected to a *serial verification* process, in which each candidate word is compared to the stimulus until a best match is determined. The verification process is similar to the search process in Forster's model; candidates are submitted for verification in descending order of frequency.

By including the concepts of activation and search, the activation-verification model represents a hybrid of the logogen and search theory approaches. Unlike these earlier models, however, the verification model is more explicit with respect to the origins of lexical representations. Becker (1976) assumed that words are represented as prototypes "composed of information abstracted from previous encounters with a given word" (page 564). As such, the verification model takes a step toward incorporating the information necessary to account for the results of the present investigation. With relatively minor modification, the verification model would resemble several episodic models proposed in the literature (see below), and might comfortably account for the effects of perceptual repetition reported here.

Cohort Theory

Marslen-Wilson's cohort theory (Marslen-Wilson, 1975, 1980b, 1987; 1989; 1990; Marslen-Wilson & Tyler, 1975, 1980; Marslen-Wilson & Welsh, 1978) assumes that acoustic-phonetic information at the beginning of an input word activates all words in memory that have the same word-initial information. These activated words constitute a "cohort," which is then pared down to a single candidate. Once a cohort is activated, all possible sources of information come to bear on the selection of the presented word. This information includes further acoustic-phonetic information as well as syntactic, or semantic information.

An important feature of cohort theory is its sensitivity to the temporal nature of speech. As such, it gives priority to the beginnings of words and assumes strict left-to-right processing of acoustic-phonetic information (see also Cole & Jakimik, 1980). Cohort theory also embraces the notion of "optimal efficiency" (Grosjean, 1980; Marslen-Wilson, 1980a, 1987; Tyler & Marslen-Wilson, 1982; Tyler, 1984). This principle states that the word recognition system selects a single candidate from the cohort at the earliest possible moment (the "isolation point"). This means that the system will commit to a decision as soon as sufficient acoustic-phonetic and higher-level sources of information are consistent with a single candidate, or at least until one candidate sufficiently diverges from its competitors (Marslen-Wilson, 1990). (Note, however, that Goodman and Huttenlocher (1988) and Taft and Hambly (1986) have shown that lexical decisions are not reliably predicted by isolation points.)

The nature of lexical representation in cohort theory is vague. Marslen-Wilson (1987, 1989; Warren & Marslen-Wilson, 1987) has argued that the theory requires no conventional linguistic units, such as phonemes, in order to function. To maintain optimal efficiency, he proposes that the word recognition system exploits coarticulatory information that crosses phonemic boundaries (e.g., nasalization of a vowel that precedes a nasal consonant) in real-time, thus avoiding unnecessary decisional delays. As Warren and Marslen-Wilson (1987) write:

> The continuity of the articulatory gestures underlying the production of speech should be reflected in a truly continuous projection of the speech input onto the mental lexicon, rather than the discontinuous process imposed by the requirements of phonemic categorization. (Page 262).

If recognition were attempted via comparison to multiple episodic representations, this non-segmental processing assumption might be a viable recognition strategy. As the cohort model currently stands, however, no such explicit representational assumptions are provided.

Neighborhood Activation Model (NAM)

The neighborhood activation model (NAM) of word recognition (Luce, 1986; Luce, Pisoni, & Goldinger, 1990; see also Goldinger, Luce, & Pisoni, 1989) assumes that word recognition reduces to the choice of a "best match" from a pool of activated, competing word candidates. Central to the model is the concept of *similarity neighborhoods*, defined as collections of phonetically similar words in the mental lexicon (Luce, 1986; see also Andrews, 1989; Coltheart, Davelaar, Jonasson, & Besner, 1977; Landauer & Streeter, 1973).

Word recognition in NAM is similar to both logogen theory and cohort theory, with two key modifications. NAM assumes that, upon stimulus input, a neighborhood of acoustic-phonetic patterns are activated in memory. The activation levels of these patterns are isomorphic to their phonetic similarity to the stimulus input. The activated phonetic patterns activate, in turn, a system of *word decision units*, conceptually similar to logogens. Word decision units are activated directly from bottom-up information in the signal, as in cohort theory. Once activated, word decision units monitor several sources of information, especially the fluctuating activation levels of the acoustic-phonetic patterns. To account for effects of similarity neighborhoods, NAM assumes that word decision units also monitor the overall level of activity in the decision system, in a manner similar to the TRACE model (Elman & McClelland, 1986; McClelland & Elman, 1986; see below). Word recognition occurs in NAM when the system of word decision units selects a best match from the activated neighborhood.

Given the abstract nature of representation in word decision units, NAM has trouble accounting for the present results. Like logogen theory, NAM seems to require some mechanism to incorporate specific perceptual details into lexical representations. Otherwise, the model has no means to explain the effects of such details on recognition.

Klatt's LAFS Model

Although Klatt's (1979; 1981) Lexical Access From Spectra (LAFS) model is included in this discussion on models of spoken word recognition, it is important to acknowledge that LAFS is actually a model of speech sound perception. LAFS is one of the few models that addresses issues in speech perception while simultaneously addressing access to lexical representations in long-term memory.

LAFS assumes direct, noninteractive access to lexical entries based on context-sensitive spectral sections (Klatt, 1979). The LAFS model assumes a precompiled, acoustic lexicon of all possible words in a network of diphone power spectra. These spectral templates are assumed to be context-sensitive units, similar to "Wickelphones," that represent the acoustic correlates of speech sounds in different phonetic environments (Wickelgren, 1969). The collection of multiple spectral representations is proposed to resolve problems of contextual variability in individual segments: LAFS resolves the problems of variability by precompiling coarticulatory effects directly into representations in memory. The listener is assumed to compute spectral representations of an input word and compare these derived spectra to prototypes in memory. Word recognition is achieved when a best match is found between the input spectrum and a prototype. Word recognition is therefore accomplished directly via spectral representations of the input, with no intermediate units corresponding to phonemes.

Klatt's model takes a step toward addressing the present data that purely abstract models, such as logogen theory, do not. LAFS embodies two assumptions that a more purely episodic model (e.g., Hintzman, 1986; Jacoby & Brooks, 1984) would also entail. First, LAFS alleviates the problems of stimulus variability by representing variability directly in the lexicon. Thus, the normalization requirement is relaxed, and potentially useful information is not discarded. Second, LAFS bypasses information reduction by assuming that recognition occurs via global matching of stimuli to representations. Both of these assumptions are consistent with the findings of the present investigation. Nevertheless, LAFS still assumes representations far more abstract than specific perceptual episodes. Like almost all other theories of word recognition, LAFS assumes that recognition occurs via matching inputs to representations in memory, but no encoding of redundant representations is assumed. Moreover, the representations in LAFS are sublexical units; the present data suggest that entire lexical episodes subserve later perception. Finally, the representations in LAFS incorporate variability due to phonetic environments, but they do not anticipate variability across speakers, speaking rates, environmental noise, etc. Klatt (1979) speculated that speaker variability could be tolerated with enough spectral templates stored in memory, but he did not consider some other important sources of variability.

TRACE and other Connectionist Models

The TRACE model of speech perception, like Klatt's LAFS model, is also concerned with both speech perception and word recognition. TRACE (Elman, 1989; 1990; Elman & McClelland, 1986; McClelland & Elman, 1986) is a completely interactive system. Developing out of the connectionist movement, and based on the interactive-activation model of visual word recognition (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982), TRACE assumes multiple levels of representation and numerous connections between processing units. TRACE also incorporates facilitative and competitive activation of units in the network, as in the earlier interactive-activation framework.

The functional units in TRACE are simple, highly interconnected processing units called nodes. There are three levels of nodes in the model: Some nodes represent phonetic features, some represent phonemes, and some represent words. When information passes upward through these levels, nodes that collect sufficient confirmatory evidence to surpass a threshold will fire and send activation along weighted links to related nodes. In this manner, information consistent with the "expectations" of the early feature detectors is proliferated upward in the network to encourage recognition of associated phonemes, and then recognition of the phonemes encourages recognition of associated words.

The architecture and interactive nature of TRACE are conceptual innovations that offer much to theories of speech perception and spoken word recognition. McClelland and Elman (1986) consider almost a dozen well-known phenomena from the speech perception literature that the model can simulate, as well as several findings from the word recognition literature. As a model of speech perception, TRACE does not treat coarticulatory effects of speech as "noise" imposed on an idealized string of phonemes. Instead, Elman and McClelland (1984; 1986) refer to "lawful variability," which serves as a rich source of information in TRACE (see also Nakatani & Dukes, 1977; Wickelgren, 1969). (Notably, the model has no means to tolerate speaker variability). Also, although the model assumes segmental representations in speech, no explicit segmentation is imposed in processing. Instead, phones and allophones are assumed in the model's architecture, and segmentation "falls out" naturally. As a model of word recognition, the inhibitive links among nodes at the lexical level allows TRACE to account for neighborhood effects in a manner similar to the neighborhood activation model. By virtue of its simple architecture and processes, TRACE captures many attributes of speech perception and spoken word recognition in an integrated system, without proliferating rules or specialized mechanisms.

With respect to the results of the present investigation, however, TRACE displays the same shortcomings as other models incorporating representational units such as logogens, word decision units, or file entries. The nodes in the TRACE network (and in the visual interactive-activation network) are abstractions, conceptually similar to logogens. It has no mechanism to encode or use detailed perceptual episodes. As such, like logogen theory, TRACE has no mechanism to account for the long-term, specific repetition effects observed in the present experiments.

Episodic & Non-analytic Models of Word Recognition

The introduction to this investigation began with a review of Richard Semon's theory of memory (1909/1923; Schacter, Eich, & Tulving, 1978). This early and influential model was predicated on the assumption that every experience leaves a unique trace in memory. These traces could then support memory for very specific experiences, or could act in harmony to support abstract cognition. For example, consider Semon's theory with respect to speech perception and production: By assuming separate, detailed traces, Semon's theory could account for the specific repetition effects observed in the present investigation; the exact details of the stimulus would access the identical trace in memory. For speech production, Semon's theory might assume that all instances of a word are accessed in memory; their physical idiosyncracies fall away and only their shared properties (e.g., lexical identity and semantic content) become available to consciousness. The previous sections reviewed a number of prominent theories of speech perception and spoken word recognition, all of which assume abstract representations and information-processing operations. In this section, we come full-circle, reviewing theories that embody the concepts Semon introduced at the turn of the century.

The progression of topics in the present discussion is representative of the current literature; the ideas of exemplar representations, detailed memories, and models combining specific and abstract information seem to be experiencing a resurgence. Underwood's (1969) article about retention of specific attributes was well-received, but was published in the same era as many influential books and articles (e.g., Miller, 1962; Neisser, 1967; Posner, 1969) that stressed the concepts of information reduction, abstraction, and economical storage of symbolic memory codes (see Tulving & Bower, 1974). In the recent literature, however, a growing number of articles stress the fine details of memory representation, and also the complex roles of these representations display across behaviors.

With respect to detailed memories, an ample review of the data was provided in Chapter I. With recent interest in implicit versus explicit memory, however, research into and theoretical importance of detailed memories has increased dramatically across many areas. Alba and Hasher (1983) carefully outline a schematic (or prototype) model of memory, including elements of many models that have shaped the field. The derived model assumed abstract, holistic representations and specialized processes that match stimuli to these representations. In an impressive review of the empirical literature, Alba and Hasher found that memory is far more detailed than such a theory would predict and that the schema theory could not possibly account for a surprising range of results.

More recently, the influences of specific experiences on memory have been catalogued and interpreted by many researchers: Smith (1990; Smith & Zarate, 1992) has proposed a theory of social judgment and attribution based on memory for specific social interactions. Smith borrows the MINERVA 2 framework (Hintzman, 1986; 1988) described below, demonstrating the versatility of the multiple-trace approach to various domains. Logan (1988a; 1988b) proposed a theory of attentional automaticity based on the collection of multiple instances in memory. Logan's theory states that automaticity develops when one changes from processing stimuli algorithmically to relying on memory for prior episodes. Recently, Logan (1990) has extended his instance-based model to account for repetition effects such as the effects observed in the present investigation.

Episodic accounts are emerging in many other domains as well. With respect to memory for novel, visual patterns, Musen and Treisman (1990) concluded:

> To account for explicit memory performance, we suggest that... a separate episodic token or exemplar is laid down for each particular occurrence of each pattern. (Page 136).

In a recent review on speaker variability effects in speech perception, attention, and memory, Pisoni (1990) concluded:

> Our findings indicate that speech signals are not encoded into a canonical segmental representation consisting of a string of idealized phonemes. The representations appear to be very detailed; they appear to be highly context-sensitive and they apparently preserve a great deal of information carried in the speech signal. (Page 1406).

These quotes provide only a small sample of the recent growth of interest in these topics (for review, see volume edited by Lewandowski, Dunn, & Kirsner, 1989, or Richardson-Klavehn & Bjork, 1988). Another important trend in the recent literature is the movement to relate performance across behavioral domains via common memory representations. For example, Nosofsky (1984; 1986; 1988b) has systematically extended the domain of his exemplar model of classification to account for findings in identification, rule learning, and recognition memory. With respect to classification and recognition memory, Metcalfe and Fisher (1986) reported an independent relationship between the processes, suggesting that they rely on separate memory systems. However, Nosofsky (1988b) replicated the orthogonal results of Metcalfe and Fisher (1986), and then proceeded to fit the disparate sets of data with a single model of exemplar representation. The model assumed what Nosofsky called a "common representational substrate" for all traces; different decision rules, optimized for the subject's task, were applied to the same set of exemplars to yield close fits of theory and data.

Research such as Nosofsky's (1988b) investigation has clear relevance to the present research. The present investigation, and numerous recent investigations, demonstrate that memory for a word affects not only memory tasks, but word recognition as well. As Jacoby and Brooks (1984) and others have argued, the present results suggest that a common set of episodes in memory are involved in perception, classification, and recognition memory. Following the lead of these previous researchers, I suggest that a subset of episodic traces corresponding to words constitutes the entity commonly known as the "mental lexicon." Unlike the collection of dictionary entries the term implies, many recent results (including the present results) suggest that the mental lexicon consists of countless episodic traces. In the following discussion, several episodic models of the lexicon are described. The models vary in their assumptions regarding levels of abstraction, and their assumptions regarding access processes, but all acknowledge the importance of specific episodes in perception.

Jacoby's Contributions

A great deal of the research and ideas underlying the theories discussed below may be attributed to Jacoby and his colleagues (Brooks, 1987; Jacoby, 1983a; 1983b; Jacoby, Baker, & Brooks, 1989; Jacoby & Brooks, 1984; Jacoby & Dallas, 1981; Jacoby & Hayman, 1987; Jacoby, Levy, & Steinbach, 1992; Jacoby & Witherspoon, 1982; Johnston, Dark, & Jacoby, 1985; Whittlesea, Jacoby, & Girard, 1990) Although Jacoby has never espoused a specific model of lexical memory or the mental lexicon, his ideas are reflected in many explicit models in the literature.

A basic premise in Jacoby's work concerns the effects of specific perceptual episodes across a variety of tasks. Jacoby and Brooks (1984; see Brooks, 1987) argued that perception, recognition memory, classification, and generalization all rely on memory for episodes. In terms of word recognition, Jacoby (1983b; Jacoby & Brooks, 1984) asserted that recognition may occur via "non-analytic" means; a stimulus may be recognized by direct similarity to prior episodes, rather than by decomposition into canonical features. Jacoby and Brooks argue that the concept of long-term repetition effects, such as the effects reported here, are consistent with memory for episodes, rather than "priming" of abstract units such as logogens. Moreover, Jacoby has frequently argued that the degree of agreement in the processes invoked at study and test mediate the degree of episodic effects in perception and memory (see also Roediger & Blaxton, 1987). Memory may influence performance unconsciously when tasks are supposedly unrelated (Jacoby & Witherspoon, 1982); no introspective awareness of prior episodes is required for them to affect later processing.

Like Logan (1988a; 1990), Jacoby assumes that processing via episodes is more efficient than analytic processing. Jacoby and Brooks (1984) note the developmental data on children's judgments of perceptual similarity to support their point: Although adults attend to specific dimensions in judging similarity between objects, children rely on overall similarity (Kemler, 1983; Smith, 1989). Jacoby and Brooks speculate that, in the absence of further learning about perceptual dimensions, the children's non-analytic approach is the easiest solution to the task. With enough experience and episodes in memory, the relevant dimensions across stimuli should emerge as abstract properties, as in Semon's theory.

It is important to note that, despite the myriad influences of specific episodes in memory and perception, it is not necessary to assume that perception or recognition memory always relies on the most similar available episode (Brooks, 1987). As in Semon's theory, Jacoby assumes that numerous similar items may be accessed by presentation of a stimulus; processing of the stimulus will reflect the sum of many traces. In effect, this harmonious processing approximates abstract or schematic processing (see Hintzman, 1986).

If there is an apparent shortcoming with Jacoby's approach, it is the lack of an explicit theoretical framework. As mentioned above, Jacoby has never proposed a specific processing model to embody the range of his theoretical ideas. Indeed, Jacoby, Marriott, and Collins (1989) write:

> As discussed by Barsalou (1990)... exemplar and abstractionist theories can be made to mimic one another in ways that make it impossible to choose between them by traditional methods of theory testing. This does not bother us because we have very little interest in testing formal theories in a traditional way. Rather, we are trying to produce a program of research that will show the advantages of thinking in terms of memory for prior episodes. (Page 120.)

It would be difficult to refute Jacoby's approach; one should not argue with success. Nevertheless, the value of specific theories is also difficult to refute. As Salasoo, Shiffrin, and Feustel (1985) note, intuitions about the behavior of complex systems are notoriously suspect. Nosofsky (1988b) could not have discovered the relation between classification and recognition without a formal model to guide the comparison. Perhaps Hintzman (1990) put the sentiment most eloquently:

> Formal (i.e. mathematical or computational) theories have a number of advantages that psychologists often overlook. They force the theorist to be explicit, so that assumptions are publicly accessible and the reliability of derivations can be confirmed.... To have one's hunches about how a simple combination of processes will behave repeatedly dashed by one's own computer program is a humbling experience that no experimental psychologist should miss. Surprises are especially likely when the model has properties that are inherently difficult to understand, such as variability, parallelism, and nonlinearity-- all, undoubtedly, properties of the brain. (Page 110-111).

Like Jacoby's position, Hintzman's position is easily understood. Indeed, a computer simulation model, like Hintzman's (1986; 1988) MINERVA 2, does guide one's thinking about phenomena in ways that introspection cannot. At the same time, an open mind, free from the concerns of parameter estimation, is perhaps more likely to conjure an experiment that reveals a counter-intuitive result. Johnston, van Santen, and Hale (1985) propose that, especially in exploratory domains, testing models for precise fits to data may be premature; models should serve to distinguish between broad classes of possible explanations for particular phenomena. The remainder of this discussion includes theories specified to various levels of detail. They should all be considered equally important contributions, regardless of their levels of formal development.

Perceptual Representation Systems (PRS)

For many years, memory theories have proposed separate memory systems that are responsible for different functions. The most commonplace distinction is the separation of short- and long-term memory (e.g., Atkinson & Shiffrin, 1968; Glanzer & Cunitz, 1966). Later distinctions include semantic versus episodic memory (Tulving, 1972; 1983), and declarative versus procedural knowledge (Anderson, 1983). Sherry and Schacter (1987; see also Schacter, 1987) argue that multiple memory systems, such as the division of implicit and explicit memory, are the inevitable consequence of natural selection-- the range of functions an organism must perform logically requires numerous, specialized capacities.

Tulving, Schacter, and Stark (1982) reported a dissociation between recognition memory and word-fragment completion. Their data were not easily explained by reference to either the episodic or semantic memory systems that Tulving (1983) supported. Tulving et al. (1982) ended their article with the following speculation:

> Since the priming effects described in this article clearly are independent of episodic memory, and since there are problems with their interpretation in terms of modifications of semantic memory, we are tempted to think that they reflect the operation of some other, as yet little understood, memory system. (Page 341).

Recently, in response to findings in implicit versus explicit memory, Schacter and his colleagues (Schacter, 1990; Tulving & Schacter, 1990) have proposed a new memory system called the *Perceptual Representation System* (PRS). PRS is assumed to underlie purely perceptual repetition effects, such as the implicit memory findings of the present investigation. Tulving and Schacter (1990) assert that laboratory demonstrations of repetition effects are representative of a set of commonplace mental occurrences: It is assumed that fluent recognition of everyday objects and symbols is an expression of PRS. The reason for its late introduction into theoretical debate is that repetition effects reflect the unconscious influences of memory. Indeed, if it were not for an experiment reported by Warrington and Weiskrantz (1968), current interest in repetition effects and separate memory systems may have never developed. Warrington and Weiskrantz discovered that severely amnesic patients could not explicitly remember recently seen information, but displayed implicit savings to a near-normal degree.

Tulving and Schacter (1990) list five sources of evidence to support the PRS hypothesis, all involving memory dissociations. Included were 1) evidence of intact repetition effects in amnesics; 2) evidence that explicit recognition develops with age, but repetition effects are relatively independent of age (Parkin & Streete, 1988); 3) evidence that alcohol impairs explicit memory, but has little or no effect on implicit memory (Hashtroudi, Parker, DeLisi, Wyatt, & Mutter, 1984); 4) independence of implicit and explicit memory in normal subjects (e.g, Tulving et al., 1982); and 5) independent effects of the same stimulus on different memory tests (e.g., Musen & Treisman, 1990). All of these dissociations suggest that repetition effects and explicit recognition have separate underlying bases.

Tulving and Schacter argue that PRS is the system used specifically for identification of perceptual objects, including words. PRS contains highly specific representations of perceptual forms, complete with all irrelevant details intact. The representations in PRS are long-lasting. PRS employs no measure of abstraction, so its effects are highly specific and inflexible; explicit memory performance (assumed to rely on more abstract episodic information) does not predict implicit memory performance. Finally, PRS is a pre-semantic system, so its effects are independent of conceptual elaboration on memory traces, such as a levels of processing manipulation (e.g., Jacoby & Dallas, 1981).

The attributes of PRS listed above bear close resemblance to the major findings of the present investigation: Implicit and explicit memory were partially dissociated, implicit memory was long-lasting and was insensitive to levels of processing. Similar findings were reported by Schacter and Church (in press), and were also interpreted as support for the PRS framework. Taken together, the review of evidence cited by Tulving and Schacter (1990) and the present findings lend considerable support to the PRS framework.

Despite the impressive agreement of the PRS framework with past and present findings, the approach has some weaknesses as well. First, the framework lacks explanatory power in much the same way as the modularity hypothesis. Faced with a set of seemingly dissociated phenomena, Tulving and Schacter hypothesized separate memory systems for each. Although the delineation of systems finds support from many unrelated lines of evidence, it basically re-describes the data, rather than predicts it. Even the new findings of Schacter and Church (in press) showing that voice-specific repetition effects were independent of levels of processing could be considered replicative of previous findings.

Second, Tulving and Schacter's interpretation of memory dissociations are too strong. Jacoby (1983a; 1983b; 1990; Jacoby & Dallas, 1981) has repeatedly shown that implicit and explicit memory measures are only partially independent; both expressions of memory are susceptible to many procedural manipulations. The theory of *transfer-appropriate processing* (Masson, 1989; Roediger & Blaxton, 1987) argues that implicit and explicit memory will mirror each other to the degree that they involve similar processes at study and test. Numerous demonstrations by Jacoby, Roediger, and others suggest that the strict demarcation of PRS from the rest of memory may be unwarranted.

Recently, Schacter (1990) addressed these and other criticisms. He argued that the PRS is not hypothesized to underly all implicit memory phenomena, nor is it hypothesized to work independently of other memory systems. Given these disclaimers, however, the original reasons for developing the PRS framework (the list of strong dissociations reviewed by Tulving & Schacter) lose much of their potency. For the present, the PRS hypothesis promises to serve the ultimate goal of many theories-- the stimulation of much concentrated research in new and unexplored areas-- but it will be difficult to establish PRS as a truly independent branch of memory.

Hintzman's MINERVA 2

If logogen theory is the "gold standard" of abstractionist theories, then Hintzman's MINERVA 2 is the gold standard of exemplar theories. Hintzman's model takes the concept of exemplar storage to its logical extreme, assuming that every perceptual experience creates an independent, analog trace (Hintzman, 1986; 1988; Hintzman, Grandy, & Gold, 1982; Hintzman & Ludlam, 1980; Hintzman, Nozawa, & Irmscher, 1982). With the multiple-trace assumption and several processing assumptions, MINERVA 2 is able to account for repetition effects such as the present data, as well as an impressive array of related findings from the categorization and memory literature.

MINERVA 2 embodies several assumptions that we have already discussed in regard to Semon's theory of memory (although MINERVA 2 was derived independently). Hintzman's model assumes that memories of specific events are stored as independent collections of "primitive properties," including perceptual details, affect, context, semantic connotation, etc. Despite their separate storage and particular attributes, aggregates of traces activated together at retrieval represent categories as a whole. Thus, the model accounts for specificity and generality of recognition via the same set of exemplar traces.

Recognition in MINERVA 2 occurs as follows: Assume that a vast collection of traces is collected in memory. Although all traces contain some idiosyncratic information, many share important properties, such as category names, contextual information, etc. When a new stimulus is presented, a probe containing all the information in the stimulus is communicated in parallel to all traces in long-term memory. All traces in memory, and all of their primitive properties, are activated in proportion to their similarity to the probe, so some traces will strongly respond and others may not respond at all. The sum total of activated traces constitutes an "echo" sent to consciousness from long-term memory. The echo may contain information not represented in the probe itself, thus providing a mechanism to associate the new probe with past information in memory.

The echo returned from long-term memory to consciousness is described in terms of two qualities: intensity and content. The intensity of an echo is a function of the both the similarity of traces to the probe, and the number of traces contributing to the echo. Echo intensity is assumed to signal the familiarity of a stimulus (e.g., in an explicit recognition task). The content of an echo is a reflection of the distinctiveness of the probe with respect to the contents of memory. If the probe is similar to many traces in long-term memory, the echo will reflect primarily their common properties, which will tend to describe the central tendency of the category of traces. If the probe is strongly similar to few traces, however, the echo will reflect more idiosyncratic information.

In computer simulations of MINERVA 2, Hintzman demonstrates that the strong multiple-trace model predicts many patterns of data typically considered hallmarks of abstract representations (Hintzman, 1986; 1988; Hintzman & Ludlam, 1980). For example, MINERVA 2 re-creates the classic results of Posner and Keele (1968) investigating the abstraction of a prototype of dot patterns. Moreover, because MINERVA 2 assumes that information is randomly lost from traces over time, it also correctly predicts that the prototype will be remembered longer than particular exemplars (Posner & Keele, 1970). In addition, Hintzman shows that the model accounts for numerous findings from the memory literature, including list length effects, orienting task effects, mirror effects, and memory for frequency. In short, MINERVA 2 accounts for data from a variety of behavioral domains using a common set of exemplar traces (see also Nosofsky, 1988b).

As an example of recognition in MINERVA 2, assume that multiple, detailed traces of visual words are stored in memory. When a familiar word is presented in a common typeface, many similar traces should strongly respond in parallel. In this condition, even if exact matches to the probe exist in memory, the large number of similar traces should still yield a non-specific echo. When an unfamiliar word is presented in an unusual typeface, however, few similar traces should respond, or perhaps many traces will weakly respond. In this condition, the presence of an exact match to the probe in memory should contribute significantly to the echo. Therefore, effects of specific repetition should be greater for unusual stimuli or stimuli presented in unusual contexts. This qualitative description is clearly compatible with the findings of Kolers and others with respect to memory for transformed text.

To account for the present results, MINERVA 2 need only assume that the experimental context contributes structure to the traces created during study. Hintzman (1988) argues that with contextual encoding, the effects of extra-experimental traces can be reduced to approximately zero. This is also a common assumption in other recognition models, such as SAM (Gillund & Shiffrin, 1984). With this assumption, the model predicts that, although all previous experiences with stimulus words will contribute to their recognition, the traces created during study will most closely match the probes presented at test.[2] The effects of specific voice details are easily predicted once the set of traces contributing to recognition is thus delimited.

It is important to note that, as it is currently described, Hintzman's model is clearly not a model of the mental lexicon. The model, as published, is more concerned with explicit recognition memory than the expression of memory via perception. Nevertheless, Hintzman (1986) discusses some of the model's potential solutions to problems in word recognition. For instance, the model stresses context-sensitive interpretation of stimuli, which would encourage the appropriate recognition of polysemous words, such as homophones. With proper modification of several assumptions, MINERVA 2 could be applied as a model of the lexicon, and would almost certainly predict a realm of frequency, neighborhood, and priming effects. Work to implement and test such a model via computer simulation is currently underway.

It is also important to note that Hintzman's model is certainly not the only multiple-trace model that could be applied to the present results. The context model of Medin and Schaffer (1978; see also Nosofsky, 1988a) also assumes that recognition occurs via summed similarity to all exemplars in memory. As such, the context model shares many predictions with Hintzman's model. Ratcliff's (1978) random-walk model also shares

---

[2] In this connection, an interesting prediction for later research is that the specific voice effects observed in the present investigation should be even stronger if pseudowords are presented as stimuli (see Salasoo, Shiffrin, & Feustel, 1985).

many basic constructs with Hintzman's model. Gillund and Shiffrin's (1984) SAM model for recognition assumes multiple traces that are accessed in proportion to their similarity to the probe. Although SAM and MINERVA 2 differ in their approach to information loss (retroactive interference in SAM versus information loss in MINERVA 2), both models share many predictions, and both appear consistent with the present results.

Finally, it is important to acknowledge that several non-exemplar models may also account for the present results. The requirements of the present data are that a model should preserve the specific perceptual details, and support the abstract linguistic functions, of words. Many non-exemplar memory models incorporate specific information into general representations. For example, the matrix models proposed by Pike (1984) and by Humphreys, Bain, and Pike (1989) and the convolution-based CHARM (Eich, 1982) and TODAM models (Murdock, 1982) all assume that multiple traces are combined in memory during learning, rather than being summated when probed. Despite the combination of traces, however, if these models maintain that any information unique to a given stimulus is preserved and available for recognition, they may predict specificity effects such as the present findings. Although Hintzman's model may be preferred for the elegant simplicity of its assumptions, any of these models may be able to mimic its predictions in one or more behavioral domains.

Hybrid Representation Model

Feustel, Shiffrin, and Salasoo (1983; also Salasoo, Shiffrin, and Feustel, 1985) have described a model of lexical representation and access that constitutes a compromise between pure abstraction and exemplar models. The aspects of their model were motivated by complex patterns of data on repetition effects on word and pseudoword identification. Feustel et al. (1983) examined word and pseudoword identification with a threshold estimation procedure: Recognition of words was superior to nonwords, and the difference between them was additive with number of repeated exposures. Notably, pseudowords were subject to large repetition effects after only 1 exposure. Indeed, after just a few repetitions, thresholds to identify pseudowords equaled those to identify words. Numerous other findings were reported, including a small decrease in repetition effects when case of type was changed (significant only for pseudowords).

In later research, Salasoo et al. (1985) examined effects of repetition on words and pseudowords in the short-term, and after a 1-year delay. Accuracy of identification improved for all stimuli with repetition. Words showed an identification advantage over pseudowords initially, but, after about five repetitions, pseudowords were identified as well as words. Subjects returned a year later, and were presented old words, new words, old pseudowords, and new pseudowords. The repetition advantage for words was no longer evident-- old words and new words were identified equally well. For pseudowords, however, this was not the case. Identification of old pseudowords was equal to old and new words, all of which were superior to new pseudowords.

Taken together, the results observed by Feustel et al. (1983) and Salasoo et al. (1985) suggested that two forms of lexical representation underly repetition effects, and by logical extension, word recognition. The authors propose a model of perceptual identification in which both abstract lexical codes and episodic traces contribute to identification. Presence of lexical units confers an advantage in identification. Repetition effects, however, are attributed to episodes of prior presentations. According to Salasoo et al., the pseudowords presented during training became *codified*, meaning that multiple episodes collect into a unitized code, similar to the codes for words hypothesized in a logogen theory. Since the complete pattern of results suggested influences of both episodic traces and codified units, Salasoo et al. suggested a hybrid model of lexical representation and memory. Like other models with episodic components, the hybrid model assumes that similarity of new stimuli to encoded episodes (including contextual information) determines performance, but codified representations provide stability and permanence to the lexicon (see also Monsell, 1991).

The hybrid model of Salasoo et al. is consistent with the present results. The most important aspect of an appropriate model for these data is the retention of episodic traces. A possible objection to the hybrid model, however, is the inclusion of unitized codes alongside episodic traces. This dual representation is "hardwired" into the model, but other models (e.g., distributed memory models or MINERVA 2) suggest these codes may arise naturally from the set of episodes. Moreover, the inclusion of exemplars plus prototypes makes the model maximally powerful and difficult to reject.

Finally, one aspect of the Salasoo et al. procedure makes evaluation of the codification hypothesis difficult. The critical finding in support of codification was the savings for old pseudowords, relative to all other stimuli. As in other models of episodically-mediated repetition effects, the hybrid model assumes that episodes include contextual information. Presumably, the words shown to subjects in the experiment were seen in many contexts before the experiment, and in many contexts during the 1-year delay. Therefore, after a year, it would not be surprising if the contextual information from the training sessions was no longer a prominent feature of the total pool of episodes for the stimulus words. The pseudowords shown during training, however, were presumably never seen before or after the experimental sessions in any context. After a year, the episodic traces of the pseudowords should maintain the original context as a prominent feature. If this was the case, more residual repetition benefit would be expected for old pseudowords than old words, with or without a codified representation. It would be interesting to replicate the Salasoo et al. procedure, including a context shift for half of the subjects when they return a year later. If the advantage for old pseudowords appeared in both the old and new contexts, the codification hypothesis would be strengthened.

Distributed Memory Models

An interesting alternative to strict exemplar models are *distributed memory models* (e.g., Hinton & Anderson, 1981; Kohonan, Oja, & Lehtio, 1981; Knapp & Anderson, 1984; McClelland & Rumelhart, 1985). McClelland and Rumelhart's (1985) model of distributed memory is discussed because it has been applied directly to findings that are similar to the present results. In McClelland and Rumelhart's distributed memory model, memory traces are created by imposing a pattern of activation over a connectionist network. The precise patterns of activation for any particular experience is unique, and each specific trace can be retrieved via repetition of the complex activation pattern. At the same time, however, the model develops abstract representations, like prototypes or logogens. By the assumption of distributed representations, McClelland and Rumelhart's model develops abstract categories by encoding countless perceptual events, but its storage is far more economical than the localist MINERVA 2 model.

As McClelland and Rumelhart (1985) write:

> Our theme will be to show that distributed models provide a way to resolve the abstraction -- representation of specifics dilemma. With a distributed model, the superposition of traces automatically results in abstraction though it can still preserve to some extent the idiosyncracies of specific events and experiences... (Page 160).

It is not difficult to imagine how a distributed network would derive a central tendency from a set of individual exemplars. With all memory traces sharing the same set of nodes and connections, however, a natural question is whether repetition effects in the distributed memory model can display the extreme sensitivity to perceptual detail found in the present data. It seems that some regression to the mean of all traces is inevitable. Can the exact repetition of an old pattern of activation have a "special" effect, even after countless similar patterns have been combined into a common substrate?

To answer this question, several aspects of the model are important: First, McClelland and Rumelhart assume that all memory traces are subject to context-sensitive encoding, meaning that aspects of the experimental session are incorporated into the pattern of activation (as in Hintzman's model). Moreover, they assume that all stimuli leave slightly different traces behind. For a brief period, highly specific repetition effects, such as the findings in the present investigation, are predicted by the distributed memory model. McClelland and Rumelhart also suggest that the model can display long-term repetition effects, such as those reported by Salasoo et al. (1985), but they offered no explicit test of this prediction. Whether the model can account for the present finding of specific voice effects after a 1-week delay is also subject to debate. Presumably, as in MINERVA 2, if the contextual encoding sufficiently delimits the traces activated during test, such a pattern may be possible. This speculation, however, can only be properly evaluated by developing an appropriate simulation model.

The distributed memory model presents a reasonable compromise between the strict exemplar approach of MINERVA 2 and the separate-systems approach of the Salasoo et al. (1985) model. If the storage problem of strict exemplar models is a legitimate concern, distributed representations offer a viable solution. Moreover, if the representation of codified linguistic units is a necessary aspect of lexical processing, distributed representations provide a means for their derivation within a common memory substrate. Indeed, McClelland and Rumelhart argue that the distributed memory model constitutes a simpler version of the model proposed by Salasoo et al., (1985).

Record-Based Models

The models of lexical representation and access proposed by Feustel et al. (1983; Salasoo et al., 1985) and by McClelland & Rumelhart (1985) are both hybrid models; they assume the lexicon contains both episodic and codified representations. An alternative hybrid model, proposed by Kirsner and Dunn (1985; Kirsner, Dunn, & Standen, 1987) assumes that the lexicon contains abstract representations and episodic procedural records. Unlike Feustel et al.'s hybrid model, however, Kirsner and Dunn's model does not assume that prototypes and exemplars both affect recognition. The basic idea of Kirsner and Dunn's model is that every episode of word recognition involves perceptual processes that operate on the stimulus and compare it to lexical entries. Upon later word recognition, detailed records of past operations may be accessed and re-applied. Kirsner and Dunn label this process *record-based* word recognition.

The record-based model of word recognition assumes several different forms of perceptual processes, all of which contribute to the structure of perceptual records. All of the fine, specific details of a perceptual stimulus combine to lend structure to the record. In effect, the record-based approach is equivalent to other models assuming episodic traces of specific stimuli. With respect to repetition effects, for example, Kirsner, Dunn, and Standen (1987) write:

> The essence of our account is that word identification is achieved by reference to a record. Similarity is the critical parameter. If the record collection includes an example that is similar to the current stimulus description, identification will be achieved easily and quickly, even under difficult viewing conditions.... The character of the record collection can be determined by repetition priming. The basic argument is that the magnitude of repetition priming depends on the extent to which the current stimulus description matches a record. Systematic manipulation of the relevant parameters should reveal the precise character of a record. (Page 151).

The record-based approach to word recognition borrows the logic from many articles written by Kolers (Kolers, 1973; 1974a; 1974b; 1975a; 1975b; 1976a; 1976b; Kolers & Magee, 1978; Kolers & Ostry, 1974; Kolers, Palef, & Stelmach, 1980; Rudnicky & Kolers, 1984). Kolers repeatedly demonstrated readers' memory for transformed text, usually indexed via fluent re-reading. From all his findings, Kolers argued that memory for transformed text actually reflects memory for the specific perceptual operations applied by the reader to encode the passages. When identical transformed text is presented, the visual patterns activate records of previous operations, thus leading to fluent re-processing and savings in reading times. The perceptual operations are encoded in fine detail, so that even slight variants of the text eliminate the re-processing benefit.

The theory of memory for operations described by Kolers was intended to address encoding of strategic, conscious processes applied to a difficult perceptual problem. In their record-based model of word recognition, Kirsner and Dunn apply the notion of procedural records to all perceptual processes, regardless of difficulty or salience. For example, recognition of a spoken word in an unfamiliar voice should entail the operations of normalization and matching procedures. Upon later presentation of the same word in the same voice, the record of these previous operations will support more fluent processing, leading to a positive repetition effect. If a different word is presented in the same voice, the necessary perceptual operations will partially overlap with the previous record, so some residual savings will be observed, but the magnitude of the repetition effect will be diminished. With increased exposure to the speaker's voice (or typeface, rotated format, speech synthesizer, foreign accent, etc.), the increasing collection of available perceptual records will lead to asymptotic performance.

With respect to the results of the present investigation, the record-based approach serves as another viable alternative to a strict exemplar theory. Indeed, the strict exemplar MINERVA 2 model and the record-based model may make the same predictions across many domains. However, there are differences that may be important. For example, experimental or environmental context cues are stored into exemplars in MINERVA 2 or SAM simply because they exist in the perceptual array. These cues are clearly irrelevant to the perceptual operations of an abstracting system, such as in the record-based model. This difference may provide a diagnostic difference between the models. Precise comparison of two simulation models would be enlightening in this regard.

Theoretical Problems and Future Challenges

As is true for any general class of models, episodic models of the mental lexicon and word recognition have several important problems that require attention. Obviously, many problems at many levels of analysis could be enumerated, simply because of the nature of psychological modeling. Theories are cast in terms of autonomous systems and simple paradigms, in stark contrast to human cognition. Forgiving these global shortcomings of all models, I focus on several problems for episodic models that enter the realm of reasonable issues for future consideration. This is *not* intended as an exhaustive list of the final issues, or a recipe for de-bugging episodic theories. I have merely selected several issues that I find troublesome; some are already addressed in the literature to some extent; others are less frequently considered:

- *Intuitive Plausibility* Episodic models enjoy great intuitive appeal in considering repetition effects, especially long-term or perceptually specific repetition effects. By encoding the exact study stimulus into memory, one provides an obvious means to explain the precision of repetition effects. Nevertheless, the intuitive appeal of the exemplar approach, in general, is weak. The abstractionist dogma in theories of memory and language is certainly due, in part, to the "psychological reality" of abstract representations (e.g., Rosch, 1973). It is hard to accept that every instance of the word "the" is independently represented in memory. Clearly, some forgetting is involved, as Hintzman's model acknowledges. Still the idea is an affront to introspective perception; McClelland and Rumelhart (1985) praise the distributed memory approach for its compromise between the need for specific representation and its reasonable storage demands. It is difficult to know how seriously this problem should be taken. The history of science is a litany of counter-intuitive discoveries, from the discovery of the earth's circling the sun to diseases arising from microscopic organisms. As an example closer to the present topic, I am absolutely certain that early speech researcher at Haskins Laboratories were surprised that phonemes, intuitively present in perception, did not reveal themselves in spectrographic displays of the physical speech signal. Intuition should guide investigation, but should not evaluate theories. However, if a more intuitive model, such as the distributed memory model, could account for all the same data as a strict exemplar model, it may be favored. In any case, direct comparison of models with respect to data should form our primary criterion for evaluation of theories.

- *Episodic Boundaries* A less philosophical problem for exemplar models regards the ambiguous nature of linguistic events. If the model assumes that episodes mediate perception, it must state what constitutes an episode (see Hintzman, 1986). In a tightly controlled and constrained laboratory situation, "episodes" conform naturally to experimental trials. Since our theories are intended to generalize outside of the laboratory, however, some principled theoretical definition is needed. Language is a hierarchical system in which words are fairly low-level entities. Memory for specific sentences, conversations, or entire novels could be more dominant episodes than single instances of individual words. How would a model such as MINERVA 2 or a distributed memory model accommodate the widely varying temporal attributes of episodes? Tulving and Schacter (1990) suggest that priming is a commonplace, everyday occurrence; perhaps a principled approach to defining episodes would allow predictions regarding priming effects in natural settings, such as the experience of deja vu, or habitual behaviors that are pervasive.

- *Episodic Cognition?* What is the relation of episodic traces to rules and creativity? If words, sentences, and perhaps conversations are encoded and maintained as episodes in memory, do linguistic rules emerge via abstraction, as McClelland and Rumelhart (1985; 1986) suggest, or are rules a separate aspect of the system? Chomsky (1959) argued powerfully and convincingly for the creativity of language and suggested that language is not mentally represented by a mere list of sentences. Can a set of episodes of grammatical (and agrammatical) sentences support the grammatical evaluation of novel sentences in the absence of rules? I do not know the answer to this. If episodic representations are considered an integral aspect of linguistic function, however, serious exploration of this question is necessary. By the PRS hypothesis (Schacter, 1990), episodes are pre-semantic representations that only process stimulus form. Perhaps this strong segregation of perceptual episodes and linguistic function is necessary. Once again, only explicit comparison of appropriate theories will provide true insight into this problem.

- *Perception and Production* The possibility of multiple episodes mediating perception raises another important question: What is the relation of speech perception to production? Most contemporary models of speech production are fully abstractionist models, such as Dell's (1986) network, which posits multiple levels of nodes corresponding to words, phonemes, etc. If one espouses an episodic model of speech perception, however, it may be parsimonious to assume a complementary episodic model of speech production. The advantages of such a model are, to my knowledge, unexplored. A possible by-product of an episodic model of production could be an account of a speaker's ability to imitate voices (Hollien et al., 1982). Further exploration of the possibilities of an episodic productive lexicon could constitute an exciting area for future research and theory.

Directions for Future Research

The results of the present investigation, taken together with related findings, suggests a non-analytic view of the mental lexicon. By the traditional analytic view, perceptual idiosyncracies across words are normalized, allowing matching procedures to identify individual tokens as types. By the non-analytic view, every word is recognized against a background of similar words, all represented via countless, perceptually detailed episodes. The visual or auditory signal should not be considered as a transparent medium of linguistic content; the vehicle itself is an integral aspect of later representation. This episodic approach to lexical representation and process foreshadows exciting prospects for continuing research. Several important lines of future investigation are suggested below; countless variations on the themes are certainly possible:

- First and foremost, the effects described in the present research should be replicated and extended. The present investigation reports novel findings regarding memory for voice, perceptual scaling in repetition effects, and dissociations between memory measures with respect to the importance of similarity relations among stimuli. All of these findings should be tested in other domains, such as visual word recognition, explicit voice learning, and others.

- Following Schacter and Church (in press), the role of hemispheric organization in mediating voice memory should be investigated. Also, as suggested above, the role of masking stimuli in episodic representations should be assessed. If stimulus masks become encoded as stimulus attributes, numerous common assumptions about masking procedures may require new consideration.

- The episodic view of lexical representation carries several implications for theories of speaker normalization and voice recognition. Any of the theories described above, whether strict exemplar models, hybrid models, or record-based models, should predict that voice familiarity will increase perceptual fluency, even for novel spoken stimuli. For example, providing listeners with extensive speech samples from a set of speakers should increase the speed or accuracy of pseudoword recognition. Moreover, the inherent and contextual variability of the training materials should modulate the benefits of voice training during generalization (see Logan, Lively, and Pisoni, 1991).

- To account for effects of perceptual details, such as voice, in repetition effects, episodic models are required to assume rich encoding of contextual information. The interaction of environmental context shifts and perceptual details in repetition paradigms has received little attention (Jacoby & Witherspoon, 1982). Empirical evaluation of this common assumption is obviously important.

- As reviewed above, the recent literature on speaker variability effects on perception, attention, and memory are ambiguous. Further research is necessary to determine the true cause of performance deficits in multiple-speaker experiments. Separately assessing the contributions of speaker normalization and mere distraction may indicate whether the concept of normalization needs to be invoked for data such as Nusbaum and Morin's (1992) results.

The topics discussed in this section constitute only a partial listing of important issues and directions for future research. As we become more familiar with the assumptions and behaviors of episodic models of speech perception and word recognition, countless other issues will no doubt arise and demand equal time and attention.

Conclusion

In a recent scathing article, Watkins (1990) argued that "memory theorizing is going nowhere." He argued that pages and pages of memory research and theory fill journals more quickly than anyone could possibly keep up with, but the growth of understanding is minimal. According to Watkins, the problem with memory research is the tenet of "mediationism" in all theories. Mediationism is the belief that memory behavior is contingent upon memory traces. He argues that the concept is empirically unassailable, logically unnecessary, and an obstruction to new ways of thinking about long-standing problems.

The present data suggest a mental lexicon completely characterized by mediationism; in contrast to traditional notions of abstract lexical representations specialized matching processes, the present results suggest that detailed, episodic traces of spoken words subserve later word recognition. In a review of theories of speech perception, Klatt (1989) notes that an optimal perceptual system maintains maximal information in its representations. Following Klatt (1989) and Jacoby and Brooks (1984) I suggest that spoken word recognition occurs against a background of detailed perceptual episodes. Watkins intended to erase (if he will pardon the metaphor) mediationist thinking from the minds of contemporary memory theorists. I would argue for mediationism at a fever pitch not found in traditional, abstractionist models of mind. The most natural account of the present results, and the most flexible assumption of mental representation, is to posit a mental system that maintains maximal information about linguistic episodes. With respect to spoken word recognition, the proper name for such a theory would be "Richard Semon's theory of the lexicon."

References

Abercrombie, D. (1967). *Elements of General Phonetics.* University of Chicago Press.

Alba, J.W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin, 93*, 203-231.

Allard, F. & Henderson, L. (1976). Physical and name codes in auditory memory: The pursuit of an analogy. *Quarterly Journal of Experimental Psychology* , *28* , 475-482.

Andersen, S.M., & Cole, S.W. (1990). "Do I know you?" The role of significant others in general social perception. *Journal of Personality and Social Psychology, 59* , 384-399.

Anderson, J.R. (1983) *The architecture of cognition* . Cambridge, MA: Harvard University Press.

Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 802-814.

Atkinson, R.C., & Shiffrin, R.M. (1968). Human memory: A proposed system and its control processes. In K.W. Spence and J.T. Spence (Eds.), *The Psychology of Learning and Motivation: Advances in Research and Theory* , vol. 2. New York: Academic Press. Pp. 89-195.

Bahrick, H.P., Bahrick, P.O., & Wittlinger, R. (1975). Fifty years of memory for names and faces: A cross-sectional approach. *Journal of Experimental Psychology: General, 104* , 54-75.

Bamber, D. (1969). Reaction times and error rates for "same"-"different" judgments of multidimensional stimuli. *Perception & Psychophysics, 6* , 169-174.

Banks, W.P. (1970). Signal detection theory and human memory. *Psychological Bulletin, 74*, 81-99.

Barsalou, L.W. (1990). On the indistinguishability of exemplar memory and abstraction in category representation. In T.K. Srull & R.S. Wyer Jr. (Eds.), *Advances in Social Cognition, Vol. III* . Hillsdale, NJ: Erlbaum. Pp. 61-88.

Becker, C.A. (1976). Allocation of attention during visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 2* , 556-566.

Becker, C.A. (1980). Semantic context effects in visual word recognition: An analysis of semantic strategies. *Memory & Cognition, 8*, 493-512.

Becker, C.A., & Killion, T.H. (1977). Interaction of visual and cognitive effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 389-401.

Begg, I. (1971). Recognition memory for sentence meaning and wording. *Journal of Verbal Learning and Verbal Behavior, 10* , 176-181.

Begg, I., Maxwell, D., Mitterer, J.O., & Harris, G. (1986). Estimates of frequency: Attribute or attribution? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12* , 496-508.

Begg, I., & Wickelgren, W.A. (1974). Retention functions for syntactic and lexical vs. semantic information in sentence recognition memory. *Memory & Cognition, 2*, 353-359.

Besner, D., & Coltheart, M. (1975). Same-different judgments with words and nonwords: Differential effects of relative size. *Memory & Cognition, 3*, 673-677.

Biederman, I., & Cooper, E.E. (1991). Priming contour-deleted images: Evidence for intermediate representations in visual object recognition. *Cognitive Psychology, 23*, 393-419.

Blosser, A.B. (1965). A performance-oriented approach to detection: Tables for detection, discrimination, and decision theory. *TRACOR Document 65-267-U* . Austin, TX: TRACOR Inc.

Bray, N.W., & Batchelder, W.H. (1972). Effects of instructions and retention interval on memory of presentation mode. *Journal of Verbal Learning and Verbal Behavior, 11*, 367-374.

Bricker, P.D., & Pruzansky, S. (1966). Effects of stimulus context and duration on talker identification. *Journal of the Acoustical Society of America, 40* , 1441-1449.

Brooks, L.R. (1987). Decentralized control of categorization: The role of prior processing episodes. In U. Neisser (Ed.), *Concepts and Conceptual Development..* New York: Cambridge University Press. Pp. 141-174.

Brown, H.L., Sharma, N.K., & Kirsner, K. (1984). The role of script and phonology in lexical processing. *Quarterly Journal of Experimental Psychology, 36A*, 491-505.

Bruce, V. (1988). *Recognising faces* . Hillsdale, NJ: Erlbaum.

Bruck, M., Cavanagh, P., & Ceci, S.J. (1991). Fortysomething: Recognizing faces at one's 25th reunion. *Memory & Cognition, 19* , 221-228.

Carr, P.B., & Trill, D. (1964). Long-term larynx-excitation spectra. *Journal of the Acoustical Society of America, 36* , 2033-2040.

Carr, T.H., Brown, J.S., & Charalambous, A. (1989). Repetition and reading: Perceptual encoding mechanisms are very abstract but not very interactive. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15* , 763-778.

Carrell, T.D. (1984). Contributions of fundamental frequency, formant spacing, and glottal waveform to talker identification. Unpublished doctoral dissertation. Indiana University.

Carrell, T.D., Smith, L.B., & Pisoni, D.B. (1981). Some perceptual dependencies in speeded classification of vowel color and pitch. *Perception & Psychophysics, 29* , 1-10.

Carroll, J.D., & Arabie, P. (1980). Multidimensional scaling. *Annual Review of Psychology, 31* , 607-649.

Carroll, M., Byrne, B., & Kirsner, K. (1985). Autobiographical memory and perceptual learning: A developmental study using picture recognition, naming latency, and perceptual recognition. *Memory & Cognition, 13* , 273-279.

Carterette, E.C., & Barnebey, A. (1975). Recognition memory for voices. In A. Cohen & S.G. Nooteboom (Eds.), *Structure and Process in Speech Perception* . New York: Springer-Verlag. Pp. 246-265.

Chananie, J., & Tikofsky, R. (1969). Choice response time and distinctive features in speech discrimination. *Journal of Experimental Psychology, 81* , 161-163.

Chance, J., Goldstein, A.G., & McBride, L. (1975). Differential experience and recognition memory for faces. *Journal of Social Psychology, 102* , 243-253.

Chomsky, N. (1959). Review of *Verbal Behavior* . *Language, 35* , 26-58.

Clarke, R., & Morton, J. (1983). Cross-modality facilitation in tachistoscopic word recognition. *Quarterly Journal of Experimental Psychology, 35A* , 79-96.

Clifford, B.R. (1983). Memory for voices: The feasibility and quality of earwitness evidence. In S.M.A. Lloyd-Bostock & B.R. Clifford (Eds.), *Evaluating witness evidence* . John Wiley & Sons.

Cofer, C.C. (1967). Conditions for the use of verbal associations. *Psychological Bulletin, 68* , 1-12.

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.

Cole, R.A., & Jakimik, J. (1980). A model of speech perception. In Cole, R.A. (Ed.), *Perception and production of fluent speech* . Hillsdale, NJ: Erlbaum. pp. 133-163.

Coltheart, M., Davelaar, E., Jonasson, J.T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* , (pp. 535-555). Hillsdale, NJ: Erlbaum.

Cooper, L.A., Schacter, D.L., Ballesteros, S., & Moore, C. (1992). Priming and recognition of transformed three-dimensional objects: Effects of size and reflection. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18* , 43-57.

Craik, F.I.M. (1981). Encoding and retrieval effects in human memory: A partial review. In J. Long & A.D. Baddeley (Eds.), *Attention and Performance IX* . Hillsdale, NJ: Erlbaum.

Craik, F.I.M., & Gemar, M.C. (1990). Evidence for long-lasting perceptual information in the transformed text paradigm. Paper presented at the 31st annual meeting of the Psychonomic Society, New Orleans.

Craik, F.I.M., & Kirsner, K. (1974). The effect of speaker's voice on word recognition. *Quarterly Journal of Experimental Psychology, 26* , 274-284.

Craik, F.I.M., & Lockhart, R.S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11* , 671-684.

Craik, F.I.M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*, 268-294.

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

Crowder, R.G., Serafine, M.L., & Repp, B. (1990). Physical interaction and association by contiguity in memory for the words and melodies of songs. *Memory & Cognition, 18*, 469-476.

Cutting, J.E., & Kozlowski, L.T. (1977). Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society, 9*, 353-356.

Cutting, J.E., Proffitt, D.R., & Kozlowski, L.T. (1978). A biomechanical invariant for gait perception. *Journal of Experimental Psychology: Human Perception and Performance, 4*, 357-372.

Cutting, J.E., & Pisoni, D.B. (1978). An information-processing approach to speech perception. In Kavanagh, J.F., & Strange, W. (Eds.), *Speech and language in the laboratory, school, and clinic* . Cambridge, MA: MIT Press.

Dell, G. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review, 93* , 283-321.

Derr, M.A., & Massaro, D.W. (1980). The contribution of vowel duration, F0 contour, and frication duration as cues to the /juz/-/jus/ distinction. *Perception & Psychophysics, 27* , 51-59.

Dunn, J.C., & Kirsner, K. (1989). Implicit memory: Task or process? In S. Lewandowski, J.C. Dunn, & K. Kirsner (Eds.), *Implicit memory: Theoretical issues* . Hillsdale, NJ: Erlbaum. Pp. 17-31.

Egan, D., Pittner, M., & Goldstein, A.G. (1977). Eyewitness identification: Photographs versus live models. *Law and Human Behavior, 1* , 199-206.

Eich, J.M. (1982). A composite holographic associative recall model. *Psychological Review, 89* , 627-661.

Eimas, P.D., Siqueland, E.R., Jusczyk, P.W., & Vigorito, J. (1971). Speech perception in infants. *Science, 171* , 303-306.

Elman, J.L. (1989). Connectionist approaches to acoustic/phonetic processing. In Marslen-Wilson, W.D. (Ed.), *Lexical representation and process* . Cambridge, MA: MIT Press. pp. 227-260.

Elman, J.L. (1990). Representation and structure in connectionist models. In Altmann, G.T.M. (Ed.), *Cognitive models of speech processing* . Cambridge, MA: MIT Press. pp. 345-382.

Elman, J.L., & McClelland, J.L. (1984). An interactive activation model of speech perception. In Lass, N.J. (Ed.), *Language and Speech* . New York: Academic Press.

Elman, J.L., & McClelland, J.L. (1986). Exploiting lawful variability in the speech waveform. In Perkell, J.S., & Klatt, D.H. (Eds.), *Invariance and variability in speech processing*. Hillsdale, NJ: Erlbaum. pp. 360-385.

Eysenck, M.W., & Eysenck, M.C. (1980). Effects of processing depth, distinctiveness, and word frequency on retention. *British Journal of Psychology, 71* , 263-274.

Fant, G. (1973). *Speech sounds and features* . Cambridge, MA: MIT Press.

Felfoldy, G.L. (1974). Repetition effects in choice reaction time to multidimensional stimuli. *Perception & Psychophysics, 15* , 453-459.

Feustel, T.C., Shiffrin, R.M., & Salasoo, A. (1983). Episodic and lexical contributions to the repetition effect in word recognition. *Journal of Experimental Psychology: General, 112* , 309-346.

Fodor, J.A. (1983). *The modularity of mind* . Cambridge, MA: MIT Press.

Fodor, J.A. (1985). Precis of *The modularity of mind. The Behavioral and Brain Sciences, 8* , 1-42.

Forster, K.I. (1976). Accessing the mental lexicon. In R.J. Wales, & E.C.T. Walker (Eds.), *New approaches to language mechanisms* . Amsterdam: North-Holland.

Forster, K.I. (1979). Levels of processing and the structure of the language processor. In Cooper, W.E., & Walker, E.C.T. (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* . Hillsdale, NJ: Erlbaum. pp. 27-86.

Forster, K.I. (1989). Basic issues in lexical processing. In W.D. Marslen-Wilson, (Ed.), *Lexical representation and process* . Cambridge, MA: MIT Press. pp. 75-107.

Forster, K.I. (1990). Lexical processing. In Osherson, D.N., & Lasnik, H. (Eds.) *An invitation to cognitive science* (Volume 1). Cambridge: MIT Press. pp. 95-131.

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

Fowler, C.A., & Rosenblum, L.D. (1990). Duplex perception: A comparison of monosyllables and slamming of doors. *Journal of Experimental Psychology: Human Perception and Performance, 16* , 742-754.

Fowler, C.A., & Rosenblum, L.D. (1991). The perception of phonetic gestures. In Mattingly, I.G., & Studdert-Kennedy, M. (Eds.), *Modularity and the motor theory of speech perception..* Hillsdale, NJ: Erlbaum. Pp. 33-59.

Galton, F. (1883). *Inquiries into human faculty and its development* . London: Macmillan. (cited in Hintzman, 1986).

Ganong, W.F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance, 6* , 110-125.

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

Geiselman, R.E. (1979). Inhibition of the automatic storage of speaker's voice. *Memory & Cognition , 7,* 201-204.

Geiselman, R.E., & Bellezza, F.S. (1976). Long-term memory for speaker's voice and source location. *Memory & Cognition , 4 ,* 483-489.

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

Geiselman, R.E., & Crawley, J.M. (1983). Incidental processing of speaker characteristics: Voice as connotative information. *Journal of Verbal Learning and Verbal Behavior, 22,* 15-23.

Gerstman, L.H. (1968). Classification of self-normalized vowels. *IEEE Transactions on Audio and Electroacoustics, au-16* , 78-80.

Gibbs, R.W. (1981). Memory for requests in conversation. *Journal of Verbal Learning and Verbal Behavior, 20 ,* 630-640.

Gibson, J.J. (1966). *The senses considered as perceptual systems.* Boston, MA: Houghton-Mifflin.

Gillund, G., & Shiffrin, R.M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91* , 1-67.

Glanzer, M., & Cunitz, A.R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior, 5 ,* 351-360.

Godden, D.R., & Baddeley, A.D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology, 66,* 325-332.

Goggin, J.P., Thompson, C.P., Strube, G., & Simental, L.R. (1991). The role of language familiarity in voice identification. *Memory & Cognition, 19* , 448-458.

Goldinger, S.D., Luce, P.A., & Pisoni, D.B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language, 28,* 501-518.

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

Goldinger, S.D., Pisoni, D.B., & Luce, P.A. (in press). Speech perception and spoken word recognition: Research and theory. To appear in N.J. Lass (Ed.), *Principles of Experimental Phonetics* . Toronto: B.C. Decker.

Goldstein, A.G., Knight, P., Bailis, K., & Conover, J. (1981). Recognition memory for accented and unaccented voices. *Bulletin of the Psychonomic Society, 17* , 217-220.

Goodman, J.C., & Huttenlocher, J. (1988). Do we know how people identify spoken words? *Journal of Memory and Language, 27* , 684-698.

Graf, P. (1982). The memorial consequences of generation and transformation. *Journal of Verbal Learning and Verbal Behavior, 21* , 539-548.

Graf, P., & Levy, B.A. (1984). Reading and remembering: Conceptual and perceptual processing involved in reading rotated passages. *Journal of Verbal Learning and Verbal Behavior, 23* , 405-424.

Graf, P., & Mandler, G. (1984). Activation makes words more accessible but not necessarily more retrievable. *Journal of Verbal Learning and Verbal Behavior, 23*, 553-568.

Graf, P., Mandler, G., & Haden, P.E. (1982). Simulating amnesic symptoms in normals. *Science, 218* , 1243-1244.

Graf, P., & Ryan, L. (1990). Transfer-appropriate processing for implicit and explicit memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16* , 978-992.

Green, K.P., Kuhl, P.K., Meltzoff, A.N., & Stevens, E.B. (1991). Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the McGurk effect. *Perception & Psychophysics, 50* , 524-536.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics, 28* , 267-283.

Halpern, A.R. (1989). Memory for the absolute pitch of familiar songs. *Memory & Cognition, 17* , 572-581.

Hasher, L., & Zacks, R.T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108* , 356-388.

Hashtroudi, S., Parker, E.S., DeLisi, L.E., Wyatt, R.J., & Mutter, S.A. (1984). Intact retention in acute alcohol amnesia. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10* , 156-163.

Hinton, G.E., & Anderson, J.A. (1981). *Parallel models of associative memory* . Hillsdale, NJ: Erlbaum.

Hintzman, D.L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93* , 411-428.

Hintzman, D.L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95* , 528-551.

Hintzman, D.L. (1990). Human learning and memory: Connections and dissociations. *Annual Review of Psychology, 41* , 109-39.

Hintzman, D.L., Block, R.A., & Inskeep, N.R. (1972). Memory for mode of input. *Journal of Verbal Learning and Verbal Behavior, 11* , 741-749.

Hintzman, D.L., Block, R.A., & Summers, J.J. (1973). Modality tags and memory for repetitions: Locus of the spacing effect. *Journal of Verbal Learning and Verbal Behavior, 12,* 229-238.

Hintzman, D.L., Grandy, C.A., & Gold, E. (1981). Memory for frequency: A comparison of two multiple-trace theories. *Journal of Experimental Psychology: Human Learning and Memory, 7* , 231-240.

Hintzman, D.L., & Ludlam, G. (1980). Differential forgetting of prototypes and old instances: Simulation by an exemplar-based classification model. *Memory & Cognition, 8* , 378-382.

Hintzman, D.L., Nozawa, G., & Irmscher, M. (1982). Frequency as a nonpropositional attribute of memory. *Journal of Verbal Learning and Verbal Behavior, 21,* 127-141.

Hintzman, D.L., & Summers, J.J. (1973). Long-term visual traces of visually presented words. *Bulletin of the Psychonomic Society, 1* , 325-327.

Hockley. W.E. (1982). Retrieval processes in continuous recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8* , 497-512.

Hollien, H., Majewski, W., & Doherty, E. (1982). Perceptual identification of voices under normal, stress, and disguise speaking conditions. *Journal of Phonetics, 10,* 139-148.

Holmberg, T.L., Morgan, K.A., & Kuhl, P.K. (1977). Speech perception in early infancy: Discrimination of fricative consonants. *Journal of the Acoustical Society of America, 62*, S76.

Homa, D., Sterling, S., & Trepel, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and Memory, 7* , 418-439.

Horton, K.D. (1985). The role of semantic information in reading spatially transformed text. *Cognitive Psychology, 17* , 66-88.

Horton, K.D. (1989). The processing of spatially transformed text. *Memory & Cognition, 17* , 283-291.

House, A.S., Williams, C.E., Hecker, M.H.L., & Kryter, K.D. (1965). Articulation-testing methods: Consonantal differentiation with a closed-response set. *Journal of the Acoustical Society of America, 37* , 158-166.

Howell, W.C. (1973). Representation of frequency in memory. *Psychological Bulletin, 80*, 44-53.

Humphreys, M.S., Bain, J.D., & Pike, R. (1989). Different ways to cue a coherent memory system: A theory for episodic, semantic, and procedural tasks. *Psychological Review, 96*, 208-233.

Hyde, S.R. (1972). Automatic speech recognition: A critical survey of the literature. In E. David & P. Denes (Eds.), *Human communication: A unified view* . New York: McGraw-Hill.

Hyde, T.S., & Jenkins, J.J. (1969). Differential effects of incidental tasks on the organization of recall of a list of highly associated words. *Journal of Experimental Psychology, 82*, 472-481.

Jackson, A., & Morton, J. (1984). Facilitation of auditory word recognition. *Memory & Cognition, 12* , 568-574.

Jacoby, L.L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior, 17* , 649-667.

Jacoby, L.L. (1983a). Perceptual enhancement: Persistent effects of an experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 21-38.

Jacoby, L.L. (1983b). Remembering the data: Analyzing interactive processes in reading. *Journal of Verbal Learning and Verbal Behavior, 22* , 485-508.

Jacoby, L.L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30* , 513-541.

Jacoby, L.L., Baker, J.G., & Brooks, L.R. (1989). Episodic effects on picture identification: Implications for theories of concept learning and theories of memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15* , 275-281.

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*. Academic Press. Pp. 1-47.

Jacoby, L.L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110* , 306-340.

Jacoby, L.L., & Hayman, C.A.G. (1987). Specific visual transfer in word identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13* , 456-463.

Jacoby, L.L., Levy, B.A., & Stelmach, K. (1992). Episodic transfer and automaticity: Integration of data-driven and conceptually-driven processing in rereading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18* , 15-24.

Jacoby, L.L., Marriott, M.J., & Collins, J.G. (1990). The specifics of memory and cognition. In T.K. Srull & R.S. Wyer Jr. (Eds.), *Advances in Social Cognition, Vol. III*, Hillsdale, NJ: Erlbaum. Pp. 111-121.

Jacoby, L.L., & Witherspoon, D. (1982). Remembering without awareness. *Canadian Journal of Psychology, 36* , 300-324.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics, 14* , 201-211.

Johansson, G. (1975). Visual motion perception. *Scientific American, 232* , 76-89.

Johnson, K. (1990). The role of perceived speaker identity in F0 normalization of vowels. *Journal of the Acoustical Society of America, 88* , 642-654.

Johnson, M.K., & Hasher, L. (1987). Human learning and memory. *Annual Review of Psychology, 38* , 631-668.

Johnson, M.K., Peterson, M.A., Chua Yap, E., & Rose, P.M. (1989). Frequency judgments: The problem of defining a perceptual event. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15* , 126-136.

Johnston, J.C., van Santen, J.P.H., & Hale, B.L. (1985). Repetition effects in word and pseudoword identification: Comment on Salasoo, Shiffrin, and Feustel. *Journal of Experimental Psychology: General, 114* , 498-508.

Johnston, W.A., Dark, V.J., & Jacoby, L.L. (1985). Perceptual fluency and recognition judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 3-11.

Joos, M.A. (1948). Acoustic phonetics. *Language, 24* , Supplement 2, 1-136.

Jusczyk, P.W., & Cohen, A. (1985). What constitutes a module? *The Behavioral and Brain Sciences, 8* , 20-21.

Jusczyk, P.W., Pisoni, D.B., & Mullennix, J.W. (in press). Effects of talker variability on speech perception by 2-month old infants. *Cognition* .

Keenan, J.M., MacWhinney, B., & Mayhew, D. (1977). Pragmatics in memory: A study in natural conversation. *Journal of Verbal Learning and Verbal Behavior, 16* , 549-560.

Kelley, C.M., Jacoby, L.L., & Hollingshead, A. (1989). Direct versus indirect tests of memory for source: Judgments of modality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15* , 1101-1108.

Kemler, D.G. (1983). Exploring and reexploring issues of integrality, perceptual sensitivity, and dimensional salience. *Journal of Experimental Child Psychology, 36* , 365-379.

Kersteen-Tucker, Z. (1991). Long-term repetition priming with symmetrical polygons and words. *Memory & Cognition, 19* , 37-43.

Kirsner, K. (1973). An analysis of the visual component in recognition memory for verbal stimuli. *Memory & Cognition, 1* , 449-453.

Kirsner, K. (1974). Modality differences in recognition memory for words and their attributes. *Journal of Experimental Psychology, 102* , 579-584.

Kirsner, K., & Dunn, J.C. (1985). The perceptual record: A common factor in repetition priming and attribute retention? In M.I. Posner & O.M. Marin (Eds.), *Attention and Performance XI* . Hillsdale, NJ: Erlbaum. Pp. 547-565.

Kirsner, K., Dunn, J.C., & Standen, P. (1987). Record-based word recognition. In M. Coltheart (Ed.), *Attention and Performance XII, The Psychology of Reading* . Erlbaum. Pp. 147-167.

Kirsner, K., Dunn, J.C., & Standen, P. (1989). Domain-specific resources in word recognition. In S. Lewandowsky, J.C. Dunn, & Kirsner, K. (Eds.), *Implicit memory: Theoretical issues* . Hillsdale, NJ: Erlbaum. Pp. 99-122.

Kirsner, K., Milech, D., & Standen, P. (1983). Common and modality-specific processes in the mental lexicon. *Memory & Cognition, 11* , 621-630.

Kirsner, K., & Smith, M.C. (1974). Modality effects in word identification. *Memory & Cognition, 2* , 637-640.

Klatt, D.H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics, 7* , 279-312.

Klatt, D.H. (1981). Lexical representations for speech production and perception. In T. Myers, J. Laver, & J. Anderson (Eds.), *The Cognitive Representation of Speech.* North-Holland. Pp. 11-37.

Klatt, D.H. (1989). Review of selected models of speech perception. In W.D. Marslen-Wilson, (Ed.), *Lexical representation and process* . Cambridge, MA: MIT Press. pp. 169-226.

Knapp, A.G., & Anderson, J.A. (1984). Theory of categorization based on distributed memory storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 616-637.

Kohonan, T., Oja, E., & Lehtio, P. (1981). Storage and processing of information in distributed associative memory systems. In Hinton, G.E. & J.A. Anderson (Eds.), *Parallel Models of Associative Memory* . Hillsdale, NJ: Erlbaum.

Kolers, P.A. (1973). Remembering operations. *Memory & Cognition, 1* , 347-355.

Kolers, P.A. (1974a). Remembering trivia. *Language and Speech, 17* , 324-336.

Kolers, P.A. (1974b). Two kinds of recognition. *Canadian Journal of Psychology, 28*, 51-61.

Kolers, P.A. (1975a). Memorial consequences of automatized encoding. *Journal of Experimental Psychology: Human Learning and Memory, 1* , 689-701.

Kolers, P.A. (1975b). Specificity of operations in sentence recognition. *Cognitive Psychology, 7* , 289-306.

Kolers, P.A. (1976a). Reading a year later. *Journal of Experimental Psychology: Human Learning and Memory, 2* , 554-565.

Kolers, P.A. (1976b). Pattern analyzing memory. *Science, 191* , 1280-1281.

Kolers, P.A. (1978). On the representations of experience. In D. Gerver & H.W. Sinaiko (Eds.), *Language Interpretation and Communication* . New York: Plenum Press. Pp. 245-258.

Kolers, P.A., & Magee, L.E. (1978). Specificity of pattern-analyzing skills in reading. *Canadian Journal of Psychology, 32* , 43-51.

Kolers, P.A., & Ostry, D.J. (1974). Time course of loss of information regarding pattern analyzing operations. *Journal of Verbal Learning and Verbal Behavior, 13*, 599-612.

Kolers, P.A., Palef, S.R., & Stelmach, L.B. (1980). Graphemic analysis underlying literacy. *Memory & Cognition, 8* , 322-328.

Kolers, P.A., & Roediger, H.L., III. (1984). Procedures of mind. *Journal of Verbal Learning and Verbal Behavior, 23* , 425-449.

Kosslyn, S.M., & Matt, A.M.C. (1977). If you speak slowly, do people read your prose slowly? Person-particular speech recoding during reading. *Bulletin of the Psychonomic Society, 9,* 250-252.

Kroll, J.F., & Potter, M.C. (1984). Recognizing words, pictures, and concepts: A comparison of lexical, object, and reality decisions. *Journal of Verbal Learning and Verbal Behavior, 23* , 39-66.

Krulee, G.K., Tondo, D.K., & Wightman, F.L. (1983). Speech perception as a multilevel processing system. *Journal of Psycholinguistic Research, 12*, 531-554.

Kruskal, J.B., & Wish, M. (1978). *Multidimensional scaling*. London: Sage University Press.

Kruskal, J.B., Young, F.W., & Sherry, J.B. (1973). How to use KYST, a very flexible program to do multidimensional scaling and unfolding. Bell Laboratories, unpublished paper.

Kuhl, P.K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *Journal of the Acoustical Society of America, 66* , 1668-1679.

Kuhl, P.K., & Miller, J.D. (1982). Discrimination of auditory target dimensions in the presence or absence of variation in a second dimension by infants. *Perception & Psychophysics, 31*, 279-292.

Ladefoged, P. (1980). What are linguistic sounds made of? *Language, 56* , 485-502.

Ladefoged, P., & Ladefoged, J. (1980). The ability of listeners to identify voices. *UCLA Working Papers in Phonetics, 49* , 43-51.

Ladefoged, P., & Broadbent, D.E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America, 29* , 98-104.

Landauer, T.K., & Streeter, L.A. (1973). Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning and Verbal Behavior, 12* , 119-131.

Laughery, K.R., Alexander, J.F., & Lane, A.B. (1971). Recognition of human faces: Effects of target exposure time, target position, pose position, and type of photograph. *Journal of Applied Psychology, 55* , 477-483.

Legge, G.E., Grosmann, C., & Pieper, C.M. (1984). Learning unfamiliar voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 298-303.

Lehman, E.B. (1982). Memory for modality: Evidence for an automatic process. *Memory & Cognition, 10* , 554-564.

Levy, B.A. (1983). Proofreading familiar text: Constraints on visual processing. *Memory & Cognition, 11* , 1-12.

Levy, B.A., & Kirsner, K. (1989). Reprocessing text: Indirect measures of word and message level processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15* , 407-417.

Lewandowski, S., Dunn, J.C., & Kirsner, K. (Eds.), *Implicit memory: Theoretical issues*. Hillsdale, NJ: Erlbaum.

Lewicki, P. (1986). *Nonconscious social information processing*. San Diego, CA: Academic Press.

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

Liberman, A.M. (1982). On finding that speech is special. *American Psychologist, 37*, 148-167.

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

Liberman, A.M., & Mattingly, I.G. (1985). The motor theory of speech perception revised. *Cognition, 21* , 1-36.

Liberman, A.M., & Mattingly, I.G. (1989). A specialization for speech perception. *Science, 243* , 489-494.

Light, L.L., Stansbury, C., Rubin, C., & Linde, S. (1973). Memory for modality of presentation: Within-modality discrimination. *Memory & Cognition, 1*, 395-400.

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

Logan, G.D. (1988a). Automaticity, resources, and memory: Theoretical controversies and practical implications. *Human Factors, 30* , 583-598.

Logan, G.D. (1988b). Toward an instance theory of automatization. *Psychological Review, 95*, 492-527.

Logan, G.D. (1990). Repetition priming and automaticity: Common underlying mechanisms? *Cognitive Psychology, 22* , 1-35.

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

Lovelace, E.A., & Southall, S.D. (1983). Memory for words in prose and their locations on the page. *Memory & Cognition, 11* , 429-434.

Luce, P.A. (1986). Neighborhoods of words in the mental lexicon. Unpublished doctoral dissertation. Indiana University.

Luce, P.A., & Carrell, T.D. (1981). Creating and editing waveforms using WAVES. *Research on Speech Perception, Progress Report No. 7* . Bloomington, IN: Speech Research Laboratory, Department of Psychology, Indiana University.

Luce, P.A., Pisoni, D.B., & Goldinger, S.D. (1990). Similarity neighborhoods of spoken words. In Altmann, G. (Ed.), *Cognitive representation of speech* . Cambridge, MA: MIT Press. 122-147.

MacWhinney, B., Keenan, J.M., & Reinke, P. (1982). The role of arousal in memory for conversation. *Memory & Cognition, 10* , 308-317.

Mann, V.A., & Liberman, A.M. (1983). Some differences between phonetic and auditory modes of perception. *Cognition, 14* , 211-235.

Manso De Zuniga, C., Humphreys, G.W., & Evett, L.J. (1991). Additive and interactive effects of repetition, degradation, and word frequency in the reading of handwriting. In D. Besner & G.W. Humphreys (Eds.), *Basic processes in reading*. Hillsdale, NJ: Erlbaum.

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

Marslen-Wilson, W.D. (1980a). Optimal efficiency in human speech processing. Unpublished manuscript.

Marslen-Wilson, W.D. (1980b). Speech understanding as a psychological process. In Simon, J.C. (Ed.), *Spoken language generation and understanding* . Dordrecht, Holland: Reidel.

Marslen-Wilson, W.D. (1987). Functional parallelism in spoken word recognition. *Cognition, 25* , 71-102.

Marslen-Wilson, W.D. (1989). Access and integration: Projecting sound onto meaning. In Marslen-Wilson, W.D. (Ed.), *Lexical representation and process*. Cambridge, MA: MIT Press. pp. 3-24.

Marslen-Wilson, W.D. (1990). Activation, competition, and frequency in lexical access. In Altmann, G.T.M. (Ed.), *Cognitive models of speech processing* . Cambridge, MA: MIT Press. pp. 148-172.

Marslen-Wilson, W.D., & Tyler, L.K. (1975). Processing structure of sentence perception. *Nature, 257*, 784-785.

Marslen-Wilson, W.D., & Tyler, L.K. (1980). The temporal structure of spoken language understanding. *Cognition, 8* , 1-71.

Marslen-Wilson, W.D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10*, 29-63.

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.

Massaro, D.W. (1972). Preperceptual images, processing time, and perceptual units in auditory perception. *Psychological Review, 79* , 124-145.

Massaro, D.W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry* . Hillsdale, NJ: Erlbaum.

Massaro, D.W. (1989). Multiple book review of *Speech perception by ear and eye: A paradigm for psychological inquiry* . *The Behavioral and Brain Sciences, 12* , 741-794.

Massaro, D.W., & Cohen, M.M. (1976). The contribution of fundamental frequency and voice onset time to the /zi/ - /si/ distinction. *Journal of the Acoustical Society of America, 60*, 704-717.

Massaro, D.W., & Cohen, M.M. (1977). The contribution of voice-onset time and fundamental frequency as cues to the /zi/ - /si/ distinction. *Perception & Psychophysics, 22* , 373-382.

Massaro, D.W., & Oden, G.C. (1980). Speech perception: A framework for research and theory. In Lass, N.J. (Ed.), *Speech and language: Advances in basic research and practice* (Volume 3). New York: Academic Press. pp. 129-165.

Masson, M.E.J. (1984). Memory for the surface structure of sentences: Remembering with and without awareness. *Journal of Verbal Learning and Verbal Behavior, 23*, 579-592.

Masson, M.E.J. (1986). Identification of typographically transformed words: Instance-based skill acquisition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12* , 479-488.

Masson, M.E.J. (1989). Fluent reprocessing as an implicit expression of memory for experience. In S. Lewandowsky, J.C. Dunn, & Kirsner, K. (Eds.), *Implicit memory: Theoretical issues* . Hillsdale, NJ: Erlbaum. Pp. 123-138.

Masson, M.E.J., & Freedman, L. (1990). Fluent identification of repeated words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16* , 355-373.

Mattingly, I.G., & Studdert-Kennedy, M.K. (Eds.) (1991). *Modularity and the motor theory of speech perception* . Cambridge, MA: MIT Press.

McClelland, J.L., & Elman, J.L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18* , 1-86.

McClelland, J.L., & Rumelhart, D.E. (1981). An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychological Review, 88*, 375-405.

McClelland, J.L., & Rumelhart, D.E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General, 114,* 159-188.

McClelland, J.L., & Rumelhart, D.E. (1986). A distributed model of human learning and memory. In J.L. McClelland & D.E. Rumelhart (Eds.), *Parallel distributed processing*, vol 2. Cambridge, MA: MIT Press. Pp. 170-215.

McDaniel, M.A. (1981). Syntactic complexity and elaborative processing. *Memory & Cognition, 9* , 487-495.

McGehee, F. (1937). The reliability of the identification of the human voice. *Journal of General Psychology, 17* , 249-271.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746-748.

McInish, J.R., & Tikofsky, R.S. (1969). Distinctive features and response latency: A pilot study. *Perception & Psychophysics, 6* , 267-268.

Medin, D.L., & Schaffer, M.M. (1978). Context theory of classification learning. *Psychological Review, 85* , 207-238.

Metcalfe, J., & Fisher, R.P. (1986). The relation between recognition memory and classification learning. *Memory & Cognition, 14* , 164-173.

Miller, G.A. (1962). Some psychological studies of grammar. *American Psychologist, 17,* 748-762.

Miller, G.A., & Nicely, P.E. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America, 27*, 338-352.

Miller, J.D. (1989). Auditory-perceptual interpretation of the vowel. *Journal of the Acoustical Society of America, 85* , 2114-2134.

Monsell, S. (1991). The nature and locus of word frequency effects in reading. In D. Besner & G.W. Humphreys (Eds.), *Basic processes in reading* . Hillsdale, NJ: Erlbaum.

Monsen, R.B., & Engebretson, A.M. (1977). Study of variations in the male and female glottal wave. *Journal of the Acoustical Society of America, 62* , 981-993.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review, 76*, 165-178.

Morton, J. (1979). Word recognition. In Morton, J., & Marshall, J.D. (Eds.), *Psycholinguistics 2: Structures and processes* . Cambridge, MA: MIT Press. pp. 109-156.

Morton, J. (1982). Disintegrating the lexicon: An information processing approach. In Mehler, J., Walker, E.C.T., & Garrett, M. (Eds.), *On mental representation* . Hillsdale, NJ: Erlbaum.

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.

Murdock, B.B., Jr. (1962). The serial position effect in free recall. *Journal of Experimental Psychology, 64* , 482-488.

Murdock, B.B. Jr. (1967). Recent developments in short-term memory. *British Journal of Psychology, 58* , 421-433.

Murdock, B.B. Jr. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 89* , 609-626.

Musen, G., & Treisman, A. (1990). Implicit and explicit memory for visual patterns. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 127-137.

Nakatani, L., & Dukes, K. (1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America, 62* , 714-719.

Nearey, T.M. (1989). Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America, 85* , 2088-2113.

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

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In Chase, W.G. (Ed.), *Visual information processing* . New York: Academic Press. pp. 283-308.

Nickerson, R.S. (1965). Response times for "same-different" judgments. *Perceptual and Motor Skills, 20* , 15-18.

Nickerson, R.S. (1967). "Same"-"different" response times with multi-attribute stimulus differences. *Perceptual and Motor Skills, 24* , 543-554.

Nickerson, R.S. (1972). Binary classification reaction time: A review of some studies of human information processing capabilities. *Psychonomic Monograph Supplements, 4*, 275-318.

Nosofsky, R.M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 104-114.

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, 14* , 700-708.

Nosofsky, R.M. (1988a). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14* , 54-65.

Nosofsky, R.M. (1988b). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14* , 700-708.

Nosofsky, R.M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance, 17* , 3-27.

Nusbaum, H.C., & Morin, T.M. (1992). Paying attention to differences among talkers. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech perception, production, and linguistic structure* . Tokyo: IOS Press.

Oden, G.C., & Massaro, D.W. (1978). Integration of featural information in speech perception. *Psychological Review, 85* , 172-191.

Oldfield, R.C. (1966). Things, words, and the brain. *Quarterly Journal of Experimental Psychology, 18* , 340-353.

Paap, K.R., Newsome, S.L., McDonald, J.E., & Schvaneveldt, R.W. (1982). An activation-verification model for letter and word recognition: The word-superiority effect. *Psychological Review, 89* , 573-594.

Palmeri, T.J., Goldinger, S.D., & Pisoni, D.B. (1992). Episodic encoding of voice attributes and recognition memory for spoken words. Manuscript under review.

Papcun, G., Kreiman, J., & Davis, A. (1989). Long-term memory for unfamiliar voices. *Journal of the Acoustical Society of America, 85* , 913-925.

Parkin, A.J., Reid, T.K., & Russo, R. (1990). On the differential nature of implicit and explicit memory. *Memory & Cognition, 18* , 507-514.

Parkin, A.J., & Streete, S. (1988). Implicit and explicit memory in young children and adults. *British Journal of Psychology, 79* , 361-369.

Parks, T.E. (1966). Signal detectability theory of recognition performance. *Psychological Review, 73* , 44-58.

Peters, R.W. (1955a). The effect of length of exposure to speaker's voice upon listener reception. *Joint Project Report No. 44* . U.S. Naval School of Aviation Medicine, Pensacola, Florida. pp. 1-8.

Peters, R.W. (1955b). 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, Pensacola, Florida. pp. 1-9.

Peterson, G.E., & Barney, H.L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24* , 175-184.

Pike, R. (1984). Comparison of convolution and matrix distributed memory systems for associative recall and recognition. *Psychological Review, 91* , 281-294.

Pisoni, D.B. (1978). Speech perception. In W.K. Estes (Ed.), *Handbook of learning and cognitive processes* . (vol 6). Hillsdale, NJ: Erlbaum. Pp. 167-233.

Pisoni, D.B. (1990). Effects of talker variability on speech perception: Implications for current research and theory. In H. Fujisaki (Ed.), *Proceedings of the 1990 International Conference on Spoken Language Processing* . Kobe, Japan. 1399-1407.

Pisoni, D.B., & Luce, P.A. (1987). Acoustic-phonetic representations in word recognition. *Cognition, 25* , 21-52.

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

Podgorny P., & Garner, W.R. (1979). Reaction time as a measure of inter- and intraobject visual similarity: Letters of the alphabet. *Perception & Psychophysics, 26*, 37-52.

Pollack, I., Pickett, J.M., & Sumby, W.H. (1954). On the identification of speakers by voice. *Journal of the Acoustical Society of America, 26* , 403-406.

Posner, M.I. (1964). Information reduction in the analysis of sequential tasks. *Psychological Review, 71* , 491-503.

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 l earning and motivation..* New York: Academic Press.

Posner. M.I., & Keele, S.W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77* , 353-363.

Posner. M.I., & Keele, S.W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology, 83* , 304-308.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85* , 59-108.

Richardson-Klavehn, A., & Bjork, R.A. (1988). Measures of memory. *Annual Review of Psychology, 39* , 475-543.

Rock, I., & Engelstein, P. (1959). A study of memory for visual form. *American Journal of Psychology, 72* , 221-229.

Rock, I., Halper, F., & Clayton, T. (1972). The perception and recognition of complex figures. *Cognitive Psychology, 3* , 655-673.

Roediger, H.L., III. (1990). Implicit memory. *American Psychologist, 45*, 1043-1056.

Roediger, H.L., III, & Blaxton, T.A. (1987). Effects of varying modality, surface features, and retention interval on priming in word-fragment completion. *Memory & Cognition, 15*, 379-388.

Roediger, H.L., III, & Srinivas, K. (1992). Specificity of operations in perceptual priming. In P. Graf & M.E.J. Masson (Eds.), *Implicit memory: New directions in cognition, development, and neuropsychology* . Hillsdale, NJ: Erlbaum.

Rogers, T.B., Kuiper, N.A., & Kirker, W.S. (1977). Self-reference in the encoding of personal information. *Journal of Personality and Social Psychology, 13*, 677-688.

Rosch, E.H. (1973). Natural categories. *Cognitive Psychology, 4* , 328-350.

Rothkopf, E.Z. (1971). Incidental memory for location of information in text. *Journal of Verbal Learning and Verbal Behavior, 10* , 608-613.

Rudnicky, A.I., & Kolers, P.A. (1984). Size and case of type as stimuli in reading. *Journal of Experimental Psychology: Human Perception and Performance, 10* , 231-249.

Rumelhart, D.E., & McClelland, J.L. (1982). An interactive activation model of context effects in letter perception: Part II. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review, 89* , 60-94.

Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology, 89* , 43-50.

Safer, M.A., & Leventhal, H. (1977). Ear differences in evaluating emotional tones of voice and verbal content. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 75-82.

Salasoo, A., Shiffrin, R.M., & Feustel, T.C. (1985). Building permanent memory codes: Codification and repetition effects in word identification. *Journal of Experimental Psychology: General, 114* , 50-77.

Samuel, A.G. (1986). The role of the lexicon in speech perception. In Schwab, E.C., & Nusbaum, H.C. (Eds.), *Perception of speech and visual form: Theoretical issues, models, and research* . New York: Academic Press. pp. 89-111.

Samuel, A.G., & Ressler, (1986). Attention within auditory word perception: Insights from the phonemic restoration illusion. *Journal of Experimental Psychology: Human Perception and Performance, 12* , 70-79.

Saslove, H., & Yarney, A.D. (1980). Long-term auditory memory: Speaker identification. *Journal of Applied Psychology, 65* , 111-116.

Scarborough, D.L., Cortese, C., & Scarborough, H.S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 1-17.

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

Schacter, D.L. (1990). Perceptual representation systems and implicit memory: Toward a resolution of the multiple memory systems debate. In A. Diamond (Ed.), *Annals of the New York Academy of Sciences, 608* , 543-571.

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

Schacter, D.L., Eich, J.E., & Tulving, E. (1978). Richard Semon's theory of memory. *Journal of Verbal Learning and Verbal Behavior, 17* , 721-743.

Schacter, D.L., & Graf, P. (1989). Modality specificity of implicit memory for new associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 3-12.

Semon, R. (1909/1923). *Mnemic psychology.* (B. Duffy, translator). London: George Allen & Unwin.

Sergent, J., & Takane, Y. (1987). Structures in two-choice reaction-time data. *Journal of Experimental Psychology: Human Perception and Performance, 13* , 300-315.

Shankweiler, D.P., Strange, W., & Verbrugge. R.R. (1976). Speech and the problem of perceptual constancy. In R. Shaw, & J. Bransford (Eds.), *Perceiving, Acting, and Comprehending: Toward and Ecological Psychology.* Hillsdale, N.J.: Erlbaum.

Shepard, R.N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior, 6* , 156-163.

Shepard, R.N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science, 210*, 390-398.

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

Sherry, D.F., & Schacter, D.L. (1987). The evolution of multiple memory systems. *Psychological Review, 94* , 439-454.

Smith, E.R. (1990). Content and process specificity in the effects of prior experiences. In T.K. Srull & R.S. Wyer Jr. (Eds.), *Advances in Social Cognition, Vol. III* . Hillsdale, NJ: Erlbaum. Pp. 1-59.

Smith, E.R., & Zarate, M.A. (1992). Exemplar-based model of social judgment. *Psychological Review, 99* , 3-21.

Smith, L.B. (1989). A model of perceptual classification in children and adults. *Psychological Review, 96* , 125-144.

Smith, S.M., Glenberg, A., & Bjork, R.A. (1978). Environmental context and human memory. *Memory & Cognition, 6* , 342-353.

Sommers, M.S., Nygaard, L.C., & Pisoni, D.B. (1992). The effects of stimulus variability on perceptual identification: Rate, talker, and amplitude. *Research on Speech Perception Progress Report No. 17*. Bloomington, IN: Indiana University.

Springer, S.P. (1973). Memory for linguistic and nonlinguistic dimensions of the same acoustic stimulus. *Journal of Experimental Psychology, 101* , 159-163.

Standing, L., Conezio, J., & Haber, R.N. (1970). Perception and memory for pictures: Single-trial learning of 2560 visual stimuli. *Psychonomic Science, 19* , 73-74.

Stevens, K.N., Williams, C.E., Carbonell, J.R., & Woods, B. (1968). Speaker authentication and identification: A comparison of spectrographic and auditory presentations of speech material. *Journal of the Acoustical Society of America, 44* , 1596-1607.

Strange, W., Verbrugge, R.R., Shankweiler, D.P., & Edman, T.R. (1976). Consonant environment specifies vowel identity. *Journal of the Acoustical Society of America, 60*, 213-221.

Studdert-Kennedy, M. (1976). Speech perception. In N.J. Lass (Ed.), *Contemporary Issues in Experimental Phonetics* . New York: Academic Press.

Summerfield, A.Q. (1975). Information-processing analyses of perceptual adjustments to source and context variables in speech. Unpublished doctoral dissertation. Belfast, Northern Ireland: Queen's University of Belfast.

Summerfield, Q., & Haggard, M.P. (1973). Vocal tract normalisation as demonstrated by reaction times. *Report on Research in Progress in Speech Perception, 2*, Belfast, Northern Ireland: The Queen's University of Belfast.

Syrdal, A.K., & Gopal, H.S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America, 79* , 1086-1100.

Taft, M., & Hambly, G. (1986). Exploring the Cohort Model of word recognition. *Cognition, 22* , 259-282.

Tanenhaus, M.K., & Lucas, M.M. (1987). Context effects in lexical processing. *Cognition, 25* , 213-234.

Tardif, T., & Craik, F.I.M. (1989). Reading a week later: Perceptual and conceptual factors. *Journal of Memory and Language, 28* , 107-125.

Tomiak, G.R., Green, K.P., & Kuhl, P.K. (1991). Phonetic coding and its relationship to talker and rate normalization. Paper presented at 122nd meeting of the Acoustical Society of America, Houston, TX.

Tosi, O. (1979). *Voice identification: Theory and legal implications.* Baltimore, MD: University Park Press.

Treisman, M. (1978). A theory of the identification of complex stimuli with an application to word recognition. *Psychological Review, 85* , 525-570.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* . New York: Academic Press. Pp. 382-403.

Tulving, E. (1983). *Elements of episodic memory.* London: Oxford University Press.

Tulving, E., & Bower, G.H. (1974). The logic of memory representations. In G. Bower (Ed.), *The Psychology of Learning and Motivation* . Academic Press. Pp. 265-301.

Tulving, E., & Schacter, D.L. (1990). Priming and human memory systems. *Science, 247*, 301-306.

Tulving, E., Schacter, D.L., & Stark, H.A. (1982). Priming effects in word fragment completion are independent of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8* , 336-342.

Tyler, L.K. (1984). The structure of the initial cohort: Evidence from gating. *Perception & Psychophysics, 36* , 417-427.

Tyler, L.K., & Marslen-Wilson, W.D. (1982). Speech comprehension processes. In Mehler, J., Walker, E.C.T., & Garrett, M. (Eds.), *Perspectives on mental representation: Experimental and theoretical studies of cognitive processes and capacities.* Hillsdale, NJ: Erlbaum.

Underwood, B.J. (1969). Attributes of memory. *Psychological Review, 76* , 559-573.

Underwood, B.J., & Eckstrand, B.R. (1968). Linguistic associations and retention. *Journal of Verbal Learning and Verbal Behavior, 7* , 162-171.

Van Lancker, D., & Kreiman, J. (1987). Voice discrimination and recognition are separate abilities. *Neuropsychologica, 25* , 829-834.

Van Lancker, D., Kreiman, J., & Emmorey, K. (1985). Familiar voice recognition: Patterns and parameters. Part I: Recognition of backward voices. *Journal of Phonetics, 13*, 19-38.

Van Lancker, D., Kreiman, J., & Wickens, T.D. (1985). Familiar voice recognition: Patterns and parameters. Part II: Recognition of rate-altered voices. *Journal of Phonetics, 13*, 39-52.

Verbrugge, R.R., & Rakerd, B. (1986). Evidence of talker-independent information for vowels. *Language and Speech, 29* , 39-57.

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.

Warren, P., & Marslen-Wilson, W.D. (1987). Continuous uptake of acoustic cues in spoken word recognition. *Perception & Psychophysics, 41* , 262-275.

Warren, R.M. (1970). Perceptual restoration of missing speech sounds. *Science, 167*, 392-393.

Warrington, E.K., & Weiskrantz, L. (1968). New method of testing long-term retention with special reference to amnesic patients. *Nature, 217* , 972-974.

Watkins, A.J. (1991). Central, auditory mechanisms of perceptual compensation for spectral-envelope distortion. *Journal of the Acoustical Society of America, 90* , 2942-2955.

Watkins, M.J. (1990). Mediationism and the obfuscation of memory. *American Psychologist, 45* , 328-335.

Weiner, F.F., & Singh, S. (1974). Multidimensional analysis of choice reaction time judgments on pairs of English fricatives. *Journal of Experimental Psychology, 102* , 615-620.

White, G.L., & Shapiro, D. (1987). Don't I know you? Antecedents and social consequences of perceived familiarity. *Journal of Experimental Social Psychology, 23* , 75-92.

Whittlesea, B.W.A. (1987). Preservation of specific experiences in the representation of general knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 3-17.

Whittlesea, B.W.A., Jacoby, L.L., & Girard, K. (1990). Illusions of immediate memory: Evidence of an attributional basis for feelings of familiarity and perceptual quality. *Journal of Memory and Language, 29* , 716-732.

Wickelgren, W.A. (1969). Context-sensitive coding, associative memory, and serial order in (speech) behavior. *Psychological Review, 76* , 1-15.

Winograd, E. (1968). List differentiation as a function of frequency and retention interval. *Journal of Experimental Psychology, 76* .

Woltz, D.J. (1990). Repetition of semantic comparisons: Temporary and persistent priming effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 392-403.

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

Young, F. (1984). Scaling. *Annual Review of Psychology, 35* , 55-81.

Zacks, R.T., Hasher, L., & Sanft, H. (1982). Automatic encoding of event frequency: Further findings. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*, 106-116.

Zadeh, L.A. (1965). Fuzzy sets. *Information and Control, 8* , 338-353.

Zaidel, E. (1978). Lexical organization in the right hemisphere. In P. Buser & A. Rougeul-Buser (Eds.), *Cerebral correlates of conscious experience* . Amsterdam: Elsevier.

Zechmeister, E.B., & McKillip, J. (1972). Recall of place on the page. *Journal of Educational Psychology, 63* , 446-453.

Zechmeister, E.B., McKillip, J., Pasko, S., & Bespalec, D. (1975). Visual memory for place on the page. *Journal of General Psychology, 92* , 43-52.