INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Auditory Repetition Priming: A Perceptual Memory of Speech

by

Maura Pilotti

A dissertation submitted to the Graduate Faculty in Psychology in partial fulfillment of the requirements of the degree Doctor of Philosophy, The City University of New York

1997

Approval Page

This manuscript has been read and accepted for the Graduate Faculty in Psychology in satisfaction of the dissertation for the degree of Doctor of Philosophy.

12-19-96
Date

Chair of Examining Committee

12-18-96
Date

Executive Officer

John S. Antrobus

Martin Chodorow

David Pisoni

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

Auditory Repetition Priming: A Perceptual Memory of Speech

by

Maura Pilotti

Adviser: Professor John S. Antrobus

The primary purpose of this investigation was to test the hypothesis that implicit memory for spoken words results from the operations of a memory system devoted to pattern recognition. Pattern recognition entails pre-lexical processes, which produce a detailed record of the acoustic characteristics of a speaker's voice along with subword information. To test this hypothesis, the relation between speech perception and implicit memory was examined. Auditory repetition priming was employed as the behavioral expression of implicit memory. Phoneme monitoring latencies were used to measure speech perception and memory processes. In addition to testing the main hypothesis, Experiment 1 investigated the detrimental effect of speaker variability on memory; Experiment 2 examined whether expectations can modulate this effect; and Experiment 3 assessed the effect of voice familiarity. The results showed that although speech perception operations entailed the use of word information, implicit memory relied on pre-lexical operations. Consistent with this finding, the number of speakers comprising the acoustic environment was found to produce a linear decrement in auditory priming. Stable acoustic changes in a multiple-speaker environment during the encoding of a set of spoken words rather than expectations regarding these changes were found to modulate the detrimental effect of speaker variability on memory. Voice familiarity, however, did not improve memory. Taken together, the present findings indicate that implicit memory of spoken words entails the encoding, storage and retrieval of speaker-specific episodic representations, which are pre-lexical in nature.

Acknowledgments

First and foremost, I would like to thank John Antrobus for his unwavering support, patience, and encouragement throughout my graduate career. Professor Antrobus is a model of unselfish mentorship, bottomless tolerance, and devoted pursuit of scientific knowledge. The other members of my committee, David Pisoni and Martin Chodorow, have earned a special mention for their constructive criticism. Edward Meehan deserves recognition for supporting and encouraging my work. I owe a handful of thanks to those students who have helped in procuring the participants for the various experiments. Finally, I would like to mention the remarkable members of the Department of Psychology at the College of Staten Island and IBR: Andy Fuller, Ekkhart Trenkner, Tina Luciano, Anita Conte, Lori Caudale, Rima Blair, Sonia Ragir, and Irene Deich, who deserve a special recognition for their support and encouragement.

Work in this dissertation was partially supported by a fellowship of the Institute of Basic Research for Developmental Neuroscience and Developmental Disabilities at the College of Staten Island, and by the APA's dissertation research award.

Table of Contents

List of Tables

Chapter I: A Brief Introduction

Purpose. The specific purpose of this investigation is to test the hypothesis that implicit memory for spoken words is the product of a memory system devoted to pattern recognition. As a result of speech perception, the pattern-recognition processes expressed by this system are assumed to develop long-term memory representations of spoken words, which, at a later point in time, are unintentionally retrieved to aid the perception of those same words. These implicit memory representations, hypothesized to contain speaker-dependent acoustic information along with phonological information at the subword level, can be thought of as pre-lexical episodic records of spoken words.

The pattern-recognition processes, hypothesized to generate implicit memory of spoken words, are assumed to be easily disrupted by an acoustic environment that requires the encoding of multiple voices, to be facilitated by voice familiarity, but unlikely to be affected by higher-order cognitive processes, such as expectations.

Problem. Several studies have shown that speech perception and memory of spoken words entail a record of a speaker's voice along with abstract-symbolic information (see Goldinger, 1992). Although these studies have been instrumental in identifying the episodic nature of implicit memory, they have not clearly separated the contribution of pattern-recognition processes to implicit memory from the lexical processes ingrained in the procedures selected to conduct such studies. The present investigation contains three experiments in which this impasse is circumvented.

Significance. The understanding of implicit memory for spoken words is particularly crucial to several research questions that the literature on memory has touched. It addresses

the question of whether memory should be regarded as a set of abstract-symbolic or episodic stimulus-specific representations of a stimulus. It provides insights into the nature of the information upon which the processes of encoding, retention, and retrieval operate. It addresses the question of whether the very same processes are to be considered automatic rather than modulated by higher-level cognitive processes.

Although some may argue that the distinction between implicit and explicit memory, primarily based on procedural differences, is an unwarranted over-generalization (J. S. Antrobus, personal communication, May 9, 1996; Willinghan & Preuss, 1995), implicit procedures remain a useful means for addressing the research questions that have been crucial to the understanding of memory. In fact, implicit procedures, by restricting the intervention of conscious processes in all the different stages of processing from encoding and retention to retrieval, provide a window on the basic functional and structural properties of memory for spoken words.

Summary. The following chapter presents a review of the pertinent literature to acquaint the reader with the empirical and theoretical issues that justify the present investigation. It begins with a brief outline of the implicit-explicit memory distinction followed by an overview of the current theoretical controversies in the conceptualization of implicit memory processes (Section 1). As memory results from perceptual processes, the literature on speech perception and spoken word recognition is then reviewed (Section 2) followed by a critical examination of the literature on explicit memory (Section 3) and implicit memory (Section 4) for spoken words and voices. This review concentrates on questioning the traditional idea that perception and long-term memory of spoken words result from the

activation of abstract-symbolic representations in which speaker-dependent information in the speech stimulus is assumed to be discarded as noise. In light of the critical role played by surface information such as voice characteristics in perception and memory of spoken words, the processes involved in the encoding, retention and retrieval of voice information are then examined (Section 5). A survey of the literature on visual word recognition attempts to establish a parallel between the auditory and the visual domains so as to ascertain whether superficial characteristics of a stimulus play a relevant role in reading as they do in spoken word recognition (Section 6). The latter sections are meant to set the stage for the present investigation. First an outline of two contrasting models of implicit memory found in the literature is presented along with an examination of the evidence that supports or opposes each model (Section 7). Subsequently, the same evidence is critically examined by considering the limitations ingrained in the experimental procedures employed to test these models (Section 8). On the basis of the evidence examined in the previous sections, a preliminary model is proposed and a series of predictions that result from this model is introduced (Section 9). Chapter III contains the experiments that test these predictions and Chapter IV presents a comprehensive discussion of the findings in light of the model proposed in Chapter II.

Chapter II: Review of the Literature

1. Implicit Memory: General

Implicit/Explicit Memory Dichotomy. One of the fundamental concerns of cognitive science is to elucidate the nature of both the functional and structural properties of long-term memory (Schacter, 1992). Models of long-term memory address this concern by arranging experimental findings into explanatory constructs. Experimental procedures, by constraining the available evidence to a predetermined set of processes, are a critical component of these findings. In fact, until recently, the investigation of human memory has used procedures, such as recall and recognition, which require the conscious recollection of prior events (Schacter, 1987; Schacter, Chiu, & Ochsner, 1993) and it has thus regarded the impact of prior experiences on observable behavior as constrained by the effortful and conscious process of retrieval.

The study of long-term memory has recently come to acknowledge the notion that the memory of a stimulus can affect performance on a subsequent presentation of that stimulus without any attempts at retention and intentional recollection (Schacter, 1987; Schacter, Chiu, & Ochsner, 1993). As in the more traditional techniques of recall and recognition, procedures in which this effect is observed entail a study and a test phase. During study, a stimulus set is presented without reference to a subsequent memory test so that retention is assumed to be unintentional. At test, a response to specific characteristics of the repeated stimuli is required without entailing effortful recollection. Facilitation in test performance as a function of prior exposure is attributed to the information acquired during study. Commonly cited examples of these procedures are

word identification, which entails recognition of words in a background noise, and word-stem completion, which solicits the completion of a stem with the first word that comes to mind (Goldinger, 1992; Schacter & Church, 1992).

Although no passive window allows for the examination of the functional and structural properties of memory without affecting them to some extent, these procedures have the advantage of being less intrusive than recall and recognition. They provide a useful tool for investigating automatic as opposed to strategically-driven memory processes so that a less adulterated view of the basic functional and structural substrate of memory can be obtained.

The present investigation acknowledges that a bicameral classification scheme appears feasible in its general claim that implicit processes can be distinguished from explicit recollection (Schacter, 1992). A bicameral classification scheme, however, disregards the idea that there may be several unique implicit processes, as many as the entire range of stimulus materials and procedures that can be used to investigate implicit processing (Willinghan & Preuss, 1995). An in-depth discussion of the multiple-domain classification issue is beyond the scope of the present manuscript so that a provisionary definition of implicit processes as those that emerge in implicit memory procedures will be adopted here (see Schacter, 1992a, 1992b, 1987). In the process of differentiating between explicit and implicit procedures, however, one encounters the problem of defining the general processes that underlie performance under these two experimental paradigms. According to Jacoby (Jacoby, 1991; Jacoby, Toth, & Yonelinas, 1993) and Mandler (1980), explicit memory performance is based on two independent operations: an

unconscious process of familiarity and an explicit process of recollection. Familiarity refers to the implicit effect of prior exposure to a stimulus. On the other hand, recollection defines the intentional use of memory. Performance on implicit memory tasks would then be the exclusive product of familiarity, if implicit tasks were not contaminated by recollection (Jacoby, 1991; Reingold & Merikle, 1990). The conceptualization of familiarity as independent from recollection, however, is a difficult claim to hold. Although Jacoby (1991) and Mandler (1980) appear to be justified in implying that the two processes are qualitatively different, their *independence* model is not adequately supported by the evidence. Joordens and Merikle (1993) propose a *redundancy* model in which performance is supported either by familiarity alone or by both familiarity and explicit recollection. Following Joordens and Merikle's *redundancy* model (1993), in the present investigation, explicit procedures are assumed to result from recollection exploiting familiarity. On the other hand, implicit procedures are conceptualized as relying primarily on familiarity. The term "primarily" is employed to account for attempts at recollection that are likely to contaminate performance in implicit memory techniques free from time constraints, such as word stem completion (Jacoby, 1991; Reingold & Merikle, 1990; Richardson-Klavehn & Bjork, 1988). It is assumed that the more disguised the implicit memory test and the more pressing the time constraints, the less likely explicit recollection is to contaminate performance.

Implicit Memory and Repetition Priming. In implicit memory procedures, *repetition priming* denotes the observable effect of a memory of a previously displayed stimulus. Assumed to result from familiarity, repetition priming emerges from diverse experimental

paradigms. The effect of prior exposure appears in a variety of procedures, besides word identification (Jacoby & Dallas, 1981) and word-stem completion (Warrington & Weiskrantz, 1974). Repetition facilitates lexical decisions (Scarborough, Cortese, & Scarborough, 1977), naming a printed word (Durso & Johnson, 1979), completion of word fragments (Tulving, Schacter, & Stark, 1982), reading sentences (Kolers & Ostry, 1974) and proofreading (Levy & Begin, 1984). These procedures are classified under the label *implicit* memory tasks to distinguish them from traditional recall and recognition, which, by requiring effortful recollection, are now identified as *explicit.*

Repetition priming occurs even when a stimulus is masked so that identification is unlikely to occur (Forster & Davis, 1984). Although repetition still facilitates identification performance, the effect is fleeting and disappears if other stimuli are inserted between the first exposure to the stimulus and its repetition. Alternatively, when the stimulus is not masked and thus can be identified, repetition priming persists over significant intervals that range from several days (Masson, 1984) to a year (Kolers, 1976). It is this latter form of priming that is of primary interest to the investigation of long-term memory. A fleeting effect is an "activity trace" (Hebb, 1961; Miller, Kasprow, & Schachtman, 1986), which is the primary domain of perception. A "structural trace" is a change that outlasts the fleeting effect of an activity trace and thus provides a technically useful window for understanding how memory is functionally and structurally organized. In the present investigation, the terms "trace" and "representation" are used interchangeably to signify a change in the human processing system as a function of an incoming stimulus.

Implicit Memory and Memory Processes. Two related dichotomous conceptualizations have been proposed to describe the memory processes that, following prior exposure to a stimulus, produce implicit memory representations and thus are responsible for "familiarity" at test: *modification* versus *acquisition* models and *abstractionist* versus *episodic* models. Although these dichotomies are largely isomorphic, the former emphasizes either the constructive or passive nature of the processes responsible for implicit memory representations, while the latter acknowledges this distinction, but places emphasis on the content of the memory representations.

Modification models portray repetition priming as the result of the "reactivation" of preexisting memory representations and their subsequent strengthening (Graf & Mandler, 1984; Morton, 1969). *Acquisition* models depict repetition priming as dependent upon the generation of an episodic memory trace. This trace may rely on preexisting information, but is unique because it contains information specific to the prior occurrence of that stimulus (Jacoby, 1983a, 1983b; Schacter, 1990; Squire, 1987). Contrary to the predictions of a modification model, an acquisition model assumes that repetition priming would extend to novel stimuli that do not have preexisting representations in memory. Although repetition priming has been observed with "novel" stimuli such as pseudowords (Bowers, 1994; see Rugg & Nagy, 1987, for contrasting findings), the meaning of "novel stimuli" is controversial, and thus results in a claim difficult to test. For example, relative to words, legal pseudowords may be considered novel stimuli in the sense that are not part of a person's mental lexicon. They may be characterized, however, in terms of preexisting subword sequences (Seidenberg & McClelland, 1989). Illegal pseudowords (e.g., xyks)

may be more novel, but they can still be identified in terms of preexisting letters. Even if one accepts the idea of "novel stimuli" in a relative sense, the evidence of repetition priming for illegal pseudowords is not sufficiently robust to support either model. In fact, Bowers (1994) found repetition priming with illegal pseudowords while other researchers did not (Dorfman, 1994; Rugg & Nagy, 1987; Schacter, Rapcsak, Rubens, Tharan, & Laguna, 1990).

Abstractionist models view repetition priming as dependent upon the reactivation of abstract representations (Carr & Pollatsek, 1985). The term "abstract" refers to representations that, by being generalizations over many individual episodes, do not contain information specific to the occurrence of a stimulus. *Episodic* models, on the other hand, postulate that the exposure to a stimulus, such as a word, produces the encoding and retention of that word as a unique and idiosyncratic memorial representation (Jacoby & Witherspoon, 1982; Hintzman, 1986; Kolers, 1975, 1976; Underwood, 1969). Because this representation contains perceptual details, the superficial similarity between the current and prior presentation of a stimulus is a critical aspect of priming. Mixed models exist that account for priming in terms of both abstract and episodic specific information (see McClelland & Rumelhart, 1985) depending on the superficial similarity between prior and current stimulus and other factors such as number and type of intervening stimuli. According to mixed models, the more physically similar the two stimuli are, the more likely priming will rely on the episodic trace left by the previous stimulus rather than on abstract information. Thus, abstractionist models can be separated from episodic and mixed models on the basis of their denial that priming possesses an episodic denotation.

The critical test for these models results from investigations that have changed the surface characteristics of a stimulus word from the study to the test phase. This experimental manipulation epitomizes one of the basic purposes of the present investigation, which is to determine the nature of the processes and representations that support repetition priming in spoken word recognition. Prior to a discussion of the findings following study-to-test surface changes in a repeated stimulus, the nature of speech perception operations will be addressed.

2. Perception of Spoken Words

A necessary but not sufficient condition for the understanding of implicit memory for spoken words is the comprehension of speech perception operations. The perception of a spoken word is herein meant to refer to two sets of processes: pattern recognition processes, which lead to the identification of an acoustic stimulus as an ordered sequence of phonemes with the distinctive "style" of a given voice, and lexical processes, which entail recognizing this sequence as a meaningful utterance (Pisoni & Luce, 1986; Pisoni, Nusbaum, Luce, & Slowiaczek, 1985). Throughout the present manuscript, the traditional dichotomy between speech perception and word recognition will be disregarded as unnecessary for the purpose of investigating memory of spoken words. Thus, the term "speech perception" will be employed here as a comprehensive label to include both pattern recognition and lexical activation.

The starting assumption for understanding the processes involved in the perception of spoken words is that an acoustic pattern distributed over time becomes, in the listener's experience, a word when it activates or generates a set of memory representations (Miller

& Eimas, 1995). The number of such representations is at the present time unknown, but one can safely isolate those that pertain to pattern recognition operations and those that concern lexical operations. Note that other representations may also be activated following the presentation of a spoken word, such as orthographic and syntactic, but they are beyond the scope of this manuscript.

The acoustic pattern, which is easily recognized as a word, however, is highly variable across and within speakers' voices. Despite the acoustic idiosyncracies of each voice, listeners easily decode several acoustic patterns as instances of the same familiar sequence of phonemes, which they have come to know as having a distinctive meaning. Even infants' perception of vowels, consonants, and syllables does not appear to be affected by the acoustic idiosyncracies of different speakers (Kuhl, 1979; Kuhl & Miller, 1982; Jusczyk, Pisoni, & Mullenix, 1992). The question is how infants as well as adults handle the variability intrinsic to spoken words, and manage to introduce in their perceptual experience a sense of phonemic stability. Consistent with Neisser's view that literal representations of acoustic stimuli rapidly fade (1976), until recently, speaker variability was conceptualized as noise. Speech perception was thought to entail a normalization process which adjusted for the variability of different voices by filtering it out as a needless byproduct of a noisy acoustic environment (see Pisoni, 1996). In this sense, normalization entailed both generalization, which allowed the listener to treat as identical the superficially different instances of an abstract-symbolic representation, and loss of information. The postulation of a normalization process, conceptualized as a filtering device, did not deny the listener's ability to process voice information. It simply

suggested that speech perception did not entail voice processing. Voice and phonemic information were simply handled independently (Liberman & Mattingly, 1985).

This traditional interpretation of how speech perception operates in the face of acoustic variability has led to several theories in which phonemic invariance is established at the expense of the superficial characteristics of spoken words, such as the speaker's voice. Current evidence has questioned this conceptualization by rejecting the idea that the fading of acoustic representations, suggested by Neisser (1967), requires that voice information be lost. Madden and Bastian (1977) demonstrated that voice changes (male versus female) between a memory set and a subsequent probe reduce recognition accuracy up to a delay of 4 seconds (secs). The study employed sets of six dichotically presented digits, which, after an interval of either 0.5, 1.4, or 4.0 secs, were followed by probe digits. On each trial, participants had to judge whether a probe digit had appeared in the preceding set. For intervals up to 1.4 secs, decrements in accuracy in different-voice trials could be accounted for by biases in responding "absent". At the longest delay, however, the decrement in accuracy was determined by reduced sensitivity. Rather than disappearing, voice characteristics, incorporated in a spoken word, persist in the memory representations that emerge from that word.

At the present time, speech perception is seen as producing more detailed representations of spoken words than was once thought. One of the sources of this complexity is the encoding and storage of a speaker's voice. It follows that speech perception, at least when an unfamiliar voice is heard, combines a process of activation of memory representations regarding phonemic sequences and meaning with a process of

generation of internal representations of the acoustic characteristics of a speaker's voice. The existence of this latter process and its functional relation with the operations that produce phonemic sequences are reviewed in the following paragraphs.

Voice Information and Pre-Lexical Effects: The Evidence. Several studies have reported that changes in the speaker's voice between temporally contiguous stimuli reduce accuracy and increase processing time in the identification of both vowels and consonants (Allard & Henderson, 1975; Assman, Nearey, & Hogan, 1982; Cole, Coltheart, & Allard, 1974; Ladefoged & Broadbent, 1957; Summerfield & Haggard, 1973). The findings that phonemic processing is impaired by speaker variability question the long-standing assumption, inherent to the postulation of the normalization process, that voice and phonemic processes are functionally independent. A recent study by Mullenix and Pisoni (1990) attempted to understand these findings by directly testing the hypothesis of functional independence. In this study, a selective attention paradigm was applied to a two-choice speeded classification task (Garner, 1974; Garner & Felfoldy, 1970). Participants were instructed to respond to a given dimension of a word stimulus while simultaneously ignoring the changes in another dimension. The two dimensions manipulated were voice and word-initial phoneme. It was thought that if voice and phonemic processes are integral aspects of speech perception, then changes in the unattended dimension should affect performance in the attended dimension. Mullenix and Pisoni found that participants processed these two dimensions in a dependent manner, and that it was more difficult for them to attend to the phonemic aspect while ignoring word-to-word voice changes than the reverse. It follows from these results that if the processing of voice characteristics is not

a listener's option, a re-conceptualization of the postulated speech normalization operations is demanded.

These studies have been instrumental in questioning the idea of normalization as a filtering process. Both acoustic voice-related characteristics and phonemic information are relevant aspects of spoken words. There is no apparent reason for postulating that the human processing system would disregard one in favor of the other (Hyde, 1972). If any normalization process is to be postulated to account for phonemic constancy, it would be better conceptualized as "compensatory process", through which the nature of the adjustments made to deal with acoustic variability is recorded as voice-related information (see Green, Kuhl, Meltzoff, & Stevens, 1991). Normalization would thus denote generalization without concurrent reduction of information. Voice-related information contained in memory could then be used to aid the perception of a speaker's subsequent utterances. Some investigations of vowel perception, however, have suggested that vowels of different speakers share speaker-independent information that renders the normalization process, either as filtering or compensatory device, an unnecessary postulation (Verbrugge, Strange, Shankweiler, & Edman, 1976; Verbrugge & Rakerd, 1986).

All of the above studies have used experimental procedures and materials that emphasize pre-lexical levels of speech processing. Thus, they have indirectly portrayed the idea that between-stimulus voice effects are confined to pre-lexical levels.

Voice Information and Lexical Effects: The Evidence. Several investigations have shown that inter-speaker variability affects word recognition. The most common procedures

employed to investigate word recognition are word identification and naming. They require the presentation of words that may vary in frequency, familiarity and lexical density (e.g., the number of words that differ from a given word by one phoneme substitution, see Greenberg & Jenkins, 1964). In word identification, words are heard at different signal-to-noise ratios or degradation levels (Horrii, House, & Hughes, 1971; Salasoo & Pisoni, 1985), and participants are generally asked to type on a computer keyboard the words that they have heard. Accuracy of identification is measured. In naming, participants repeat the words they have heard. Response latencies and accuracy of identification are recorded. The advantage of the latter procedure is that it allows for an examination of the time-course of a perceptual process (see Mullenix, Pisoni, & Martin, 1989).

One of the earliest studies to indicate that the effects of a multiple-speaker acoustic environment are not confined to pre-lexical processing was an investigation conducted by Creelman (1957). His procedure entailed measuring the accuracy with which monosyllabic words presented in noise were identified. Creelman found that words presented in multiple-speaker lists were identified less accurately than the same words in single-speaker lists. Other experiments, conducted after Creelman's study (see Pisoni, 1990, for a review) have consistently reported two findings. First, word recognition, both in terms of accuracy and response latency, is impaired by multiple-speaker stimulus materials. Second, this effect is more robust than that produced by word frequency and lexical density, factors that have been traditionally recognized as important in spoken word identification (Mullenix, et al., 1989). Voice familiarity, however, can improve word identification in a multiple-speaker acoustic environment. Nygaard, Sommers, and Pisoni (1994, 1996)

trained participants for several days to explicitly recognize a set of speakers' voices. Following voice training, participants were asked to identify novel words presented at different signal-to-noise ratios. Identification accuracy was found to be higher for novel words spoken by familiar speakers than for the same novel words spoken by unfamiliar speakers (note that this effect is limited to conditions in which training and test materials overlap in type so that either words or sentences are used in both phases).

The evidence reviewed above strongly suggests that speaker variability affects lexical processing. The finding, however, does not warrant unambiguous conclusions regarding the locus and nature of voice effects. Even if word identification is impaired by speaker variability, it is still possible for some normalization process to operate at the early pre-lexical level. The effect of speaker variability on word identification would then be the byproduct of an early impairment passed on to higher levels of processing. This impairment could be conceptualized as stemming from the additional operations necessary to remove voice-related noise or to compensate for such noise without long-term retention of voice characteristics. Alternatively, it could be interpreted as stemming from the additional operations necessary to encode and perhaps store specific voice characteristics. These additional operations could be assumed to be processes of compensation that, as a byproduct of their workings, produce voice-related traces (see Johnson, 1990, and Miller, 1989, for qualifications on vowel perception) or processes directly involved in gathering speaker-dependent acoustic information (Goldinger, 1992). Although the alternative conceptualizations within each model of speech perception in a multiple-speaker acoustic environment are still grounds for debate, experiments concerned with memory of

spoken words and voices have tested the idea of voice as noise as opposed to voice as useful information retained in memory to aid perceptual and memory processes.

3. Explicit Memory of Spoken Words and Voices: The evidence

Explicit Memory for Words: The Evidence. Several investigations have studied the effect of multiple-speaker acoustic environments on memory for spoken words to determine whether voice characteristics are a component of the long-term representations of spoken words. A set of experiments used serial-ordered recall to study memory for lists of words (Goldinger, Pisoni, & Logan, 1991; Logan & Pisoni, 1987; Martin, Mullenix, Pisoni, & Summers, 1989). This procedure entails presentation of words in single- or multiple-speaker lists, followed by recall of the words in the order of their presentation in the lists. Baseline recall performance is provided by the single-speaker lists. The findings showed that, compared to a single-speaker, the multiple-speaker acoustic environment impaired recall of words heard in the primacy portion of the lists to be recalled. No effects were detected in the middle or recency portions (Goldinger, et al., 1991; Logan & Pisoni, 1987; Martin, et al, 1989). In Goldinger, et al.'s investigation (1991), when the inter-stimulus interval (ISI) between successive words was lengthened from 250 to 4,000 msecs, words in the primacy portion of multiple-speaker lists were actually recalled more accurately than the same words in single-speaker lists.

These serial-recall findings support a model of speech perception in which voice is encoded and retained in long-term memory. Mullenix and Pisoni's findings (1990) suggest that voice characteristics are concurrently processed with word information. At short delays between contiguous words, it is reasonable to hypothesize that insufficient

processing time hinders the encoding of both voice and word information in the primacy portion of a multiple-speaker list. Consequently, retrieval of the words in this portion is impaired. At longer delays, voice information, once encoded, is used to aid recall. Obviously, multiple voices can better assist recall than one voice. Because words occurring early in the list are more likely to have been thoroughly processed than words occurring later, voice information can serve as a more effective retrieval cue in the primacy portion of the multiple-speaker list.

A competing model assumes that voice characteristics disrupt the encoding of word information by requiring additional processing time to normalize across speakers' voices. At short delays, insufficient processing time would impair the processing of words in the primacy portion of the list, and thus their recall. At longer delays, the detrimental effect of the normalization process should be attenuated. The encoding and thus recall of multiple- and single-speaker word lists across all the serial positions should be indistinguishable. Although the evidence at short delays may be interpreted as supporting the "speaker normalization" model, the finding at longer delays is clearly inconsistent with this view.

Additional support for the proposal that voice characteristics along with word information comprise the long-term memory record of speech perception comes from a set of experiments using a continuous-recognition memory procedure (see Craik & Kirsner, 1974; Palmeri, Goldinger, & Pisoni, 1993). In continuous recognition memory (Shepard & Teghtsoonian, 1961), participants hear a list of spoken words. The number of speakers may vary from two to several. Some of these words are subsequently repeated

within the same list by either the same or a different speaker. The interval between the first presentation of a word and its subsequent repetition may vary from zero (no lag) to several intervening words (64 was the maximum number of intervening words in Palmeri, et al.'s study). Participants are asked to discriminate between repeated (old) and non-repeated (new) words. In one version of the task, the speakers' voices are irrelevant to the judgment to be made. In another version, participants are required to judge for each word identified as "old", whether or not the speaker is the same.

Craik and Kirsner's study (1974) employed two speakers, a male and a female, and lags of 1 to 32 intervening words. The superior recognition accuracy of same-voice repetitions as opposed to different-voice strongly supports the idea that voice and word information, processed in parallel during speech perception, form combined rather than independent representations in long term-memory. Palmeri, et al. (1993) extended the experimental procedure to include from 2 to 20 voices and lags from 0 to 64 intervening words. Although Palmeri, et al. replicated Craik and Kirsner's findings, they also found that same-voice repetitions produced better recall than different-voice repetitions, irrespective of gender. Consistent with Craik and Kirsner's finding, the instructions to identify word and voice repetitions rather than simply word repetitions did not enhance the advantage in recognition performance of same-voice repetitions. This advantage persisted across lags and across multiple-speaker environments. If voice and word information were processed and stored independently, the increment in the number of acoustic voice changes between words and/or the instructions requiring the processing of both types of information should have had an impact on same-voice repetitions. Contrary to the

evidence collected with the serial-ordered recall procedure (Goldinger, et al., 1991), Palmeri, et al's findings suggest that the additional processing required by multiple voice changes does not impair the encoding and storage of voice and word information. Procedural differences between serial-ordered recall and continuous recognition procedures may account for this apparent discrepancy. Serial-ordered recall instructions stress the retrieval of word information in a specified arbitrary order without the aid of the acoustic-stimulus words. In continuous recognition memory, instructions emphasize, to a certain extent, the match between a memory representation and an acoustic-stimulus word. The processing demands of serial-ordered recall are costly and, at short lags, voice changes between successive words in the primacy portion of the list to be remembered simply increase the burden. Similar time constraints and processing demands are not present in continuous recognition memory. Thus, in this latter procedure, the effects of the additional processing of voice information do not emerge.

Explicit Memory for Voices: The evidence. The examination of long-term memory for voices has generally entailed the recognition of a voice or voices previously heard within a set of unfamiliar voices (Carterette & Barnebey, 1975; Papcun, Kreiman, & Davis, 1989; Thompson, 1985). The delay between the prior exposure to a voice and the recognition test has ranged from 1 day to 5 months.

In general, participants are found to be quite reliable in explicit recognition of unfamiliar voices after short delays that range from approximately one day to one week. Reliability drastically decreases at longer delays (McGehee, 1937). After short delays, however, the ability to recognize a novel voice depends upon the awareness that a

recognition test for that voice will follow (Saslove & Yarney, 1980). Recognition accuracy is also affected by the number of voices and the duration of the speech samples available to the listener (Legge, Grosman, Pieper, 1984; Pollack, Pickett, & Sumby, 1954). Two additional factors have been shown to be relevant to voice recognition accuracy: language and voice familiarity. When the utterances produced by a speaker are in the language of the listener, voice recognition accuracy is higher than when the same utterances are produced in a foreign language (Goggin, Thompson, Strube, & Simental, 1991; Goldstein, Knight, Baisis, & Conover, 1981). Accented voices are recognized less accurately than non-accented voices especially when brief speech samples are available (Goldstein, Knight, Bailis, & Conover, 1981). Voice familiarity aids voice recognition (Bricker & Pruzansky, 1966; Ladefoged & Ladefoged, 1980; Van Lancker, Kreiman, & Emmorey, 1985; Van Lancker, Kreiman, & Wickens, 1985). This finding cannot be related to differential hemispheric processing. In fact, in dichotic listening, both familiar and unfamiliar voices are processed by both hemispheres (Kreiman and Van Lancker, 1988; Bartholomeus, 1974a, 1974b). The ability of voice familiarity to aid not only voice recognition but also the perceptual processes involved in word identification (Nygaard, et al., 1994), along with the findings of no hemispheric lateralization for voice, support the conclusion that voice and word information are processed as integral perceptual dimensions. The voice-word compounds result in the episodic traces encountered in studies of memory for spoken words.

The Content of Memory for Voices. Although it has been established that speech perception entails the processing of voice characteristics and that memory representations

preserve the result of this processing, two alternative hypotheses have been formulated to predict which voice characteristics are retained in long-term memory. Palmeri, et al. (1993) argue for a detailed representation of a speaker's voice. Geiselman and Bellezza (1976, 1977) and Geiselman and Crawley (1983) assume that only a "summary" representation of the primary properties of that voice is retained in memory. In this latter case, gender connotation would be one of these primary properties.

Gender connotation refers to those properties of the acoustic stimulus that differ between men and women, such as fundamental frequency (the average female f_0 is 1.7 times that of the average male), breathy voice quality, dynamic intonation contour, and gender-specific dialects (Kahn, 1975; Cooper & Sorenson, 1981; Peterson and Barney, 1952; Thorne, Kramearae, Henley, 1983). It has been proposed that the speaker's gender is an important source of information for decoding spoken words (Geiselman, 1979; Geiselman & Bellezza, 1976, 1977; Geiselman & Crawley, 1983). It modifies the semantic interpretation of the message to be conveyed. According to this proposal, labeled the "voice connotation hypothesis" by Geiselman, voice information is automatically encoded but not automatically retained in long-term memory. "Automatic" here refers to unintentional, strategy-free processes. If a speaker's voice constitutes an additional source of information regarding the meaning of a message, then it will be remembered. Alternatively, if it does not, or if it conflicts with the message, the storage of a speaker's voice in long-term memory will be inhibited. The "voice connotation hypothesis" implies that voice gender-related information is encoded as part of the semantic rather than phonological representation of spoken utterances. Investigations that have examined this

hypothesis (Geiselman, 1979; Geiselman & Bellezza, 1976, 1977; Geiselman and Crawley, 1983) employed sentences rather than words as stimulus material, and free recall or recognition as memory procedure. In Geiselman and Bellezza's study (1977), participants heard a set of sentences and were asked to remember either the sentences they had heard (incidental voice-encoding) or the sentences plus the speaker's voice (intentional voice-encoding). Sentences consisting of an agent, an action verb and a complement, could be either neutral as to the gender of the agent, or not (e.g., citizen versus actress). The "voice connotation hypothesis" was supported by recall performance. Retention of the speaker's voice was detected in neutral sentences, but not in sentences in which the gender of the agent was specified, and it was largely independent of the encoding instructions. Geiselman (1979) found no retention of the speaker's voice when it conflicted with the gender of the agent.

Palmeri, et al. (1993) rejected Geiselman's "voice connotation hypothesis" by showing that the repetition of a word in a different voice irrespective of the speaker's gender weakens the memory of that word. A critical analysis of Geiselman's procedure solves the apparent contradiction between these two sets of results. Although Geiselman (1979) has not denied that voice is automatically encoded with word information, he has argued that the long-term retention of voice is under the listener's strategic control. The listener may reject or retain this additional information on the basis of its usefulness in understanding the message. Palmeri, et al.'s evidence (1993), however, suggests an alternative interpretation of Geiselman's findings. Rather than assuming that the switch from a temporary to a long-term representation of a speaker's voice is under a listener's

strategic control, it is more parsimonious to hypothesize that long-term memory of speakers' voices is utilized strategically in an experimental setting in which sentences rather than words are employed. In fact, the meaningfulness of sentential context as opposed to isolated words, in addition to the emphasis that task instructions place on this context, may lead to maximizing the use of semantic information at the expense of voice information, which is primarily pre-semantic in nature.

If a detailed representation of voice characteristics is encoded and retained in long-term memory, it remains to be understood which perceptual characteristics of a speaker's voice comprise the memory of that voice. Bradlow and Nygaard (1996) found that, in a word recognition procedure, recognition was facilitated by repetitions in the same voice and the same rate of speech. No effect of amplitude variation was observed. Similarly, Church and Schacter (1994) found that in implicit memory procedures, changes in voice, intonation, or simply fundamental frequency between a word and its repetition, have a detrimental effect on performance. Again, no amplitude-related change effects were observed. Taken together, these findings suggest that factors such as pitch, speaking rate and intonation comprise the auditory voice component of the memory representations of spoken words.

Investigations of voice recognition, however, have identified several, but not necessarily overlapping characteristics that are employed to differentiate voices. Voiers (1964) identified clarity, roughness, magnitude, and animation. Holmgren (1967) selected intensity, quality, pitch, and rate. Clarke and Becker (1969) listed pitch, pitch variability, rate, clickiness, breathiness and sibilance. Carterette and Barnebey (1975) found that

intensity, pitch and intonation pattern are all employed to define a speaker's voice and suggested that these attributes may be stored in long-term memory along with word information. It is unclear whether all the attributes obtained in voice recognition studies are indeed employed in on-line speech perception and become part of the long-term memory record of spoken words. When perceiving speech, listeners' attention is devoted to phonemic information rather than to voice. In voice recognition, on the other hand, the opposite pattern of allocation of attentional resources is obtained. In addition, participants are encouraged to generalize over different instances to construct a "template" for each voice. Thus, it is reasonable to hypothesize that a more fine-grained processing of voice would be performed in voice recognition studies. This latter assumption does not imply that voice information in on-line speech perception is strategically encoded, it simply suggests that the type of task required by an investigation on voice recognition may compel listeners to add more details to the "mental picture" of a voice, which, in on-line speech perception, would be generated automatically.

As procedural differences may lead to differences in encoding of voice information, the type of stimulus material employed in investigations on word recognition (isolated words versus words embedded in sentences) may dictate which voice characteristics are available for encoding. In fact, sentences provide additional voice-related information, such as prosody and rhythm, which cannot be gathered from isolated words. In addition, sentences may emphasize the encoding of these additional characteristics relative to those embedded in isolated words. In a recent study on the effect of voice familiarity on word recognition, Nygaard and Pisoni (1996) supported this assumption by finding that voice

familiarity acquired from hearing sentences improves identification of words in sentences, but it does not easily transfer to the identification of isolated words. If the encoding of voice characteristics is sensitive to the stimulus material that carries voice information, it remains to be ascertained the extent to which voice-related information gathered from isolated words can be exploited when words embedded in sentences are to be identified.

In summary, the above-mentioned studies have shown that voice characteristics along with abstract-symbolic information comprise the memory representations produced and used in explicit memory procedures. These representations are "episodic" in the sense that they are unique to the experience that has produced them. Voice characteristics determine this uniqueness. The type of stimulus material defines not only which voice characteristics are available but also the extent to which they will be encoded and retained in memory. It is unclear, however, whether the extensive multidimensional encoding of voice information suggested by voice recognition studies applies to on-line speech perception and explicit memory for spoken words. The following section examines these findings in more detail as they apply to implicit memory.

4. Implicit Memory of Spoken Words: The evidence

Study-To-Test Voice Changes. The investigation of implicit memory of spoken words has primarily concentrated on determining whether the memory representations produced by prior exposure to a set of spoken words contain instance-specific voice information rather than merely abstract-symbolic word information.

The standard procedure used in implicit memory investigations generally entails a study phase in which participants are asked to respond to a set of spoken words, and a test

phase in which the same words are repeated generally intermixed with novel words. A study-to-test voice change refers to a word spoken by a speaker at study and repeated at test by another speaker. In the test phase, participants are asked to respond to words or to their fragments without referring to the study phase. The study and test phases are separated by a distractor task. In the following paragraphs, a brief examination of the evidence will be presented. The terms "auditory repetition priming" and "implicit memory" will be used interchangeably to refer to the effects of prior exposure. Repetition priming, however, denotes a behavioral phenomenon. Implicit memory refers to the unobservable memory representations that are assumed to support this behavioral phenomenon.

Schacter and Church (1992) found that study-to-test voice changes reduced auditory repetition priming in a stem completion task. However, when white noise was added to the words of the test phase, performance in both word stem completion and word identification was unaffected by voice changes. In a subsequent series of experiments, Church and Schacter (1994) showed that study-to-test voice changes had a detrimental impact on priming in word identification and stem completion. Contrary to the prior set of experiments, word identification did not entail hearing words in a background noise. Words were simply degraded with a low-pass filter. Similarly, Goldinger (1992) found that in an identification task with words presented in noise, priming was sensitive to study-to-test voice changes. This effect was larger when only two voices were presented. In addition, Goldinger found that the perceptual similarity of speakers' voices, determined through a scaling procedure, had a significant effect on priming. The magnitude of priming decreased as a function of voice dissimilarity. In contrast to Schacter and Church's

findings (1992), noise did not wipe out the effect of voice repetition. Schacter and Church attributed the lack of significant voice effects to the presence of noise. In Goldinger's study, a signal-to-noise ratio of + 5 dB was implemented for fifty msecs after the onset of each word. Although Schacter and Church (1992) did not specify the signal-to-noise ratio and the duration of noise relative to word length, the noise might have been higher and completely overlapped each word. Schacter and Church's hypothesis finds support in Jackson and Norton's (1984) findings of no study-to-test voice changes when noise completely masked each word and lasted until after word offset. An alternative explanation, suggested by Goldinger (1992), focuses on the presence rather than the nature of noise. In Goldinger's word identification procedure, the presence of noise both at study and at test enhanced the perceptual similarity between a stimulus and its repetition. This does not apply to Schacter and Church's study, in which no noise was used in the study phase. Turner and Pisoni (1993-1994) employed a forced-choice continuous identification task of words either degraded or presented in the clear. Response latencies to same-voice repetitions were found to be faster than different-voice repetitions independently of the degradation level (note that in this task accuracy is not sensitive to study-to-test voice changes). This finding supports Goldinger's hypothesis by demonstrating that when a stimulus is equally disrupted or not disrupted across its several repetitions, voice information remains a relevant component of priming.

Sheffert (1995) assessed the effect of noise versus no-noise at study on identification performance at test. An identification task was employed at study with words entirely embedded in noise (the signal-to-noise ratio was of - 5 dB) or presented in

the clear. At test, the same identification task was carried out on studied and non-studied words, either embedded in noise or low-pass filtered. Study-to-test voice effects on identification performance were found only when words were embedded in noise at study. Although, Sheffert's (1995) findings indicate that noise does not have a detrimental effect on the encoding and retrieval of voice information, they do not entirely support Goldinger's "perceptual similarity" hypothesis. In fact, voice effects were found even when noise had embedded words at study and filtering had disrupted the acoustic signal carrying the same words at test. Sheffert's findings, rather, suggest that the increased attention, required by the identification of unclear stimuli, may improve perceptual analysis both at study and at test, and then be responsible for the observed voice effects.

Taken together, the above-mentioned findings lead to two general conclusions regarding implicit memory for spoken words. First, voice information appears to be encoded into long-term memory along with abstract-symbolic information. Second, the retrieval of these episodic memory representations does not require explicit recollection. Whether the encoding of voice information, and thus its contribution to these memory representations, is determined by the instructions provided at study has been investigated in several studies.

Study-To-Test Voice Changes and Levels of Processing Manipulations. Goldinger (1992) manipulated the study task by employing three levels of processing. In this experiment, the term "level" implies that the operations emphasized by the study task can be organized along a continuum defined as a progression from a superficial to a deep level of processing (Lockhart, Craik, Jacoby, 1975). Depth is defined on the basis of the extent to which an

operation relies on information regarding a word's meaning and its syntactic properties rather than information more directly gathered from the auditory and phonological form of the word. In Goldinger's experiment, participants classified words on the basis of the speaker's gender, word-initial phoneme, or syntactic category at study. Goldinger hypothesized that if intentionality is not a necessary condition for the encoding of voice, attention to word information, such as the phonemic and syntactic properties of words rather than voice information, should not affect priming. Indeed, Goldinger found that same-voice repetitions were faster and more accurate than different-speaker repetitions at all levels of processing. Although this difference was stable across levels of processing, priming scores increased as a function of deeper levels of processing.

Schacter and Church (1992) manipulated the study task by using either a semantic (meaning rating) or a pre-semantic (clarity rating) task. Word-stem completion performance was employed to measure priming effects. The effect of study-to-test voice changes on priming was irrespective of the study task. However, when the manipulation of the study task entailed either judgments of the pleasantness of a word (semantic task) or pitch ratings (pre-semantic task), priming effects were larger following pleasantness rating than pitch rating, and study-to-test voice effects were observed only following pleasantness rating. The fact that more priming was also observed after pleasantness rating suggests that the latter finding may be an artifact due to the magnitude of the priming effects. Sheffert (1995) also manipulated the study task by employing a semantic (meaning rating) and pre-semantic task (clarity rating). At test, participants identified low-pass filtered words or words in noise. Although Sheffert found that priming was insensitive to

the task used at study, no significant study-to-test voice changes were obtained. The fact that words were presented in the clear at study and were degraded at test explains these latter findings, which thus do not shed light on the issue at hand.

Lively (1994) did not find study-to-test voice change effects on priming in a lexical decision task employed both at study and at test. He found, however, more robust priming for low frequency words than high frequency words and no priming with pseudowords. In another experiment, Lively (1994) utilized a lexical decision procedure at study and a gating task at test. In gating (Grosjean, 1980), participants hear progressively extensive portions of a spoken word until the whole stimulus is presented. The task is to identify the word as early as possible. Although Lively found that repeated words were recognized earlier in time than new words, no study-to-test voice changes were observed. When the study task was changed to word identification, the same outcome was obtained. A word frequency effect, however, was observed. When word identification was employed at study followed by stem completion at test, priming without study-to-test voice change effects was also found.

Not all the above-reported findings suggest that implicit memory of spoken words entails the encoding of voice information irrespective of the type of processing encouraged by the instructions at study. Two independent investigations support the "automatic encoding" hypothesis. Goldinger (1992), by implementing three different levels of processing both at study and at test, found stable study-to-test voice changes across the three levels of processing. Schacter and Church (1992) obtained the same results with meaning and clarity rating tasks at study and a word-stem completion at test. Sheffert's

(1995) and Lively's (1994) findings of no effect of study-to-test voice changes on priming do not support this hypothesis. Procedural differences explain these discrepancies. Sheffert's null results may be due to the presence of noise in the test phase (see above discussion on the effects of noise). The evidence collected by Lively (1994) remains an anomaly that does not challenge the postulation of episodic voice-specific implicit memories, but rather questions either the techniques employed at test or the stimulus material used at study. One may argue that lexical decision entails deep levels of processing that render the task insensitive to voice information at test. It is difficult, however, to attribute the same lack of sensitivity to gating and word stem completion. An examination of Lively's investigation highlights an interesting procedural detail that can explain the lack of study-to-test voice effect. At study, each word was repeated by seven different speakers. In Goldinger's (1992) and Schacter and Church's (1992) investigations each word was spoken by one talker at study. Multiple repetitions by different speakers in Lively's investigation may have encouraged the generation of a "template" for each voice generalized over the different instances in which that voice was heard. This generalization process may have then de-emphasized the salience of voice information in the instance-specific word representations employed at test or simply depleted these representations of the voice characteristics unique to each speaker. In addition, gating, word identification and lexical decision are procedures that enhance the processing of stimuli as whole words and this enhancement may also lead to a decrement in the retrieval of voice information at test. The absence of study-to-test voice changes in naming (Brown, Fowler, & Rueckl, 1996) may be attributed to the combined effect of the word-

level response and motor-vocal response required by the task. In fact, a motor-vocal response may render the participants' voices more salient than the speaker's voices and thus override the speaker-specific voice characteristics of the stimulus material.

In summary, despite procedural artifacts and idiosyncrasies, the evidence suggests that implicit memory for spoken words entails the encoding, retention and implicit use of voice information. There is not sufficient evidence to support the notion that depth of processing, originally employed to explain improved recollection in explicit memory, leads to a differential encoding of voice information. The extent to which the processes involved with the encoding, retention, and retrieval of voice information are under the listener's control is examined in the following section in more details.

5. The Processing of Voice Information and Automaticity

The term "automaticity" in cognitive psychology refers to processes that, initiated unintentionally, are performed without awareness, do not deplete limited attentional resources and are autonomous in the sense that they cannot be interrupted or modified before completion (Shiffrin & Schneider, 1977). In general, automatic processes develop through extensive training and operate on well-known rather than novel stimuli. Posner and Rogers (1978) indicated that the notion of automatic processing implies an invariant relationship between a stimulus and a set of internal representations.

This monolithic interpretation of automaticity implies an all-or-none criterion, so that if a process does not possess all the defining characteristics of automaticity it is not automatic, but rather controlled. Bargh (1992) has pointed out that the defining characteristics of automaticity "do not hang together in an all-or-none fashion, but rather

seem to be able to co-occur in just about any combination" (Bargh, 1992, p. 183). Thus, the encoding of voice is automatic in the sense that it does not entail intentionality to be initiated and awareness to be executed (Geiselman & Bellezza, 1977). The encoding and retention of a speaker's voice, however, require attentional resources and thus it is not fully automatic. Several findings support this conclusion. Mullenix, et al. (1989) and Mullenix and Pisoni (1990) showed that in speech perception, the encoding of word information is affected by a multiple-speaker acoustic environment. Martin, et al. (1989) found that serial recall of the primacy portion of a multiple-speaker word list was less accurate than in a single-speaker list. The recall of digits visually presented prior to a multiple-speaker word list was also less accurate than the recall of digits followed by a single-speaker list. Goldinger, et al. (1991) found that the serial recall of a multiple-speaker list was affected by presentation rate. At short ISIs, recall was more accurate in the early portions of a single-speaker list. Goldinger (1992) found that the switch from two to six or ten voices produced a decrement in auditory repetition priming. Therefore, the detrimental effects of speaker variability are not limited to speech perception and explicit memory, but they involve the generation of implicit memory representations as well.

In summary, this pattern of results suggests an increment in processing demands in multiple-speaker acoustic environments. With the exception of Palmeri, et al.'s findings (1993), the evidence collected in speech perception and memory investigations suggests that the encoding and retention of several voices deplete the attentional resources necessary for the generation of long-term memory representations of both voice and word

information. Goldinger (1992) attributed the effect of speaker variability on memory to a diminished encoding of voice. This reduction of attentional resources results in memory representations that are not strong enough to be efficiently used in either explicit recollection or implicit performance. The impairment in recollection or in the implicit use of these representations can be thought of as a byproduct of the early impairment. Whether the use of these representations is also affected by a multiple-speaker environment at test is, at the present time, unclear.

Prolonged practice in several tasks, however, makes information processing in those tasks effortless (Schneider, Dumais, Shiffrin, 1984; Schneider, Shiffrin, 1977; Shiffrin & Schneider, 1977). The fact that voice familiarity improves word identification (Nygaard, et al., 1994), suggests that extensive exposure to a specific voice renders that voice a familiar stimulus and its encoding less attention demanding. Lightfoot's finding (1989) that voice familiarity improves recall of multiple-speaker lists, indicates that the impairment in recall of multiple-speaker lists is primarily due to the processing costs associated with the encoding of voice information. Although no evidence is available at the present time, the same effect of voice familiarity should be observed in implicit memory. Lightfoot's finding, however, was obtained through explicit encoding of voice (e.g., participants were asked to associate voices with fictitious names). The hypothesis that the beneficial effect of voice familiarity on memory could be easily replicated without explicit encoding of voice information appears to be plausible.

The issue of automatic processing bears on the classical distinction between automatic activation of pathways versus effortful, conscious attention to pathways (Posner

& Rogers, 1978). The concept of "pathway" implies an invariant relationship between a stimulus and a set of internal representations. Effortful processing can be conceptualized as "processing interference" (Miller, Kasprow, Schachtman, 1986). In the encoding and retention of voice, effortful performance can be linked to "processing interference", which arises from competition for attentional resources. The decrement in priming as a function of number of voices supports not only the idea of effortful encoding and retention, but also the proposal that repeated exposure to a speaker's voice may reduce "processing interference". Indeed, as the number of voices in the acoustic environment increases, exposure to a given voice decreases and thus does familiarity. Effortless performance would be expected to emerge in a restricted acoustic environment in which voice familiarity can be rapidly achieved. With increments in speaker variability, familiarity decreases. Attentional resources are required and thus processing interference is observed (Mullenix & Pisoni, 1990).

Whether the encoding and retention of voice are subjected to individual processing strategies was directly tested by Palmeri, et al. (1993). One of the properties of automaticity is lack of conscious control over a process. This means that irrespective of the listeners' intentions, an autonomous process cannot be interrupted or modified before completion. Palmeri, et al. (1993) utilized a continuous-recognition memory procedure with instructions to base recognition judgements on either word information alone or both word and voice information. Words were repeated either in the same voice or in a different voice. They hypothesized that if the encoding of voice and its long-term retention are controlled by listeners' strategies, instructions to respond to word and voice information

as opposed to word alone would have facilitated recognition performance of same-voice repetition and increased the gap between same-voice and different-voice repetitions. A consistent same-voice repetition advantage, irrespective of instructions, was observed. They also hypothesized that if the encoding and retention of voice are not automatic, an increment in acoustic variability (from 2 to 20 speakers) would impair word recognition. No significant effect of acoustic variability was observed. According to Palmeri, et al., these two findings suggest that the encoding and long-term retention of a speaker's voice are not performed strategically. This conclusion does not deny the presence of "processing interference" relative to the concurrent processing of voice and phonemic information. In Palmeri, et al.'s study (1993), presentation rate was slow so that the effect of the additional processing load due to the multiple-speaker environment was less likely to surface. Indeed, participants were allowed a maximum of 5 secs to produce a response and a 1-sec delay separated a response from the onset of the subsequent stimulus.

The discussion of automaticity as it refers to the encoding and retention of voice-related characteristics in the previous paragraphs, leads one to question the notion of automatic use of these characteristics in memory tasks. In fact, it is reasonable to assume that, at test, priming stems not only from speech perception and retention processes, but also from the processes involved in the use of long-term memory representations of voice and word information. According to Posner and Rogers (1978), automatic activation would be expected to facilitate the processing of a stimulus and of stimuli that are similar in important ways to that stimulus. On the other hand, effortful activation would entail inhibition of similar but not identical stimuli. An example of effortful activation in

priming is found in Ratcliff, McKoon, and Verwoerd's study (1989) in which the identification of a visually presented word is facilitated by prior exposure to the same word and inhibited by prior exposure to a similar but not identical word. In implicit memory for spoken words, the idea of automatic use of voice information can be tested without the contamination of conscious retrieval processes. If voice information is employed in performance in an automatic manner, the detrimental effect on priming of a voice change between study and test should be positively related to voice dissimilarity (e.g., the more different the voice used at test is from the voice used at study, the larger should be the decrement in priming). On the other hand, the detrimental effect of study-to-test voice changes should be inversely related to voice dissimilarity if effortful processing is performed. It follows that the more similar the two voices are, the more interference should emerge. Goldinger's findings (1992) support the former pattern of results rather than the latter. Although the encoding of multiple voices may require attentional demands and produce interference (Mullenix & Pisoni, 1990), at test, this interference is probably attenuated by the presence of memory representations for the incoming stimuli. In this context, the use of preexisting lexical representations containing voice may not require attentional resources and thus it may not entail inhibitory processes. In other contexts using sentences and explicit memory tests, however, voice information encoded in memory representations may be subjected to strategic use (Geiselman, 1979). Whether or not strategies such as hypothesis generation and expectations may affect the retrieval of implicit memory representations at test, however, is still an open question.

6. Detailed Implicit Memory Representations: Evidence from the Visual Domain

The evidence presented in the previous sections suggests that the processes of encoding and retention of spoken words produce episodic memory representations. These representations are detailed in the sense that they contain information regarding the speaker's voice as well as abstract-symbolic representations. Whether the same pattern applies to the visual domain is discussed in the following paragraphs.

Letters. The well-known series of experiments conducted by Posner (Posner, Boies, Eichelman, & Taylor, 1969; Posner & Keele, 1967; Posner & Mitchell, 1967) on perceptual matching of letters (AA versus Aa) showed that the physical characteristics of visually presented material affect name-match decision times only when both letters are simultaneously presented. When an increasing interval is inserted between the display of the first letter and the appearance of the second letter of the pair, the advantage of a physical match progressively disappears. These findings, which suggest that surface characteristics of written material are temporarily encoded, but lost within a couple of seconds, would lead one to think that implicit memory of visual stimuli does not entail detailed representations as in the auditory domain. While in both domains superficial changes in repeated patterns affect immediate processing, unnecessary details in visual processing appear to be discarded in favor of more abstract and stable symbolic representations.

Sanocki (1987, 1988) suggests that this assumption may be incorrect. He argues that the perception of visual characters entails the encoding of both abstract information, such as letters, and surface characteristics, such as the font in which these letters are typed. This encoding allows perceptual processes to become "tuned to the details" of a font

(Sanocki, 1987, p. 267) so that reading can improve in speed and accuracy. Mullenix and Pisoni (1990) have shown that in speech perception, a multiple-speaker environment has a detrimental effect on the activation of abstract-symbolic information. Sanocki (1987) demonstrated a similar effect for font variability in reading by presenting strings of four characters in either the same font (regular condition) or two different fonts (mixed condition). A string could contain either four letters or three letters and a foil (e.g., a non-letter), and participants were required to indicate which of the two types of strings was presented on each trial. Results showed that mixed-font strings have a negative effect on letter perception. This effect depends on the magnitude of the adjustment that perceptual processes make to adapt to different fonts. The more different two fonts are in the mixed-font strings, the longer is the processing time. In another experiment, Sanocki (1988) employed strings of four letters followed by a two-alternative forced-choice identification task. Sanocki (1988) found the same pattern of results in accuracy of identification. In addition, he showed that perceptual adjustments are not confined to conditions in which fonts change within a string, but they also occur when fonts change between strings.

Although Sanocki (1987, 1988) has argued that the perceptual processes involved in letter perception entail the encoding surface characteristics, he has neither argued for nor tested the hypothesis, supported in the auditory domain, that instance-specific episodic representations are constructed for each stimulus. This hypothesis is examined in the following paragraphs where the findings of long-term retention of surface characteristics in reading words in text are examined

Words. The evidence of implicit memory for surface characteristics has been gathered by

using study-to-test changes in type case, type font, orientation of text and handwriting. This evidence is contradictory.

Several experiments have changed text orientation as to obtain inverted, upside-down or backwards text. Kolers and Ostry (1974) tested the effect of inverted sentences (180°) on reading times. At study, participants read normally oriented or inverted sentences and, after delays ranging from 3 minutes to 32 days, read the same sentences intermixed with others. The orientation of half of the original sentences, however, was changed. Although, at test, reading times of sentences normally oriented was largely independent of their orientation at study, reading times of inverted sentences was found, at short delays, to be facilitated by a physical match. Horton (1985) and Tardif and Craik (1989), however, did not observe a similar pattern. Graf and Ryan (1990) tested the effect of study-to-test orientation changes on visual word identification. Rather than simply inverting text, they presented sentences upside-down. No significant effect was observed. Graf and Ryan (1990), however, found an effect of backwards orientation.

Other experiments have changed the type case of words from study to test. Scarborough, Cortese, and Scarborough (1977) tested the effect of study-to-test changes of type case in lexical decisions (lowercase versus uppercase). No significant effects were obtained except in one experiment. Similarly, Levy and Kirsner (1989) did not find an advantage of same-form repetition on both word identification and reading times. Jacoby and Witherspoon (1982) and Jacoby and Hayman (1987) found that same-form repetition facilitated word identification performance. The effect was limited, however, to words presented in lowercase letters both at study and at test as opposed to words studied in

uppercase letters and tested in lowercase. Woltz (1990) observed a facilitatory effect for same-form repetition in a semantic task. Participants, required to determine whether words in a pair shared the same meaning, were faster, but not more accurate, when pairs were repeated in the same type case.

Additional experiments have changed the type font. Levy and Kirsner (1989) found that reading times of repeated passages were not improved by same-font print. Levy (1983) and Levy, DiPersio, and Hollingshead (1992) asked participants to read a passage four times, and then read the passage again in either the same or a different type font. Levy (1983) tested type font changes in error detection and found that participants, exposed to passages repeated in the same font, were able to detect more errors than when exposed to the same passages printed in a different type font. Similarly, Levy, et al. (1992) found that both accuracy and time in error detection of repeated passages were facilitated by same-font text. Graf and Ryan (1990) used unfamiliar type fonts and found that the study task determined whether information regarding surface features, such as type fonts, would be encoded. Perceptual identification was facilitated by same-font words when participants were asked to rate the readability as opposed to the pleasantness of study words. Graf and Ryan (1990) argued that the readability task enhanced the encoding of surface information, which is unlikely to be relevant in a semantic task such as pleasantness rating. Brown and Carr (1993) used familiar type fonts and did not find an effect of study-to-test font changes in lexical decision and naming. They argued that type-font familiarity determined whether surface features of words would be encoded. Usual fonts have previously been encoded and thus are unlikely to be recorded again. This argument does not easily apply to

experiments in which the study-to-test change entailed typewritten versus handwritten words. Clarke and Morton (1983) did not find a facilitation in tachistoscopic thresholds for typewritten words typed at study as opposed to handwritten. Similarly, Levy and Kirsner (1989) found no effect of study condition entailing handwritten versus typewritten words on word identification and reading times of typed passages.

The conflicting results of study-to-test changes in the surface features of repeated words compared with the clear pattern of results of font changes in letter identification (Sanocki, 1987, 1988) suggest that the role of surface information in reading and speech perception may differ not in terms of encoding and retention but in terms of retrieval processes. In reading, the processing of visual features, such as type font or type case, appears to be necessary for the activation of representations entailing letters (Posner & Keele, 1967; Posner & Mitchell, 1967; Sanocki, 1987, 1988), as the processing of voice characteristics appears to be for the perception of abstract-symbolic representations (Goldinger, et al., 1991). Voice characteristics, however, frequently provide additional useful information in decoding a message (Geiselman, 1979; Geiselman & Bellezza, 1976, 1977; Geiselman & Crawley, 1983). The specific form of these letters is generally irrelevant to the act of decoding a linguistic message. Graf and Ryan's finding (1990) supports this conclusion by showing that attention to visual features during study allows for the encoding of surface characteristics such as type fonts. Sanocki's evidence (1987, 1988) demonstrates that this finding cannot be interpreted as suggesting that surface features are strategically encoded in reading. Surface information may be simply strategically retrieved. According to Carr, Brown, and Charalambous (1989), visual

features experienced in daily reading, such as type font, type case and handwriting, are irrelevant variations that are unlikely to be automatically encoded in long-term memory. However, because reading text entails gathering the invariant characteristics of written symbols, it is reasonable to hypothesize that surface information, such as type fonts, may be encoded but not employed in on-line reading. Alternatively, the tasks employed to detect study-to-test surface changes, such as reading words in a passage, may not be sensitive enough for detecting the retention of surface information.

7. Memory System Versus Processing Conceptualizations of Implicit Memory

The models. In the previous sections, experimental findings have been the primary focus. This section will address two alternative conceptions of implicit memory formulated in the literature. As the main interest is implicit memory for spoken words, explicit reference to this domain will be made whenever possible.

Numerous investigations have favored the *transfer-appropriate processing* model, which postulates that auditory repetition priming results from the processing operations that overlap between study and test. As different processes are reflected in different tests, priming is seen as an adaptable rather than a rigid behavioral phenomenon (Roediger, 1990). In contrast, clinical neuropsychological investigations have preferred the *multiple-memory system* model, which portrays auditory repetition priming as rigidly dependent upon a cortically-based memory system (Squire and Knowlton, 1995). According to Tulving (1985), a memory system can be thought of as an organized structure, which can be neurologically and functionally separated from other structures. It entails a set of processes that characterize that system as distinct from others. Since different tests display

different processing requirements, the ability of a memory test to engage that system will result in a more or less uncontaminated form of auditory repetition priming. The *transfer-appropriate processing* model fits either an episodic or an abstractionist conceptualization of implicit memory. Similarly, it can accommodate either a modification or an acquisition account of implicit memory phenomena. The *multiple-memory system* model is clearly episodic. Embedded in this model is the idea of acquisition as a descriptive term for the processes of memory formation.

The *transfer-appropriate processing* model stems from the original "levels of processing" conceptualization of explicit memory (Craik & Lockhart, 1975), which assumed that a stimulus could be processed at different levels of analysis. These levels were thought of as sequentially ordered from superficial to semantic. Moreover, they were assumed to be governed by task instructions. "Depth" of encoding referred to how close an analysis of a stimulus was to the semantic level. Thus, if task instructions required that a response be given to the type case in which a word was printed, a superficial level of analysis was assumed to have been performed. On the other hand, if a response to the word's meaning was required, a deeper level of analysis was postulated. A subsequent modification of the idea of "depth" of encoding led to the notion of "domains" of encoding (Lockhart, Craik, & Jacoby, 1975), which deprived the original conceptualization of its rigid serial connotation in favor of a more flexible account of how instructions guide encoding processes. This model accounted for the evidence that showed that instructions to process a word's meaning led to a minimal rather than thorough processing of the superficial characteristics of that word. Only those operations necessary to gather sufficient

information for the activation of the word's meaning were carried out. As a consequence of this modification, at the present time, "depth" or "levels" of encoding are primarily employed to index the type of information upon which task instructions focus participants' attention.

The *transfer-appropriate processing* model (see Roediger, Srinivas, & Weldon, 1989) applied to auditory repetition priming leads to the assumption that, at study, task requirements determine how thoroughly each speech perception operation is performed, and thus define the relative strength of the memory representations that these operations produce. For example, lexical decision, which requires participants to distinguish between a word and a similar sounding pseudoword, promotes a minimal pre-lexical analysis of a spoken word (e.g., only those operations necessary to provide evidence for subsequent lexical processes) and emphasizes lexical analysis. Task requirements at test simply define the sensitivity of the procedure to the memory representations previously developed. Consequently, the degree of similarity between the operations executed at study and those performed at test is crucial in determining which memory representations contribute to auditory repetition priming (Roediger & Blaxton, 1987). For example, if a lexical decision task is utilized both at study and at test, auditory priming is assumed to be based on lexical representations and not to involve memory of the speaker's voice (Lively, 1994). The *transfer-appropriate processing* model has received support primarily from studies in the visual domain (see Lewandowsky, Dunn, & Kirsner, 1989).

The *multiple-memory system* model, which has primarily found support in the auditory domain, claims that task-directed speech perception operations, such as

performing a semantic task or rating the pitch of the voices that produce the study words, do not affect the strength and quality of the memory representations developed in the study phase. Schacter and Church (1992) claim that implicit memory of spoken words results from the operations of a cortically-based memory system exclusively devoted to pattern-recognition processing. Because pattern-recognition operations are pre-lexical in nature (Pisoni, et al., 1985), this system is assumed to process voice characteristics and phonological information at the pre-lexical (e.g., subword) rather than lexical (e.g., word) level. At test, emphasis is placed on how task-directed speech perception operations affect the ability of the memory procedure to apprehend the representations developed during the study phase. For example, although word identification and stem completion require lexical processing, these procedures are also known to emphasize pre-lexical processing, and thus they are appropriate for capturing the memory representations upon which auditory repetition priming relies (Church & Schacter, 1992; Goldinger, 1992; Church & Schacter, 1994). Lexical decision, on the other hand, is not (Lively, 1994). Therefore, a voice change between study and test, such as the word "boat" spoken by a female voice versus a male voice, is expected to have a detrimental impact on auditory repetition priming in stem completion and word identification, but not in lexical decision where emphasis is placed entirely on lexical processing.

The *multiple-memory system* model agrees with the *transfer-appropriate processing* model that implicit memory procedures can be placed along a continuum defined by the extent to which each procedure relies on data-driven rather than semantically-driven operations. This notion originates from the idea of "depth" of processing (Lockhart, Craik,

& Jacoby, 1975). Data-driven tests are assumed to enhance pre-lexical processes and thus are expected to be primarily sensitive to surface characteristics such as voice in the auditory domain and fonts in the visual domain. Examples of data-driven procedures are word identification and stem completion. Semantically-driven tests are assumed to depend primarily on the meaning of a stimulus and thus are less likely to be sensitive to a stimulus' surface characteristics. Lexical decision is an example of a semantically-driven procedure. The *multiple-memory system* model postulates that implicit memory in the auditory domain must be investigated through data-driven procedures. The *transfer-appropriate processing* model allows for both procedures to be utilized. Both models, however, assume that the sensitivity of the memory test to the different "components" of the memory representations constructed in the study phase will determine which component will surface at test. Note that the *transfer-appropriate processing* model defines sensitivity in terms of processes that overlap between study and test, but the overall idea is fundamentally the same.

Clinical and neuropsychological evidence appears to support the *multiple-memory system* model. Patients with global amnesia have been found to perform poorly on tasks that require explicit recollection (Scoville & Milner, 1957). They perform normally, however, on implicit tasks (Warrington & Weiskrantz, 1970). Neuroanatomical investigations have shown that damage of the medial temporal lobe (i.e., hippocampus, entorhinal, parahippocampal and perirhinal cortices), basal forebrain, and diencephalon does not affect the formation and use of implicit memories (Squire, 1992; Squire & Zola-Morgan, 1991). Rather, implicit memory of visual and auditory stimuli in data-driven tests

relies on the occipital and posterior temporal cortices, where the recognition of visual and auditory patterns is assumed to be performed. The *multiple-memory system* model, however, encounters the problem of defining exactly the functional properties of a memory system as distinct from those of other systems. Brain imaging techniques, such as positron emission tomography (PET) and magnetic resonance imaging (MIR), provide a measure of neuronal activation during task performance. Evidence of neuronal activity, which per se is not sufficiently informative, combined with measures of observable performance, has permitted the identification of a considerable amount of region-specific activity in the brain following certain task operations (Gazzaniga, 1989; Posner, 1993). As different tasks engage several dissimilar operations, however, it is difficult to isolate the processes of one system from the operations of other systems. Thus, the test of the *multiple-memory system* model must rely upon the arduous task of finding memory procedures that are unlikely to engage other systems.

The criteria of functional and stochastic independence have been employed to isolate memory systems in conjunction with the search for suitable memory procedures. As these criteria have been primarily employed to distinguish between implicit and explicit memory systems, the present discussion will refer to the obstacles that have emerged from this undertaking.

Functional independence is inferred when an independent variable, such as study-to-test voice changes, has a differential effect on two measures of performance, such as implicit versus explicit. Schacter and Church (1992) demonstrated functional independence by reporting an effect of study-to-test voice changes in a stem completion

task, but not in a cued recall task. Lively (1994), however, found the opposite outcome in a lexical decision task as opposed to a recognition task. Similarly, Turner and Pisoni (1993-1994) found study-to-test voice change effects on accuracy in a recognition task but not in a word identification task (note that word identification latencies were sensitive to study-to-test voice changes). Thus, evidence of functional independence is questionable in several ways. Of primary importance is the underlying assumption that there is a one-to-one mapping between a task and a process or processes. Jacoby (1991), however, has reminded researchers that tasks are unlikely to be "process-pure". Even if the findings are homogenous across studies (note that this is not the case here), functional independence can be interpreted as due to the different processes engaged by the two memory tasks rather than as dependent upon a separable memory system. If the findings are heterogenous, functional independence leads to contradictory conclusions regarding memory systems.

Stochastic independence has been described as a more compelling criterion than functional independence. Stochastic independence means that the operations entailed by the two memory tasks do not overlap. On the basis of the assumption that each task taps into a different memory system, stochastic independence signifies separate memory systems. The validity of this assumption has been questioned by acknowledging that tasks are not process-pure (Jacoby, 1991). Nevertheless, in the visual, but not in the auditory domain, researchers have attempted to gather evidence showing that the effect of prior exposure is stochastically independent from the combined effect of prior exposure and conscious recollection. For example, Tulving, Schacter, and Stark (1982) have shown that

at test, the repetition of a stimulus word enhances implicit memory performance in a word fragment completion test. Such an enhancement is independent of recognition performance. In fact, facilitation following repetition was the same for remembered and not remembered stimuli. However, since the recognition test was administered before the word fragment completion test, the recognition test provided a second study phase in which new words had the opportunity to become, to a certain extent, old. The finding of stochastic independence is thus not a surprise.

In spite of the obstacles in isolating explicit and implicit memory systems, the claims of the *multiple-memory system* model for auditory repetition priming rely on a testable hypothesis. The hypothesis of a cortically-based memory system devoted to the processing of predetermined characteristics of auditory stimuli conforms to Fodor's conceptualization of a module (1983), which is a domain-specific, informationally-encapsulated system. Domain specificity refers to the fact that the module deals only with one type of information. Informational encapsulation implies that the functioning of the system cannot be modified by strategic processing. A module is thus an independent processing system. Whether it possesses its own unique set of acquisition, retention, and retrieval processes is a matter of debate (see Sherry & Schacter, 1987). Schacter and Church (1992) state that this seems to be the case in the auditory domain. According to this conceptualization, implicit memory for spoken words can be depicted as entailing a pattern-recognition module that captures only the superficial aspects of the information produced by speech perception and transcribes them into an episodic trace. On the basis of this model, during study, the manipulation of speech perception operations does not

modify the functioning of this module. The memory test, however, is instrumental in capturing the memorial episodes obtained from the study session. Consistent with the idea of a module, repetition priming is modality specific. Repetition priming does not easily transfer to another modality (Bassili, Smith, MacLeod, 1989; Gipson, 1986; McClelland & Pring, 1991; Tulving & Schacter, 1990). Contrary to this conceptualization of implicit memory, the *transfer-appropriate processing* model fits the idea of information processing as an array of operations governed by task demands and thus to a certain extent by an individual's conscious and strategically determined operations (Graf & Ryan, 1990).

The Manipulation of the Encoding Task as a Test for These Models. The crucial test for the *multiple-memory system* and the *transfer-appropriate processing* models stems from the effects of encoding manipulations on priming. For the purpose of the present investigation, the evidence from the auditory domain will be discussed. The manipulation of the encoding condition is usually performed through task instructions under the assumption that task-directed operations modify the manner in which the stimulus material is encoded. Generally, two general categories of tasks have been employed: semantic versus pre-semantic. The effects of this manipulation have been measured on the magnitude of priming and its sensitivity to study-to-test voice changes. Note that the sensitivity of priming to study-to-test voice changes is an index of the functioning of the above-mentioned module.

In favor of the *multiple-memory system* model, Schacter and Church (1992) showed that, in data-driven implicit memory tasks, auditory priming is affected minimally or not at all by encoding manipulations. In addition, changes in voice between study and test have

a detrimental effect on priming. Sheffert (1995) also found no effect of encoding manipulations. Goldinger (1992) manipulated both the study and test tasks and found an increment in the magnitude of priming as a function of deeper levels of processing and stable study-to-test voice effects. The concurrent manipulation of both study and test procedures makes the task of isolating the effects on encoding and retention from those on retrieval arduous. Thus, Goldinger's findings cannot clearly speak in favor of either model. The same argument applies to Lively's experiment, in which a lexical decision task was employed both at study and at test (1994). Lively's experiments on levels of processing (1994), however, are relevant to the present discussion. As described previously, Lively administered three combinations of study and test procedures: lexical decision and gating, word identification and gating, and word identification and stem completion. Except for lexical decision, the tasks are, for lack of a better terminology, "data-driven". The combinations of interest are those that share the same data-driven memory task. The *multiple-memory system* model would predict no effect of the manipulation of the study task. Alternatively, the *transfer-appropriate processing* model would predict larger priming effects in the experiment in which word identification is employed as a study task based on the fact that word identification and gating are more likely to share operations than lexical decision and gating. In support of the *transfer-appropriate processing* model, the administration of a word identification task followed by gating produced 76 msecs priming, while the administration of a lexical decision task followed by gating only produced a 31-msecs priming effect. Although no direct test for this difference in magnitude was reported, Lively (1994) argues that the consistency of the

operations performed at study and at test yielded a larger priming effect. However, the *transfer-appropriate processing* model would also predict a decrement in priming following study-to-test voice changes in the presence of study and test procedures known to be sensitive to surface features such as voice information. No voice change effects were observed. If the difference in the magnitude of priming effects is taken into account and assumed to be statistically significant, this evidence both supports and questions the *transfer-appropriate processing* model. However, if one takes into account the fact that studied words were repeated several times in different voices rather than once at study, the absence of voice change effects does not challenge the model.

The Assumptions underlying Encoding Manipulations and Task Sensitivity: The Problem. Some findings support the *multiple-memory system* model (Schacter & Church, 1992; Sheffert, 1995) and others support the alternative model (Lively, 1994). A possible explanation for this discrepancy arises from the largely untested assumption, made to justify encoding instruction manipulations, that explicitly directing participants to focus on a specified type of information rather than another during study guarantees that the underlying encoding processes will be qualitatively different (see Schacter & Church, 1992, for an exemplification of this assumption). This is to say that if the study task emphasizes the processing of meaning as opposed to the processing of perceptual details, the encoding of stimulus information will be similarly biased. A comparable assumption applies to the task instructions at test. In fact, the sensitivity of a task to different types of information is defined by the memory test instructions, which, thought to engage or simply emphasize the retrieval of certain types of information, bias the processing of the repeated

stimuli. As for the study phase, a direct test of this assumption, which measures the informational content of perceptual operations, is largely absent from the literature. Instead researchers have measured the type of information that supports priming as key evidence for classifying a memory procedure as either data- or semantically-driven (Rajaram & Roediger, 1993). For example, if a memory test produces evidence for study-to-test voice changes, the test is data-driven (see Schacter & Church, 1992).

8. Methodological Considerations: A brief note on contamination

The Need for On-Line Testing of Perceptual Operations. The understanding of implicit memory in the auditory domain requires the conceptualization of speech perception and memory processes as operations distributed over time. There is no passive window, however, that allows for the examination of speech perception and memory without affecting, to a certain extent, the operations upon which they rely. Both the *multiple-memory system* and the *transfer-appropriate processing* model embody the idea that tasks are not "process pure" (Jacoby, 1991). The choice of experimental procedures is thus crucial in determining the set of interpretations that can be made out of the evidence collected. Evidence is only the final result of a series of processes that cannot be directly observed. Some of these processes reflect spurious operations, which are unique to a memory task and of little interest to the investigation of memory.

Both the *multiple-memory system* and the *transfer-appropriate processing* models depict memory as a set of post-perceptual processes that produce structural traces. The *multiple-memory system* argues that these processes produce the traces necessary for pattern recognition, which are, at different degrees, captured by memory tasks. The

transfer-appropriate processing model postulates that task requirements modulate processes both at study and at test. It is the overlap of operations between the two phases that determines the content of implicit memory.

A test for these models must recognize the validity of Jacoby's idea (1991) that tasks deprive researchers of an uncontaminated picture of a psychological phenomenon. It also has to recognize that knowledge of the operations required by a task can provide an intelligible picture of that psychological phenomenon. This knowledge can be gathered from comprehensive comparisons of the priming effects obtained with different tasks (see Rajaram & Roediger, 1993). The distinction between "data-driven" and "semantically-driven" tasks is an attempt in this direction. However, as previously suggested, an accurate analysis of perceptual operations, both at study and at test, is a necessary step for obtaining a more explicit conceptualization of this dichotomy. This argument applies to the processes involved in reading printed text and perceiving speech.

The present investigation embodies the notion that without an on-line test for the type of information that is processed both at study and at test, the basis for encoding manipulations and the argument on test sensitivity rest on shaky grounds. A stringent test for the *multiple-memory system* and *transfer-appropriate processing* models can only result from the concurrent analysis of the information processed during speech perception both at study and at test and the information that supports priming. The *transfer-appropriate processing* model predicts isomorphism. On the other hand, the *multiple-memory system* model predicts that, despite the information that comprises the activity traces of speech perception at study, priming will rely on pre-lexical information (voice and phonemic

sequences at the subword level). It predicts, however, that the content of the activity traces of speech perception at test, guided by task instructions, has to be largely isomorphic to the structural traces that support priming for these structural traces to emerge in performance. Thus the memory test has to be, as much as possible, data-driven. The present investigation provides a test of the conflicting predictions that the *multiple-memory system* and the *transfer-appropriate processing* models make regarding speech perception operations and priming by selecting a memory task that satisfies this requirement.

The Need for a Memory Procedure that Removes the Word-Level Response Bias. The choice of the appropriate procedure for testing these models results from first re-conceptualizing the dichotomy between data-driven and semantically-driven procedures as a continuum along which tasks that are primarily sensitive to certain types of information can be theoretically and practically arranged. The arrangement of tasks on the data-driven end of the spectrum is then based on the extent to which lexical information is ingrained in the performance of each task. Data-driven memory tasks, such as stem completion, word identification, gating, etc., constrain participants to respond to stimuli as whole words and thus it coerces them into explicitly activate and use lexical information. This behavioral restriction (e.g., bias) incorporated in the procedures employed by the investigations reviewed in the previous sections, with the exception of Goldinger's experiment on levels of processing, produces two notable and problematic effects, which are reviewed in the next paragraphs.

It is hypothesized that the word-level response bias produced by these memory procedures modulates the speech perception operations performed at test, so that the

activation of lexical rather than pre-lexical information is enhanced. This enhancement is produced in different degrees by each procedure and may lead to a substantial decrement in the sensitivity of the memory task to pre-lexical information such as voice characteristics. The extreme example of this decrement is lexical decision (see Lively, 1994). Indeed, the concept of "sensitivity" of a memory task to different forms of information refers to the modulation of speech perception operations at test. In the context of the *multiple-memory system* and *transfer-appropriate processing* models of auditory repetition priming, speech perception operations at test are then, in different degrees, unable to engage the pre-lexical information that might have produced priming. Contrary to the *transfer-appropriate processing* model, however, the *multiple-memory system* model implies that this is the only available information.

In addition to underestimating the contribution of voice information to priming, it is reasonable to hypothesize that the findings obtained in the investigations reviewed above are contaminated by the word-level response bias to different degrees. In fact, if participants are asked to respond to stimuli as whole words, it is to be expected that repetition priming will rely on lexical information (Weldon, 1991). Although the supporting evidence stems from the visual domain, this argument can easily apply to the auditory domain. The enhancement of lexical information in word-level procedures may thus introduce spurious lexical information into the priming effect. This means that priming will be contaminated by the lexical information activated during speech perception at test, but not included in the memories that sustain priming. Recall that the *multiple-memory system* model claims that priming results from pattern-recognition operations,

which are pre-lexical processes. The test of this claim cannot be carried out when the memory procedure constrains responses to the word level.

The present investigation employs a phoneme monitoring procedure (Foss & Swinney, 1973), which can enhance the processing of pre-lexical information, such as voice characteristics, in speech perception (Mullenix & Pisoni, 1990) without constraining repetition priming to a lexically-based behavioral response and, under certain experimental conditions, without inhibiting lexical processing in on-line speech perception. In fact, phoneme monitoring responses, which are indicative of speech perception operations, can be made sensitive to lexical variables, such as word frequency and lexical status, by embedding words in sentential contexts or by increasing momentary processing load when isolated words are employed (Eimas, Hornstein, & Payton, 1990; Eimas & Nygaard, 1992; Foss & Swinney, 1973; Rubin, Turvey, & Van Golder, 1978). Phoneme monitoring thus represents a unique opportunity to test the alternative predictions of the *multiple-memory system* and *transfer-appropriate processing* models. It allows researchers to measure the effect of pre-lexical and lexical perceptual processes on priming without constraining priming to a lexically-based behavioral response. According to the *multiple-memory system* model, if pre-lexical information is the only component of the memories that sustain priming, in a phoneme monitoring task, used both at study and at test, lexical information should affect speech perception, but not priming. The *transfer-appropriate processing* model, on the other hand, predicts isomorphism between the information processed by speech perception and the information that produced priming. A discussion of this procedure will be presented in the introductory section of the first experiment that

comprises this investigation.

9. Implicit memory for speech: A preliminary modular model

The Model. Most of the findings reviewed above suggest that implicit memory of spoken words entails encoding, retention and retrieval operations that preserve relevant physical characteristics of spoken words. The implicit memory of a spoken word can be thought of as a unique memorial episode. The present investigation is based upon a preliminary model of auditory repetition priming. Its basic assumptions result from or represent a specification of Schacter and Church's model of implicit memory for spoken words (1992). The intent is to develop a model of the most basic memory processes that must exist according to all available evidence.

It is hypothesized that in the study phase, speech perception operations assemble in short-term memory a unique set of internal representations of a speech stimulus by blending preexisting linguistic knowledge from long-term memory with the acoustic information gathered from that stimulus. It is further hypothesized that since no intentional learning strategies, such as overt or covert rehearsal, are likely to be used during study, only two general cognitive operations are involved in the perception of speech: a process of generation that gathers from the speech stimulus a set of auditory characteristics that pertains to a speaker's voice and a process of activation of preexisting knowledge regarding legal sequences of phonemic sounds and semantic properties that, in the listener's language, represent words.

This unique blend of information in short-term memory, produced by speech perception operations at study, can be thought of as an "activity trace" (Hebb, 1949; 1961),

that is, a trace that does not entail a structural, long-lasting change in long-term memory. During the study phase and, perhaps, even during the distractor phase, portions of the unique blend of information that comprises the representations developed by speech perception processes become part of long-term memory by producing a structural change (Hebb, 1949; 1961). This change is structural in the sense that it outlasts the initial perceptual operations carried out on the speech stimulus. Long-term memory consists of distinct systems devoted to the processing of different types of information (Schacter and Church, 1992). It is hypothesized that the long-term memory representations produced by the study phase involve structural changes in a system devoted to pattern recognition operations, such as those necessary to perceive a speech stimulus as a familiar sequence of phonemes and to recognize a voice as a familiar sound. The structural changes accommodated by this system are thus conceptualized as episodic memory representations in which phonemic invariance is augmented with the uniqueness of voice information.

At test, these structural changes facilitate speech perception operations on repeated words (studied) relative to novel words (non-studied). Speech perception operations for studied words therefore entail reactivation of preexisting long-term episodic representations. The degree of matching between episodic representations and speech stimuli determines the extent to which speech perception operations will be facilitated by repetition. This facilitation is thus characterized by selectivity in the sense that it depends on the information that has defined the structural memory changes. Selectivity, however, is simply a necessary, but insufficient condition for the emergence of facilitation. Facilitation of speech perception operations at test is contingent on the ability of the

memory procedure to gather the information that has produced structural changes. If the test is implicit (e.g., it does not require additional processing such as intentional recollection) as well as sensitive to these changes, the advantage of studied words over non-studied will entail pattern-recognition information. On the other hand, non-studied words will be subjected to speech perception operations that do not rely on preexisting episodic representations.

If one accepts the assumption that repeated exposure to a speaker's voice leads listeners to extract from the several episodic representations of individual utterances, a memory of that voice (Nygaard, et al., 1994), then voice familiarity should enhance priming by adding to the strength of the episodic representation of a repeated word the impact of the "abstract" memory of that voice. Similarly, if a multiple-speaker environment disrupts the generation of episodic memories, sequentially-ordered voice changes at study would diminish processing interference and thus improve priming. However, if the memory system fully functions as a module, the ability to predict voice changes, gathered at study and employed at test, should not have any effect on priming.

This modular model leads to an important assumption regarding implicit memory for spoken words. Speech perception operations guided by the study task produce activity traces that contain a richer blend of information than the structural traces (e.g., the long-term representations) gathered by the implicit memory test. Although speech perception operations may lead to representations that contain voice characteristics along with lexical information, such as legal sequences of phonemes comprising the sound of a word and clusters of semantic features embracing the meaning of that word, the long-term memory

representations that support auditory repetition priming will be pre-lexical. That is, they will entail both voice and legal sequences of phonemes at the subword level.

The Modular Model in a Larger Context. This model is consistent with an episodic conceptualization of implicit memory in the auditory domain. The postulation of an informationally encapsulated and autonomous system, which produces perceptually detailed structural traces of incoming speech stimuli, renders this model a refinement of the *multiple-memory system* conceptualization of implicit memory. Ingrained in this model is the assumption that priming is insensitive to both lexical processing and strategic operations. The assumption that acquisition and activation of preexisting representations contribute to priming, renders this model a hybrid in reference to the acquisition versus modification dichotomy of memory.

The present model, however, is different from the hybrid model proposed by Feustel, Shiffrin, and Salasoo (1983) and Salasoo, Shiffrin, and Feustel (1985), which attempted to account for the effects of repetition on words and pseudowords. Words were found to be identified more accurately than pseudowords after one repetition. After additional repetitions, however, the difference in accuracy disappeared. After a one-year delay, repeated words and pseudowords were recognized as accurately as novel words, but better than novel pseudowords. On the basis of these findings, Feustel, et al. (1983), and Salasoo, et al. (1985) suggested that episodic traces, that is, memories of the individual encounters with a stimulus in the experimental setting, contribute to repetition effects. These episodic traces, however, contain all sorts of abstract-symbolic information and only a limited amount of surface information. Feustel, et al. (1983), and Salasoo, et al. (1985)

claimed that the advantage that words have over pseudowords depends on the fact that words benefit from preexisting lexical representations. Memory of pseudowords was assumed to entail a process of "codification" through which a unitized trace (e.g., a pseudo-lexical representation) is developed. When the process of codification is accomplished through about five repetitions, both repeated words and pseudowords can benefit from unitized traces. After a long delay, all the stimuli that can rely on these unitized traces (e.g., repeated words, repeated pseudowords and novel words) display an advantage over novel pseudowords. The modular model presented here postulates that priming in the auditory domain is based on pre-lexical memory traces and the findings support this model. Contrary to Feustel, et al. (1983) and Salasoo, et al. (1985), it is argued that these traces entail detailed representations of the physical characteristics of a stimulus word that are the same for words and pseudowords, even if speech perception is sensitive to such a distinction. The differential effect of one repetition on words and pseudowords described by Feustel, et al. (1983) and Salasoo, et al. (1985) can be accounted for by word-level biases, and thus it does not contradict the predictions of the modular model.

The notion that priming is based on both stimulus-specific and preexisting abstract information, ingrained in the modular model, is consistent with the episodic approach of MINERVA 2 (Hintzman, 1986, 1988; Hintzman, Nozawa, & Irmscher, 1982). Although MINERVA 2 was developed to account for explicit memory performance, it is relevant to the present discussion because it assumes that each stimulus generates an episodic trace. Even additional repetitions of that stimulus are assumed to produce individual episodic

traces. Perceptual details are assumed to comprise the record of that stimulus along with other types of information including meaning. A trace is a collection of these different types of information ("primitives" in Hintzman's language). Hintzman distinguished between primary and secondary memories (James, 1983), which, in the modular model introduced here, correspond to the activity traces of speech perception and to the structural traces in long-term memory, respectively. A repeated stimulus produces an episodic trace in primary memory. In the modular model, this is the trace that results from speech perception. The trace in primary memory activates traces in secondary memory on the basis of similarity. The more traces activated, the more likely the memory for that stimulus will be to reflect common rather than distinctive properties. MINERVA 2 can thus account for Lively's findings (1994) of no voice effects following several repetitions of a word, each time in a different voice, by postulating that the individual traces for that word activate a memory that reflects the characteristics that are shared by all the different traces. Because abstract-symbolic information rather than voice-specific information is what all the different traces have in common, the memory for that word appears to be generalized over the different voices so that priming does not display voice specificity. MINERVA 2 also predicts that the more unique a repeated stimulus is, the more likely it will be to activate its detailed trace rather than a collection of traces. It follows that study-to-test voice changes should be stronger for pseudowords than for words. This prediction will be examined in the present investigation. Although MINERVA 2 appears to display an overall affinity with the modular model, it is inconsistent with the assumption that not all the "primitive properties" of a trace in primary memory will transfer to secondary

memory.

The general properties of McClelland and Rumelhart's (1985) distributed memory model accounts for both abstract and episodic representations of repeated stimuli at short delays and it is thus only partially consistent with the modular model. This model postulates that memories of a stimulus are patterns of activation over a network of interconnected units. Units are organized into modules, which are also interconnected with each other. As the pattern of activation for each stimulus is somehow unique, the model predicts that for brief periods of time, repetition priming will depend on a detailed representation of that input. However, this effect should not last long. The network is an adaptable structure that adjusts to incoming stimuli. Individual stimuli have considerable effects on the pattern of activation of the network. With time, decay will consistently minimize the residual activation of that pattern. After several repetitions of a stimulus, a robust pattern of activity will stabilize the activation pattern into a structural trace. After several different stimuli, however, the pattern of specificity following one repetition of a stimulus will probably be lost. After several slightly different repetitions of a stimulus, an overall generalized pattern of activation should produce priming, which then loses its specificity. As priming's sensitivity to voice can be observed at long delays (see Goldinger, 1992), McClelland and Rumelhart's model encounters the problem of explaining how this effect can be obtained after long delays and several intervening stimuli. This model could be improved by modifying the learning algorithm responsible for memory traces. If the learning algorithm that governs the network's ability to adapt to incoming stimuli is made to reflect two components, one that produces small, slow, but

relatively permanent changes and another that leads to large, fast, but temporary changes, priming based on episodic representations can be made to last for longer periods of time (see Cleeremans & McClelland, 1991, for additional qualifications). The modular approach proposed here is also consistent with McClelland and Rumelhart's conceptualization of processing modules.

Testing the Modular Model: The Experiments. In the following chapter, three experiments are presented. All of them provide a test for the different predictions that result from the modular model introduced in this section. First, it is of interest to test the hypothesis, supported by previous findings, that implicit memory for spoken words results from episodic memories that entail detailed representations of speech stimuli. These episodic memories are conceptualized as structural traces that contain information about the speaker's voice. Study-to-test voice changes in a repeated word will be employed to test this prediction. Second, the hypothesis that the activity traces produced by speech perception operations are not exact duplicates of the structural changes that represent implicit memories of spoken words will be tested. Speech perception operations at both study and test will be made sensitive to word-level factors, such as word frequency and lexical status, without constraining responses to explicit word level processing. The effect of these factors on repetition priming will then be investigated. In addition to these tests, Experiment 1 will assess whether Goldinger's finding (1992), relative to repetition priming in a multiple-speaker acoustic environment, can be replicated in a more restricted acoustic

environment where the implicit memory procedure is not biased toward the word level. Experiment 2 will investigate whether this effect can be modulated by the ability of listeners to predict word-to-word voice changes at test. Experiment 3 will estimate the effect of voice familiarity. In the context of an experimental procedure that is specifically designed to eliminate word-level biases and enhance pre-lexical processing, the modular model leads to the following assumptions:

1. Implicit memory for spoken words results from episodic memories that entail voice-specific representations of speech stimuli.
2. Implicit memory for spoken words is the result of pattern-recognition processes, which produce pre-lexical long-lasting memory representations.
3. The temporary memory representations produced by speech perception at study and at test and the long-term memory representations that support priming are not isomorphic.
4. The detrimental effect of speaker variability on speech perception (Mullenix & Pisoni, 1990) results in weak memory traces largely unable to sustain priming.
5. The pattern-recognition processes responsible for implicit memory of spoken words are insensitive to expectations. At test, the predictability of word-to-word voice changes in a multiple-speaker environment does not improve implicit memory of spoken words. Stable word-to-word changes at study, however, strengthen the encoding of voice information and thus memory.
6. Voice familiarity, however, facilitates the encoding of voice and thus produces more robust structural traces.

Chapter III: Experiments

Experiment 1

The effects of Voice Repetition and Acoustic Environment on Repetition Priming

Auditory repetition priming occurs when a response to a repeated spoken word is facilitated without intentional recollection (Church & Schacter, 1994). Auditory priming is therefore regarded as the behavioral correlate of implicit memory for spoken words.

It is generally assumed that, at study, speech perception entails operations on the speech stimulus, which generate or activate internal representations (Pisoni & Sawusch, 1975). At test, the operations performed on that same stimulus are facilitated as a result of these memory representations (Squire & Knowlton, 1995). For the purpose of this experiment, operations are broadly classified as pre-lexical and lexical. Pre-lexical operations consist of pattern-recognition processes that develop representations of auditory characteristics such as voice and phonemic sequences at the subword level. Pattern recognition processes are insensitive to word-level properties, such as word frequency and lexical status (Pisoni, et al., 1985). Lexical operations produce representations pertaining to words, including a word's phonemic sequence, its semantic and syntactic properties.

Traditionally, investigations of auditory priming have relied on implicit memory procedures that require participants to report whole words, rather than subword units (Church & Schacter, 1994; Goldinger, 1992; Jackson & Morton, 1984; Lively, 1994; Schacter & Church, 1992). This response bias, generated by procedures such as word stem completion, word identification and lexical decision, renders priming dependent on the activation of lexical information (Weldon, 1991; Lively, 1994), while focusing

performance away from the acoustic component.

The acoustic characteristics of a speaker's voice are encoded in speech perception (see Goldinger, et al., 1991, for additional qualifications), and become a component of the implicit memory record of spoken words in conjunction with more abstract symbolic information (Goldinger, 1992). The contribution of voice information to priming, however, has been found to depend on the degree of sensitivity of the memory task to surface rather than word-level processing. In fact, a voice change between a spoken word and its repetition has been shown to have a detrimental effect on auditory priming in stem completion and word identification (Church & Schacter, 1994; Goldinger, 1992; Schacter & Church, 1992), but not in lexical decision (Lively, 1994).

The purpose of this experiment is to introduce an implicit memory procedure designed to eliminate the word-level response bias and enhance the sensitivity to acoustic voice processing. To focus attention on the subword level, the present experiment employs a phoneme monitoring task at both study (encoding) and test. In phoneme monitoring, participants are required to respond to word-initial phonemes, such as the phoneme [b] of the word [boat]. Spoken words presented at study are repeated at test along with other words. Auditory priming is indexed by faster response latencies to previously studied words than to non-studied words. A one-speaker reference condition, in which study words are repeated at test in the same voice, is utilized to ascertain the presence of priming with this procedure.

To provide a test for the hypothesis that auditory repetition priming results from pattern recognition, speech perception operations in phoneme monitoring are made to

include word-level information. Although phoneme monitoring does not bias responses toward this level, earlier research suggests that responses to word-initial phonemes involve lexical operations when the task is made complex by increasing momentary processing load (Eimas, Hornstein, & Payton, 1990; Eimas & Nygaard, 1992; Rubin, Turvey, & Van Gelder, 1976). In this experiment, processing load is increased by requiring participants to make a choice response to words beginning with one of four target phonemes: /b/, /p/, /d/, and /t/ (i.e., participants respond to each of the stimuli on the basis of the presence or absence of a target phoneme rather than producing a response to only stimuli that begin with a given target phoneme). Under these processing load requirements, pilot work has shown that speech perception operations in phoneme monitoring involve lexical processing (see Lively & Pisoni, 1990, for contrasting evidence in a slightly different procedure). Thus, phoneme monitoring, used at both study and test, represents a unique experimental setting in which speech perception operations can be made sensitive to both voice and lexical processing without explicitly emphasizing the latter. It is hypothesized that, even if speech perception operates at the lexical level, auditory repetition priming should entail phonological information at the subword level along with voice information. Thus, if priming is not sustained by word units, comparable priming effects should be obtained with high and low frequency words.

Although it has been recognized in the literature that the detection of voice effects on auditory priming depends on the sensitivity of a memory task to surface processing, the detrimental impact of acoustic variability on the voice component of priming has not been fully appreciated. The effects of voice information on priming have been primarily

obtained in multiple-speaker acoustic environments (Church & Schacter, 1994, Exp. 1; Goldinger, 1992; Lively, 1994; Schacter & Church, 1992), although, these environments are also known to impair word identification (Mullenix, Pisoni, & Martin, 1989) and to minimize the encoding of voice (Goldinger, 1992). In the present experiment, it is hypothesized that by de-emphasizing word information, as in phoneme monitoring, the sensitivity of the task to the acoustic voice component of priming would be enhanced. Consequently, the detrimental impact of acoustic variability would be accentuated. In fact, previous studies have shown that performance in a phoneme monitoring task is disrupted by voice changes between words (Mullenix & Pisoni, 1990). Voice changes between contiguous words are employed to assess the effect of acoustic variability on priming. Voice changes are produced by either two, three, or four speakers. On the other hand, the contribution of voice characteristics to priming is measured by introducing voice changes between study and test.

In summary, the experiment first examines whether an implicit memory procedure devoid of the word-level response bias produces priming, and whether priming, if any, is influenced by acoustic variability. Second, the episodic and pre-lexical nature of the memory traces that support priming is estimated by observing the effects of study-to-test voice changes and word frequency, respectively. The concurrent analysis of word frequency effects on speech perception allows for an examination of the modular pre-lexical nature of implicit memory processes.

Method

Participants

Three hundred and twenty undergraduate students from the College of Staten Island volunteered. They were native speakers of English, and reported no history of speech or hearing disorder. Thirty-two participants served in each of the between-subjects conditions. Three additional participants, who failed to reach a 90% accuracy rate in preliminary training, were excluded from the analyses.

Stimuli and Apparatus

The stimuli consisted of 48 monosyllabic English words divided into two sets of 24, in which 12 stimuli began with the phoneme h and 12 stimuli began with either t, d, or p (4 of each phoneme). Each set was matched for frequency (Francis & Kucera, 1982), such that half of the words were high frequency (mean frequency = 226.88, median = 151.5) and half low frequency (mean frequency = 3.25, median = 3.00). Each set was presented as a study list equally often. The test list included all 48 words, so that half were previously studied and half were not.

Each of the 48 stimuli was recorded by four speakers, two males and two females, using a Unidyme II microphone. These four speakers and two additional speakers, one male and one female, recorded 72 similar stimuli to be used for preliminary training. Lists consisted of words spoken by either one, two (a male and a female), three, or four speakers.

Each stimulus was low-pass filtered at 4.8 kHz and digitized at an 11-kHz sampling rate by using a 16-bit analog-to-digital converter equipped with anti-aliasing filters,

installed on an IBM-compatible 486 personal computer. A signal-processing package was utilized to create a digital file for each word, mean length = 501 msecs. Words were rerecorded until two naive independent observers rated them to be intelligible and clear. The presentation of the stimuli, filtered at 4.8 kHz, on calibrated Koss TNT/77 headphones at 75 dB and the recording of the participants' responses were accomplished by a custom-made program.

Design

Four same-speaker baseline conditions, consisting of either one, two, three, or four speakers, were utilized. Each studied word was spoken by the same speaker during the test phase. The speakers of the preliminary training were always different from those utilized in the study and test phases. The length of the study and test lists was kept constant. Therefore, as the number of speakers increased, the total number of exposures, or familiarity, to each speaker decreased. To obtain an unconfounded measure of speaker variability, the multiple-speaker baseline conditions were replaced with familiarity conditions. This was accomplished by using the speakers of the study and test phases in preliminary training and thus, where necessary, extending training. The 72 exposures to a voice in the same one-speaker condition served as the reference point (i.e., 24 study words and 48 test words). There were 36, 24, and 18 exposures to each speaker in the two-, three-, and four-speaker conditions, respectively. The number of exposures to a given voice was equated to the reference by presenting, in preliminary training, each voice 36, 48, and 54 times, respectively.

To assess the importance of voice repetition, when priming was found in any of the

above baseline or familiarity conditions, corresponding different speaker conditions were devised. In these conditions, each studied word was spoken at test in a different gender voice. Voice changes in the familiarity conditions entailed familiar voices.

In summary, the experimental design consisted of the following factors: Priming (studied and non-studied: within-subjects); Number of Speakers (one, two, three, and four speakers: between-subjects); Voice (same and different: between-subjects); Word Frequency (high and low frequency: within-subjects); and Target Phoneme (presence of b and absence of b: within-subjects).

Procedure

The experiment was presented to participants as a study on perception consisting of a series of brief tasks. Participants were first exposed to preliminary training, with feedback, to assure a high level of accuracy and to minimize speed-accuracy tradeoffs. They were asked to listen to a list of stimuli and to indicate, as quickly and accurately as possible, whether or not each stimulus began with the phoneme b by pressing the right or left mouse key respectively. An overall accuracy rate of 90% was required to proceed to the study phase of the experiment.

Following exposure to the study list, a 10-minute distractor task modeled after Posner, Boies, Eichelman, and Taylor's visual matching task (1969) was presented. Upon completion, the instructions regarding the phoneme monitoring task were repeated and the test list was presented without reference to the prior study list. No feedback was provided during either the study or the test phase.

The sequence of each trial consisted of a 500-msec, 1000-Hz tone, a 500-msec

silent interval and the presentation of the stimulus. The order of stimulus presentation was independently randomized for each participant and response latencies were computed from word onset. A trial ended and the next trial began after a response was made or if no response occurred after 3000 msecs.

In the one-speaker condition, words from both study and test lists were randomly presented. In all the multiple-speaker conditions, the words were presented in a quasi-random order, such that contiguous words were never in the same voice. The speaker's voice assigned to each word was counterbalanced across participants within speaker conditions so that any word was uttered equally often by each of the four speakers. In multiple-speaker lists, an equal number of high and low frequency words were spoken by each voice.

Results and Discussion

Preliminary training maximized accuracy (overall mean accuracy = 98.65%), and permitted response latencies to be the primary dependent variable (see Table 1). Latencies two standard deviations above and below each participant's mean were excluded from the final analyses (4.38%). In this experiment, *priming* was operationally defined as statistically significant faster response latencies to studied than to non-studied words. When priming was found in any of the experimental conditions, priming scores were analyzed. *Priming scores* were obtained by subtracting the mean latency of studied words from the mean latency of non-studied words. An alpha level of .05 was applied to all a priori pairwise comparisons (Kirk, 1982).

Table 1
Mean Latency and Standard Deviation of Correct Identifications and Mean Percent Correct as a Function of Priming and Speaker Condition

Priming	Response Latency		Priming Score	% Correct
	Mean	SD		Mean
Same One-Speaker Condition				
Non-studied	714	121		98.96
Studied	677	107	+ 37	99.09
Same Two-Speaker Condition				
Non-studied	744	172		98.31
Studied	720	170	+ 24	98.44
Same Three-Speaker Condition				
Non-studied	709	161		98.18
Studied	696	146	+ 13	98.83
Same Four-Speaker Condition				
Non-studied	745	193		98.57
Studied	741	172	+ 4	98.96
Same Two-Speaker Familiarity Condition				
Non-studied	755	155		98.31
Studied	726	130	+ 29	98.44
Same Three-Speaker Familiarity Condition				
Non-studied	775	184		98.31
Studied	760	172	+ 15	98.18
Same Four-Speaker Familiarity Condition				
Non-studied	718	140		98.70
Studied	707	121	+ 11	98.05
Different One-Speaker Condition				
Non-studied	703	121		98.83
Studied	693	106	+ 10	98.70
Different Two-Speaker Familiarity Condition				
Non-studied	758	217		97.66
Studied	753	205	+ 5	98.96
Different Three-Speaker Familiarity Condition				
Non -studied	746	181		98.83
Studied	757	180	- 11	99.22

Priming at different levels of speaker variability. To obtain a preliminary estimate of the effect of the increased variability and diminished voice familiarity of the acoustic environment on same-voice repetitions, a test for linear trends was conducted on the priming scores of all the baseline conditions (one-, two-, three-, and four-speaker). Priming scores linearly decreased as a function of the number of voices heard during test, $F(1, 124) = 13.54$, $p < .01$, $MSe = 1440.18$.

To determine whether auditory repetition priming occurred in each of the four same-speaker baseline conditions, a priori pairwise comparisons were conducted between the response latencies of non-studied and studied stimuli at each level of the Number of Speakers factor. The one-, two- and three-speaker conditions displayed a significant priming effect. The additional operations required to process an increasingly variable and progressively less familiar acoustic environment (Mullenix & Pisoni, 1990) disrupted the ability of voice and word repetition to sustain priming in the four-speaker condition. To evaluate whether the magnitude of priming varied as a function of the number of voices comprising the acoustic environment, the priming scores of the speaker conditions that displayed a significant effect of same-voice repetition were analyzed. A one-way ANOVA with Number of Speakers as factor suggested that priming decreased as a function of the acoustic environment, $F(1, 93) = 3.51$, $p < .034$, $MSe = 1349$. Thus, in an implicit memory test emphasizing operations at the subword level, the advantage of studied words over non-studied was dependent upon the acoustic environment.

To partial out the effect of acoustic variability from the impact of decreased voice familiarity, the two-, three- and four-speaker conditions were replaced with their

corresponding familiarity conditions (see Table 1). A test for linear trends on the priming scores of the one-speaker baseline and the three familiarity conditions showed that acoustic variability alone still produced a linear decrement on same-voice repetitions, $F(1, 124) = 5.59$, $p < .05$, $Mse = 2514.79$. Pairwise comparisons between the response latencies of non-studied and studied stimuli at each level of the Number of Speakers factor also showed that priming occurred in the one-, two-, and three-speaker familiarity conditions, as indicated earlier. When voice familiarity was equated to the one-speaker reference, the detrimental impact of acoustic variability on priming was limited to the four-speaker familiarity condition. To assess whether priming in these three conditions was a function of the number of speakers, a one-way ANOVA was conducted on priming scores. No effect of Number of Speakers was observed, $F(2,93) = 2.77$.

The effect of study-to-test voice changes on priming. To investigate whether priming was sensitive to study-to-test voice changes, response latencies of the same speaker conditions that produced significant priming effects (one speaker, two familiar speakers, and three familiar speakers) were combined with those obtained in the corresponding different speaker conditions. These response latencies were submitted to a 2 X 2 X 3 ANOVA with Priming (studied and non-studied), Voice (same voice and different gender voice between study and test) and Number of Speakers (one speaker, two speakers, and three familiar speakers) as factors. There was a main effect of Priming, $F(1,186) = 16.96$, $p < .0001$, $MSe = 1132.62$, and a significant interaction of Priming with Voice, $F(1,186) = 13.86$, $p < .0001$, $MSe = 1132.62$, and with Number of Speakers, $F(2,186) = 3.38$, $p < .036$, $MSe = 1132.62$ (all the other $Fs < 2.82$). The significant interaction of Priming and

Voice demonstrated that priming was sensitive to study-to-test voice changes. Pairwise comparisons indicated that these voice changes eliminated priming and thus showed that voice information was a critical component of auditory priming.

The effect of word frequency on speech perception and priming. Further analyses were conducted to determine the effects of Word Frequency and Target Phoneme on priming. A 2 X 2 X 2 ANOVA with Priming (studied and non-studied), Word Frequency (high and low), and Target Phoneme (presence and absence) was performed on the response latencies of the conditions that produced significant priming (same one speaker, two speakers, and three familiar speakers). This analysis yielded a main effect for Priming, $E(1,95) = 86.79$, $p < .0001$, $MSe = 7648.38$, Word Frequency, $E(1,95) = 24.16$, $p < .0001$, $MSe = 5764.72$, and Target Phoneme, $E(1,95) = 27.12$, $p < .0001$, $MSe = 3422.41$, but no significant interactions ($Es < 1.65$). Response latencies of words beginning with b were, on the average, 59 msecs faster than those of words beginning with another phoneme. Responses were, on the average, 22 msecs faster to high frequency words than to low frequency. The absence of a significant interaction between Priming and Word Frequency showed that the memory traces that supported auditory priming were independent of a word-level factor such as frequency of occurrence in the language. Priming appeared to rely on pre-lexical memory representations.

The main effect of Word Frequency on the response latencies of the test phase indicated that the processing of stimuli involved the lexical level. To determine whether, at study, stimuli were also encoded at the word level, a 2 X 2 X 3 ANOVA was conducted on the response latencies of the study phase of the conditions that produced priming at test

(one-, two- and three-speaker familiarity conditions) with Word Frequency, Target Phoneme, and Number of Speakers as factors. A main effect of Word Frequency and Target Phoneme were observed, $F(1,92) = 74.06$, $p < .0001$, $MSe = 5201.19$; and $F(1,92) = 4.17$, $p < .044$, $MSe = 5170.37$, along with their significant interaction, $F(1,92) = 4.30$, $p < .041$, $MSe = 3311.05$. At study, responses to high frequency words were 64 msecs faster than to low frequency. Responses to words beginning with b were 16 msecs faster than to words beginning with another stop consonant. Taken together, these findings indicated that, both at study and at test, stimuli were processed at the word level, but the memory traces that were responsible for auditory priming resided at the subword level.

Accuracy. The absence of priming effects in the analyses of the mean percent accuracy depleted further analyses of any theoretical justification. Because the stimuli were clearly presented and the task relatively easy, a ceiling effect had rendered this measure uninformative. Thus, no further analyses are reported on this dependent measure.

General Discussion

In summary, the same one-speaker condition produced faster response latencies to studied stimuli, in which voice repetition and word repetition were combined, than to non-studied stimuli, for which only voice repetition occurred. When voice and word repetition were maintained in a multiple speaker environment, where voice changes occurred between words, acoustic variability and reduced voice familiarity had a detrimental effect on priming. Increased acoustic variability and reduced voice familiarity, as entailed by two and three voices, preserved priming. When acoustic variability further increased and

familiarity decreased with four voices, priming was eliminated. Apparently, when the encoding of acoustic information required the processing of different and progressively less familiar voices, voice and word repetition were not sufficient to support priming. When voice familiarity was equated to the one-speaker reference, the detrimental effect of acoustic variability remained restricted to four voices. When voice repetition was removed from both the studied and non-studied stimuli, equivalent response latencies were obtained. At the level of processing entailed by this task, word repetition was dependent on voice repetition to produce auditory priming. This evidence is consistent with the reliance of phoneme identification performance on voice information (Mullenix & Pisoni, 1990) and supports the notion that, compared with word level tasks, such as word identification, word stem completion, and lexical decision (Church & Schacter, 1994; Goldinger, 1992; Lively, 1994; Schacter & Church, 1992), phoneme monitoring emphasizes surface levels of processing such as those that operate on acoustic and subword information.

Although phoneme monitoring directed attention to subword units, the effect of word frequency on response latencies showed that speech perception operations entailed word-level information. The absence of an effect of word frequency on priming strengthened the idea that implicit memory for spoken words, at the level of processing entailed by phoneme monitoring, is based on surface rather than lexical processing (see Schacter and Church, 1992, for additional qualifications).

The conceptualization of implicit memory of spoken words as stemming from a memory system devoted to pattern recognition (Schacter & Church, 1992) is strongly supported by the findings that speech perception operations produce an "activity trace",

which is informationally richer than the episodic memories that sustain priming. Whether these speaker-dependent episodic memories are the product of a memory system that is resistant to the modulation of higher-order strategic processing is investigated in the following experiment in which the role of expectations regarding voice changes is examined.

Experiment 2

The Effect of Voice-Sequence Predictability on Priming in a Multiple-Speaker Acoustic Environment

Experiment 1 replicated the findings of recent investigations that have shown that auditory repetition priming is impaired by study-to-test voice changes, such as those occurring when a word is spoken by a female at study and then by a male at test (Goldinger, 1992). This impairment in repetition priming has been interpreted as evidence that voice information is a component of the implicit memory representations of spoken words. The notion of an acoustic component to priming is consistent with the idea that speech perception entails the encoding of voice information, in addition to the processes that extract from the acoustic stimulus the abstract-symbolic information that listeners identify as phonemic sequences (Goldinger, et al., 1991). The encoding of voice information can be said to be automatic in the sense that it is initiated without deliberate attention to the voice characteristics of the speech stimulus and is executed without awareness (Palmeri, et al., 1993). However, it is not automatic in the sense that it can be performed without interfering with other concurrent operations. Mullenix and Pisoni (1990) demonstrated that in a multiple-speaker environment, the perceptual operations devoted to the encoding of voice characteristics interfere with those necessary to extract phonemic information from acoustic stimuli (Mullenix & Pisoni, 1990). The detrimental effect that speaker variability has on speech perception coincides with a similarly adverse effect on implicit memory. Goldinger (1992) showed that in a word identification task used as implicit memory procedure, auditory repetition priming decreased as a function of

speaker variability. Experiment 1 replicated Goldinger's finding in a more restricted acoustic environment.

Although intentionality and awareness are not necessary conditions for the encoding of a speaker's voice, it is unclear whether the operations devoted to the use of episodic memory traces are automatic in the sense that they are unaffected by control processes, such as expectancies, hypothesis generation, or any type of response strategies (see Norman & Shallice, 1986, and Bargh, 1992, for additional qualifications regarding automaticity). This experiment is primarily concerned with assessing whether the detrimental effect that speaker variability has on implicit memory can be attenuated by increasing the predictability of the acoustic changes produced by different speakers. Predictability diminishes uncertainty, which is determined by the number of different predictions that can be generated (Peterson & Pitz, 1988). Palmeri, Lively, and Pisoni (1991) manipulated voice predictability in a four-speaker acoustic environment by presenting stimuli in a fixed sequence where each stimulus could be spoken either by any of the four voices (uncorrelated condition) or by only a given voice (correlated condition). The stimuli were four digits arranged in a repeating sequence (e.g., 3-1-2-4-2-1-3-2-1-4). Palmeri, Lively, and Pisoni (1991) showed that implicit memory of the repeating sequence did not benefit from the pairing of each digit with a given voice. It is unclear whether implicit memory of the individual digits would benefit from the sequential order of voices irrespective of the pairings. Memory of stimulus sequence as opposed to individual stimuli adds to the representations of each stimulus the memory of their co-occurrence in the physical structure of the acquisition list, such as in associate learning (Humphreys, 1976).

This added information may make higher-order relational information more relevant and thus render voice either redundant when voice-stimulus pairings exist or superfluous when they do not. In fact, although a multiple-speaker acoustic environment has a detrimental effect on implicit memory for individual stimuli such as spoken words (Goldinger, 1992), no difference is detected between a multiple-speaker and a single-speaker acoustic environment in implicit memory of a repeating stimulus sequence (see Palmeri, Lively, & Pisoni, 1991). It follows that in a priming paradigm, if the listener, at test, can predict not only the possible values of the dimension "speaker's voice" (e.g., the voices utilized), but also their occurrence, the interference that a multiple-speaker acoustic environment may have on the re-activation of the episodic traces entailing voice information might be reduced. As a consequence, the detrimental impact of speaker variability on auditory priming would be lessened.

To investigate the effects of voice predictability on implicit memory of spoken words, the present investigation utilizes a phoneme monitoring task both at study and at test. The basic assumption is that phoneme monitoring will emphasize the disruptive effect that speaker variability has on auditory repetition priming (Mullenix & Pisoni, 1990). Speaker variability is operationally defined as voice changes between contiguous words. At study, participants are presented with a set of words (studied). At test, these same words, repeated in the same voice, are intermixed with other words (non-studied). Latency to respond to word-initial phonemes is recorded. Auditory repetition priming is indexed by faster response latency to studied than to non-studied words.

It is first hypothesized that a decrement in auditory repetition priming would be

observed in a four-speaker acoustic environment relative to a single-speaker. In a single-speaker acoustic environment (one-speaker baseline condition), studied and non-studied words are uttered by only one speaker. Memories of the studied words along with the voice that utters them should develop at study and be implicitly used at test. In a multiple-speaker acoustic environment, both at study and at test, words are quasi-randomly selected in order to avoid having two contiguous words spoken in the same voice (four-speaker baseline condition). Experiment 1 demonstrated that speaker variability disrupts the processes that produce and use these memories. As a result, in this condition, the episodic memory traces should not be able to support priming.

It is then hypothesized that if voice predictability mitigates the effect of speaker variability, an improvement in priming effect would be detected in a four-speaker acoustic environment (four-speaker predictable sequence condition). At study, words are quasi-randomly selected so as to have voice changes between contiguous words following a fixed sequence (e.g., speaker [1], [2], [3] and [4]). Thus, memories of the words and voices used in the study phase should develop along with the knowledge of the order in which the speakers are presented. At test, studied and non-studied words are quasi-randomly selected so as to preserve the voice sequence of the study phase. It is hypothesized that prior to the test phase, the implicit knowledge of the voices utilized at study and their sequential order should generate an expectation regarding how the voices of the test phase will be ordered. If auditory repetition priming is the behavioral expression of a memory system that functions as a pattern recognition module, voice-sequence predictability should not improve priming relative to the four-speaker baseline condition.

In the four-speaker predictable sequence condition, however, any improvement of priming may indicate an effect of ease of processing either at study or at test, rather than the mitigating effect of voice predictability. It is reasonable to hypothesize that, at study, the stability of acoustic changes between contiguous words would reduce the interference between the encoding of voice and the processing of phonemic information and thus contribute to a more robust encoding of both types of information. Similarly, the stability of acoustic changes between contiguous words at test may minimize the interference that a multiple-speaker environment is known to have on speech perception and simply facilitate the emergence of auditory repetition priming. Even if Goldinger's findings (1992) have demonstrated the absence of inhibitory processes in the use of implicit memories in a multiple-speaker environment, the notion of interference between encoding of voice and phonemic information demonstrated by Mullenix and Pisoni (1990) may still apply to the test phase. To rule out these alternative explanations for the hypothesized effects of voice predictability, two conditions are examined in which a fixed voice sequence is presented either at study or at test (fixed voice sequence at study and fixed voice sequence at test conditions).

Method

Participants

One hundred and sixty students from the College volunteered. Thirty-two participants served in each between-subjects condition. They reported English as their first language and no history of speech or hearing disorder at the time of testing.

Stimuli and Apparatus

The stimuli and the apparatus of Experiment 1 were employed.

Procedure

Participants were exposed to the same general procedure employed in Experiment 1, which entails preliminary training, study, distractor, and test. The experimental conditions, however, were different except for the replication of the one- and four-speaker baseline conditions (1-B and 4-B)necessary to obtain a measure of the effect of speaker variability. In the one-speaker condition, stimuli of both study and test lists were randomly presented and speakers' voices were counterbalanced across participants. In the four-speaker condition, stimuli were presented in a quasi-random order, so that contiguous words were never spoken by the same speaker.

To estimate the effect of voice-sequence predictability on auditory repetition priming, three additional four-speaker conditions were implemented. In the predictable sequence condition (4-P), words were quasi-randomly selected in order to present the four voices in a fixed sequential order both at study and at test. To separate the effects of voice-sequence predictability from ease of encoding at study, a condition was devised in which voices were presented in a fixed sequential order at study, but not at test (4-FS). Similarly,

to rule out ease of processing at test, a condition was implemented in which the fixed sequential order of voices was employed at test only (4-FT). A Latin square design was utilized to counterbalance voice sequences.

In the multiple-speaker conditions, the voice assigned to each word was counterbalanced across participants within speaker conditions so that all the words were uttered equally often by each of the four speakers. In all the conditions, studied words were repeated at test in the same voice. The speakers used in preliminary training were always different from those utilized in the study and test phases. The experiment comprised the following factors: Priming (studied and non-studied: within-subjects) and Speaker Condition (1-B, 4-B, 4-P, 4-FS, and 4-FT).

Results and Discussion

Response latencies were employed as the main dependent variable (see Table 2). Latencies two standard deviations above or below each participant's mean response latency were omitted (4.26 %). The operational definition of auditory priming as significantly faster responses to studied than to non-studied words guided the following analyses and interpretations of the data.

Table 2
Mean Latencies of Correct Identifications, Mean Percent Correct, and Standard Deviations as a Function of Speaker Condition and Priming

Speaker Condition	Priming	Response Latency		Priming Score	% Correct
		Mean	SD		Mean
1-B	Non-Studied	701	108	+35	98.67
	Studied	666	91		98.94
4-B	Non-Studied	715	132	+3	97.53
	Studied	712	140		98.44
4-P	Non-Studied	715	151	+15	98.05
	Studied	700	154		98.83
4-FS	Non-Studied	724	129	+15	97.53
	Studied	709	124		98.70
4-FT	Non-Studied	716	164	+ 9	98.18
	Studied	707	153		98.31

To determine whether priming varied as a function of the acoustic environment, a priori pairwise comparisons were conducted on the response latencies of studied and non-studied words at each level of the Speaker Condition factor ($p < .05$). The hypothesis that acoustic variability has a detrimental effect on auditory repetition priming was supported by the finding that a change from one to four speakers eliminated priming in the baseline conditions (1-B and 4-B). In fact, while the one-speaker acoustic environment produced a significant priming effect of 35 msecs, the four-speaker acoustic environment displayed a nonsignificant 3 msecs difference. The hypothesis that voice-sequence predictability lessens the effect of acoustic variability was not supported. Although a significant priming effect of 15 msecs was displayed by the condition in which four speakers were heard in a fixed sequential order both at study and at test (4-P), the condition in which the speakers were heard in a fixed order at study (4-S) also exhibited a significant 15-msec priming effect. When a fixed voice sequence was implemented at test only (4-T), however, a nonsignificant 9 msecs was observed. Taken together, these results suggest that ease of encoding at study attenuates the impairment that acoustic variability has on phonemic processing and thus on auditory repetition priming.

The priming scores (non-studied minus studied) of the conditions that produced a significant priming effect (1-B, 4-P, and 4-S) were submitted to a one-way ANOVA with Speaker Condition as factor. Priming did not vary as a function of the speaker condition, $E(2, 93) = 2.71$, $p < .07$.

As no specific expectations regarding accuracy had been formulated, the percentage accuracy data were submitted to a 2 X 5 ANOVA with Priming and Speaker Condition as

factors (see Kirk, 1982). Priming was defined as significantly higher accuracy for studied words than for non-studied. A significant effect of Priming was found, $E(1, 155) = 6.88$, $p < .010$ (the other Es < 1). Although the ability to correctly identify a word-initial phoneme was not affected by the speaker manipulation, it varied as a function of repetition. Responses to studied words were on the average 1% more accurate than to non-studied. The lack of a significant interaction between Priming and Speaker Condition suggests that accuracy may not be sufficiently sensitive to speaker variability (Mullenix & Pisoni, 1990).

In summary, Experiment 1 demonstrated that the switch from one to four speakers has a detrimental effect on auditory repetition priming. The present experiment replicated this finding and further supported the conclusion that a multiple-speaker acoustic environment produces interference between the activation of preexisting phonemic sequences and the generation of the auditory representations of a speaker's voice. According to Goldinger (1992) this interference diminishes the encoding of voice information. As a result, auditory repetition priming is not adequately supported by the implicit memory traces constructed during study.

A fixed voice sequence at study is sufficient to minimize this effect. It is reasonable to hypothesize that a more robust encoding of voice and possibly of word information can be performed when fixed voice changes occur at study. Once robust structural traces are developed, acoustic variability at test does not have a significant effect on the re-activation of these traces. Similarly, the re-activation of structural traces containing voice information does not appear to be dependent on controlled processes, such as hypothesis generation and

expectations. This latter finding strongly supports the idea that implicit memory of spoken words relies on a memory system that functions as a module (Fodor, 1983). One of the primary properties of a module is its resistance to "cognitive penetrability". Its functioning is automatic in the sense that the module operates as an informationally-encapsulated device that is unyielding to strategic processing such as expectations. However, it is sensitive to the diminished acoustic variability of the study session. This finding is predicted by the modular model. In fact, if the activity traces of speech perception are weakened by word-to-word acoustic changes, there is no reason for their corresponding structural traces not to be similarly impaired.

Experiment 3

The Effects of Lexical Status and Speaker Familiarity on Priming of Words and Pseudowords

The present experiment further examines the hypothesis, supported by the previous two experiments, that implicit memory for spoken words depends on episodic speaker-specific memory representations, which, although deprived of their lexical content, are enriched by voice-related acoustic details. Recent investigations (Church & Schacter, 1994; Schacter & Church, 1992) have shown that auditory repetition priming is not affected by study task instructions that are intended to direct speech perception operations either to the lexical level (e.g., semantic ratings) or to the pre-lexical level (e.g., pitch rating). On the other hand, study-to-test voice changes, such as a word spoken by a female voice at study and by a male voice at test, generally have a detrimental effect on priming. Church and Schacter (1994) have interpreted these findings as support for the proposal that auditory repetition priming results from a cortically-based memory system exclusively devoted to the processing of voice and phonological information. They have concluded that the basic function performed by this memory system is pattern recognition in the auditory domain. The pattern-recognition operations hypothesized to be performed by this system are pre-lexical processes, which are insensitive to word-level properties, such as word frequency and lexical status. Lexical status refers to the distinction between a word and a pronounceable pseudoword (Pisoni, et al., 1985).

These investigations, however, have employed, as implicit memory tests, word identification and stem completion tasks, which are sensitive to voice processing, but

require participants to respond to speech stimuli as whole words. This response bias toward the word-level is likely to render lexical processing an unwanted byproduct of the memory procedures rather than an expression of the actual workings of the pertinent memory system (Lively, 1994; Weldon, 1991). Word-level implicit memory procedures are thus unable to adequately test the hypothesis that auditory repetition priming results from pattern-recognition operations which produce pre-lexical memory representations including voice and phonological information at the subword level.

In the present experiment, a procedure devoid of the word-level response bias, phoneme monitoring, is employed both at study and at test. Under certain processing-load conditions, phoneme monitoring can preserve lexical processing in speech perception (Eimas, Hornstein, & Payton, 1990; Eimas & Nygaard, 1992; Rubin, Turvey, & Van Gelder, 1976) and emphasize, at the same time, the encoding of voice information (see Experiment 1). Phoneme monitoring, by eliminating the lexically-based behavioral constraint of word-level procedures, thus provides a unique experimental context for testing the hypothesis that auditory repetition priming is based on pre-lexical representations. Experiment 1 provided support for this hypothesis by showing that although speech perception is sensitive to word frequency, priming is not. In this experiment the same prediction is tested by assessing the potential effects of lexical status on speech perception and memory. Lexical status is manipulated by employing words, which are meaningful utterances, along with pronounceable pseudowords, which are utterances devoid of meaning. It is assumed that if phoneme monitoring latencies are sensitive to the stimuli's lexical status, speech perception involves lexical operations. On

the basis of the findings of Experiment 1, even if speech perception entails lexical information, auditory repetition priming is not expected to vary as a function of lexical status.

Experiment 1 showed that with a memory test that emphasizes subword-phonemic information, auditory operations such as the encoding of acoustic voice characteristics are maximized. Acoustic-voice information is thus expected to be an essential component of the representations that sustain auditory repetition priming. If voice information is a substantial component of priming, a voice change between study and test will eliminate acoustic repetition and consequently abolish priming. In the present experiment, to further emphasize the encoding of acoustic voice characteristics, one speaker is employed at study. At test, a two-speaker acoustic environment is used. Same-speaker conditions entail studied words produced by the speaker of the study phase. Contrary to Experiment 1, non-studied words are always produced by a novel speaker. Thus, rather than having two voices previously heard simply switched at test, the potential effect of prior exposure to a voice is eliminated from non-studied words. Different-speaker conditions involve two novel speakers. Thus, a more undiluted measure of priming due to voice and phonemic repetition in same-speaker conditions can be contrasted with priming due exclusively to phonemic repetition.

If the contribution of voice to implicit memory of spoken words is robust, voice familiarity will improve priming. In fact, repeated exposure to a voice has been shown to improve word identification (Nygaard, et al., 1994) and explicit memory (Lightfoot, 1989). In the latter case, explicit memory for spoken words was improved through explicit

encoding of voice. The hypothesis that the beneficial effect of voice familiarity on memory can be easily replicated without explicit encoding of voice information appears to be plausible (Palmeri, et al., 1993). This experiment investigates whether voice familiarity, obtained through the implicit encoding of voice characteristics, improves implicit memory of a set of stimuli (e.g., spoken words and pseudowords).

In summary, the present experiment will first re-assess whether attention to word-initial phonemes induces priming. The contribution of voice information to auditory priming will then be estimated in a context in which a speaker's voice is either unfamiliar or familiar to the listener. Lastly, the manipulation of lexical status in the context of an implicit task will assess whether the operations performed in speech perception, indexed by decision latencies, are mirrored in the memory representations that support priming.

Method

Participants

One hundred and twenty-eight students from the College volunteered. Thirty-two additional participants were utilized in the voice familiarization pilot study. They reported English as their first language, and no history of speech or hearing disorder. Thirty-two participants were assigned to each of the between-subjects conditions.

Stimuli and Apparatus

The stimuli consisted of 24 words (mean frequency = 118, median = 95, Francis and Kucera, 1982) and 24 pronounceable pseudowords (e.g., BOPE). Half of the words and pseudowords began with b, and half began with either t, d, or p (4 for each phoneme). The stimuli were arranged in two study lists of 24 stimuli containing 12 words and 12 pseudowords matched for initial phoneme. The words in the two study lists were also matched for frequency. The test list contained all the 48 stimuli, so that half were previously studied and half were not.

The 48 stimuli were selected from a larger pool of 120 monosyllabic stimuli entailing a set of 60 words and a matched set of 60 pronounceable pseudowords. The word/pseudoword distinction was based on a change in the final phoneme of a word. Each set comprised 30 stimuli beginning with the phoneme b and 30 with either the phoneme t, d, or p.

Each of the 48 stimuli was recorded by four speakers, two males and two females. These four speakers also recorded 72 similar stimuli to be used for preliminary training. Study lists consisted of stimuli spoken by one speaker. Test lists entailed two speakers:

a male and a female. Each stimulus was digitized at 11-kHz sampling rate by using a 16-bit analog-to-digital converter equipped with anti-aliasing filters and installed on an IBM-compatible 486 personal computer. A signal processing package was utilized to create a digital file for each stimulus, mean length = 484 msecs (485 msecs and 481 msecs for words and pseudowords, respectively). Stimuli were rerecorded until two independent observers rated them to be intelligible and clear. The presentation of the stimuli, filtered at 4.8 kHz, on calibrated Koss TNT/77 headphones at 75 dB and the recording of the participants' responses were accomplished by a custom-made program.

Procedure and Design

The experiment entailed preliminary training, study phase, distractor task, and test phase (see Experiment 1 for additional qualifications). To estimate the contribution of voice information to auditory priming, same-speaker conditions in which studied stimuli were repeated at test in the same voice were matched with their corresponding different-speaker conditions in which a gender voice change was introduced between study and test. In the *same-speaker baseline* condition, at test, studied stimuli were produced by the speaker of the study phase and non-studied stimuli were produced by a novel speaker. Therefore, while studied stimuli benefitted from voice and word repetition, no repetition was involved in non-studied. Any priming effects in this condition could be attributed to the combined effect of voice and word repetition. In the *different-speaker baseline* condition, the speaker employed in the study phase did not produce studied stimuli at test. Two novel speakers were utilized to produce both studied and non-studied stimuli. The change in voice between study and test always implied a gender change. Any priming

effects could be attributed to word repetition only. In both conditions, the speaker of preliminary training was always different from any of the speakers of the study and test phases.

To assess the facilitatory impact of voice familiarity on auditory priming, the above conditions were replicated with a familiarization procedure utilized as preliminary training. In the *same-familiar speaker* condition, preliminary training involved familiarization with a voice that was subsequently employed to generate the study list. At test, the same voice produced the studied stimuli and a novel voice always produced the non-studied stimuli. In the corresponding *different-familiar speaker* condition, preliminary training involved familiarization with one voice. A novel voice uttered the study list. At test, the familiar voice produced the study stimuli and a novel voice uttered the non-studied stimuli.

In all the conditions, the stimuli were presented in a quasi-random order, such that contiguous stimuli were never in the same voice. The speaker's voice assigned to each stimulus was counterbalanced across participants within speaker conditions so that an equal number of words and pseudowords were spoken by each voice. A 2 X 2 X 2 X 2 X 2 mixed-factorial design was implemented with Priming (studied and non-studied: within-subjects), Voice (same and different: between-subjects), Voice Familiarity (familiar voice and unfamiliar voice: between-subjects), Lexical Status (word and pseudoword: within-subjects), and Target Phoneme (presence of b and absence of b: within-subjects) as factors.

Pilot Study on Voice Familiarization

Prior to running the voice-familiarity conditions, an ancillary study was conducted to obtain a measure of voice familiarity. This study, modeled after Nygaard, et al. (1994),

involved two phases: voice familiarization and test. The stimuli of preliminary training were divided into two sets of 36. When one set, spoken by one of the four speakers, was employed in the familiarization phase, the other set was employed in the test phase. At test, participants (none subsequently utilized in the experiment) heard 36 stimuli spoken either in the same voice as the one employed in the voice familiarization phase or in a different gender voice. Speakers and study lists were counterbalanced across participants. In both phases, phoneme monitoring was performed. Voice familiarity was operationally defined as the facilitatory effect of prior exposure to a voice on the latency to detect a word-initial phoneme in novel stimuli.

This pilot study suggested that if the familiarization set was repeated 5 times, faster response latencies to the same-voice set than to the different-voice set would be obtained at test. Thirty-two participants, randomly assigned to one of the between-subjects conditions, were exposed to 180 (36 X 5) exposures to a voice. The results showed that, at test, the familiar voice condition produced faster response latencies than the different voice condition, $t(30) = 2.49$, $p < .019$, SEdm = 45.40. This familiarization procedure was applied to the *same-familiar speaker* and *different-familiar speaker* conditions.

Results and Discussion

Response latencies to stimulus-initial phonemes were utilized as the main dependent variable (see Table 3). Latencies two standard deviations above or below each participant's mean response latency were omitted (4. 90 %). The operational definition of auditory priming as significantly faster responses to studied than to non-studied words guided the following analyses and interpretations of the data.

Table 3
Mean Latencies and Standard Deviations of Correct Identifications and Mean Percent Correct in the Study and Test Phases as a Function of Speaker Condition, Priming, Lexical Status

	Response Latency					% Correct	
	Word		Pseudoword			Word	Pseudoword
	Mean	SD	Mean	SD	Priming	Mean	Mean
Unfamiliar Voice: Study Phase							
Same Speaker	785	157	836	200		99.22	98.44
Different Speaker	787	172	820	216		98.44	99.48
Unfamiliar Voice: Test Phase							
Same Speaker					+26		
Non-Studied	759	147	786	173		97.66	98.44
Studied	723	133	770	153		97.92	98.96
Different Speaker					- 7		
Non-Studied	737	155	767	188		97.40	98.44
Studied	743	151	775	185		97.66	98.44
Familiar Voice: Study Phase							
Same Speaker	733	149	778	236		98.70	97.66
Different Speaker	779	200	789	227		97.66	95.31
Familiar Voice: Test Phase							
Same Speaker					+ 29		
Non-Studied	750	152	770	182		98.96	99.22
Studied	723	128	739	153		99.74	98.44
Different Speaker					- 9		
Non-Studied	711	163	745	186		97.66	95.58
Studied	720	190	754	218		98.44	95.84

The effect of study-to-test voice changes on priming. To assess the occurrence of priming and its sensitivity to acoustic repetition, a priori pairwise comparisons were conducted on the response latencies of studied and non-studied words ($p < .05$) in the baseline and familiarity conditions. In the same-speaker baseline condition, studied stimuli were on the average 26 msecs faster than non-studied. Similarly, in the same familiar speaker condition, response latencies of studied stimuli were on the average 29 msecs faster than non-studied. The advantage of studied stimuli over non-studied was eliminated in both different-speaker conditions. Taken together, these findings indicate that priming, in a task devoid of the word-level bias, relies heavily on acoustic voice-specific information. When acoustic repetition is eliminated from the studied stimuli, priming does not occur.

The effect of voice familiarity on priming. To determine whether voice familiarity affected priming, priming scores (response latencies of non-studied minus those of studied) of the same-speaker conditions were submitted to a one-way ANOVA with Voice Familiarity as factor. Priming did not benefit from voice familiarity, $F < 1$. There are three possible explanations for this finding. The most parsimonious interpretation is that, in a simple acoustic environment such as that produced by one voice, the contribution of voice information to the implicit memory record of speech perception may have reached an asymptotic level. Alternatively, the presence of one speaker in the familiarization procedure along with the implicit processing of voice encouraged by the phoneme monitoring task may have hindered a fine-grained perceptual encoding of voice information. In fact, studies that have found an effect of voice familiarity on word identification and explicit memory employed several speakers and an explicit voice

identification procedure in the familiarization training (Lightfoot, 1989; Nygaard, et al., 1994; Nygaard & Pisoni, 1995). The use of a varied acoustic environment along with explicit voice-identification training may elicit fine perceptual details for each speaker's voice which then may magnify the difference between these voices and novel voices. In a one-voice training procedure, this reasoning suggests that although the voice information that comprises the representations that sustain auditory priming has been repeatedly activated during training, it may not contain a more complex or refined set of characteristics that makes that voice qualitatively different from the record of that same voice without the familiarization training. If the effect of voice familiarity on implicit memory is due to the complexity of the voice record rather than to magnitude of prior activation, then a one-speaker environment may not augment the contribution of voice information to priming. The third interpretation for the lack of an effect of voice familiarity on priming obtained in the present experiment relies on the assumption that repeated exposure to a voice produces a memory trace for that voice generalized over all the instances previously heard. Then the memory of the acoustic details of that voice would be a separate memory from the episodic memory of a stimulus constructed during study. Memory for voices would lead to processing voices as gestalts and thus generalize over the individual instances of the study phase. Memory for words would lead to analytic processing of each instance. This interpretation is suggested by a study on the processing of familiar sentences. Kempler and Van Lancker (1987) argued that familiar sentences are processed as gestalts rather than analytically. The voice gestalt and the memory of a given studied stimulus may be functionally independent in the sense that they do not concurrently

contribute to priming. However, voice gestalts may aid word identification (Nygaard, et al., 1994). In fact, in word identification, a representation of a voice, irrespective of the instances that produced that representation, is likely to diminish the burden placed upon the processes of voice and phonemic encoding. This explanation, however, is inconsistent with Lightfoot's findings that voice familiarity aids explicit memory for spoken words.

The effect of lexical status on speech perception and priming. To assess whether speech perception processes occurred at the pre-lexical or lexical level, at study and at test, the response latencies of the same speaker conditions, which were the only conditions that displayed significant priming effects, were submitted to a 2 X 2 ANOVA with Lexical Status and Target Phoneme as factors. The response latencies at study displayed a main effect of both Lexical Status and Target Phoneme, $F(1,63) = 16.69$, $p < .001$, $MSe = 8958.70$, and $F(1,63) = 23.46$, $p < .001$, $MSe = 6854.75$, respectively, but not an interaction, $F < 1$. The analysis of response latencies at test displayed a main effect of Lexical Status and Target Phoneme, $F(1, 63) = 13.40$, $p < .001$, $MSe = 3537.73$, and $F(1, 63) = 46.70$, $p < .0001$, $MSe = 3788.60$, respectively, but not a significant interaction, $F = 2.54$. In both cases, response latencies for words were faster than those for pseudowords. Similarly, responses to stimuli beginning with the phoneme b were faster than those to stimuli beginning with another stop consonant. The sensitivity of response latency to lexical status (Rubin, Turvey, & Van Gelder, 1976) indicated that although the task required participants to respond to stimulus-initial phonemes, speech perception involved lexical analysis.

To estimate the level of processing upon which priming relied, the same analysis

was performed on the priming scores (response latency of studied stimuli minus those of non-studied) of the same speaker conditions. No significant results were found (Es < 1.89). The absence of an effect of Lexical Status on priming suggest that the representations upon which priming relies are pre-lexical. Taken together, these results indicate that regardless of the depth to which stimuli are processed in speech perception, priming emerges as a pre-lexical phenomenon.

Accuracy. A priori pairwise comparisons were performed on the mean percentage accuracy of the studied and non-studied words in the baseline and familiarity conditions. No significant results were observed ($p < .05$). The absence of priming effects in the analyses of the mean percent accuracy depleted further analyses of any theoretical justification. Thus, no further analyses are reported on this dependent measure.

General Discussion

In summary, the present experiment demonstrated that, although speech perception operations guided by the task of the study and test phases entail lexical processing, auditory repetition priming results from pre-lexical operations. The sensitivity of priming to voice changes between study and test indicates that voice information is one of the components of implicit memory of spoken words. Its insensitivity to lexical status suggests that the phonological component of priming reside at the subword level. In a simple encoding environment of one voice, voice familiarity apparently does not augment priming. The most parsimonious explanation for this null finding is that, in this environment, the contribution of voice to priming rapidly reaches an asymptotic level. An alternative explanation assumes that voice familiarity leads to the encoding of a detailed

representation of a speaker's voice rather than simply increasing the level of activation of the record of that voice. A one-voice familiarization procedure along with explicit voice-encoding instructions may not encourage this fine-grained encoding and thus voice familiarity does not augment priming. Another interpretation for the absence of an effect of voice familiarity on priming is based on the assumption that repeated exposure to a voice produces a memory of the acoustic details of that voice that is functionally independent from the episodic memory of a stimulus constructed during study. This interpretation is questionable because it applies to the findings on word identification (Nygaard, et al., 1994; Nygaard & Pisoni, 1995), but it does not explain why explicit memory is facilitated by voice familiarity (Lightfoot, 1989).

Although the present experiment is unable to cull these explanations of the mechanisms responsible for the lack of an effect of voice familiarity on implicit memory, it has been successful in demonstrating that auditory repetition priming is the product of pre-lexical processes, which blend subword phonological and speakers' voice information into a unique memory trace. These findings clearly support the theoretical approach conventionally referred to as the *multiple-memory system* model of implicit memory (Church & Schacter, 1994; Schacter, 1987; Schacter & Church, 1992). Implicit memory of spoken words draws upon a memory system devoted to pattern recognition in the auditory domain. This system functions as a module (Fodor, 1983; Moscovitch, 1992) in the sense that the richness of the activity traces produced by speech perception operations is deprived of semantic connotation, but preserves acoustic details. Contrary to the predictions of *abstractionist* and *modification* models of implicit memory phenomena (Carr

& Pollatsek, 1985; Graf & Mandler, 1984; Mandler, 1980; Morton, 1979), priming is not simply due to passive activation processes of preexisting memory representations (see Bowers, 1994). The memory system responsible for auditory repetition priming produces a unique memory for each spoken word and thus appears to combine acquisition processes of voice information with processes of activation of preexisting phonological representations at the subword level. Although the present findings strengthen an instance-specific approach to memory, the similar effect of study-to-test voice changes on pseudowords and words does not support the prediction of Hintzman's model (1986) that the more unique a repeated stimulus is, the more likely it will be to activate its detailed trace rather than a collection of traces.

Chapter IV: General Discussion

Introduction, Findings and Comparisons

The predictions of the modular model: A reminder. The main purpose of this investigation was to examine the predictions that result from a modular model of implicit memory for spoken words. The modular model, here presented, is a specification of the *multiple-memory system* conceptualization of implicit memory phenomena (see Schacter & Church, 1992). At the core of this model is the distinction between speech perception operations, which produce activity traces, and implicit memory precesses, which are responsible for structural traces (see Hebb, 1961; Miller, Kasprow, Schachtman, 1986). Activity traces can be thought of as temporary representations that result from the encoding operations used to perceive spoken words. Structural traces, on the other hand, can be conceptualized as long-lasting memory representations, which aid speech perception at test.

The model postulates that the transfer from activity traces to structural traces (e.g., the act of retention) is performed by a memory system characterized by domain specificity (i.e., it regards only a restricted range of information) and informational encapsulation (e.g., it is resistant to the effect of higher-order cognitive processes such as expectations and hypothesis generation). On the basis of these properties, the memory system selectively excludes from retention information that it is unable to process. If speech perception operations have produced activity traces that contain lexical information, this system, devoted to pattern recognition, will reject lexical information. Thus, activity and structural traces are not necessarily isomorphic. They both preserve, however, an

instance-specific episodic record of a perceptual experience. Voice information is responsible for this specificity. In light of the instance-specific nature of the modular model, the speech perception processes that are responsible for activity traces are assumed to combine processes of generation of internal representations of a speaker's voice and processes of activation of preexisting abstract-symbolic information.

The model assumes that repetition priming occurs when, during test, an activity trace of speech perception matches a structural trace. Since the structural trace contains pre-lexical information such as voice and phonemic sequences at the subword level, the match will be based on this information only. The magnitude of auditory repetition priming will depend on the degree of match of voice and phonemic information at the subword level between a test stimulus and its structural trace.

Although the modular denotation of the proposed model is primarily based on the postulation of an information reduction mechanism that, by operating from activity traces to structural traces, discards lexical information, the pattern-recognition function that this model attributes to implicit memory heavily relies on the assumption of voice-specific structural traces. The encoding of voice information performed by perceptual processes and its retention by memory processes is thus inherent to the postulation of the pattern-recognition function. The assumption of voice-specificity for activity and structural traces leads the model to predict that a multiple-speaker acoustic environment, by compelling perceptual processes to encode additional voice information and memory processes to retain this additional information, would, under time constraints, weaken first activity traces and then structural traces. The weakness of the resulting traces would then emerge

behaviorally in a decrement in repetition priming. In light of the postulation of a pattern-recognition memory system informationally encapsulated, that is, resistant to the modulating effect of higher-order cognitive processes, the model predicts that the disruptive effect of the multiple-speaker environment would not be attenuated by the ability of the listener to predict voice changes at test. On the other hand, voice familiarity, obtained through passive repeated exposure to a voice, would facilitate encoding and retention processes and thus strengthen these traces.

The evidence. The present investigation represents an attempt to provide a test for the modular model here proposed. Although much remains to be done, the evidence obtained in the experiments presented in chapter III provides initial support for the model.

As predicted, although speech perception operations were sensitive to word-level factors, such as word frequency and lexical status, in Experiment 1 and Experiment 3, no effect of such factors was observed on auditory priming. Study-to-test voice changes, however, impaired priming. Taken together, these findings portray implicit memory for spoken words as dependent on speaker-specific episodic memory representations, which are to be considered pre-lexical in nature. Experiment 2 supported the postulation of informationally encapsulated memory processes by demonstrating that strategic operations, such as the ability to predict the order in which voice changes can occur at test, did not affect priming. Experiment 1 showed that a multiple-speaker acoustic environment weakens the structural traces that embody implicit memory. In Experiment 2, the assumption that this impairment is due to the disruptive effect of acoustic variability at study rather than during retrieval, was supported by the finding that a diminished acoustic

variability at study only, obtained through stable word-to-word voice changes, improved priming. The absence of an effect of voice familiarity, however, questioned the assumption that voice familiarity simply strengthens the structural traces that embody implicit memory for spoken words. This finding was first interpreted as due to the restricted acoustic environment of one speaker employed at study. In this environment, the encoding of voice may reach an asymptotic level independently of the amount of prior exposure to that voice. Alternatively, if voice familiarity improves the encoding of voice by developing a more refined set of voice characteristics rather than a more robust trace for these characteristics, a one-speaker acoustic environment may not lead to such a detailed encoding and thus voice familiarity has no effect on priming. Lastly, independently of the one-speaker acoustic environment at study, if voice familiarity emphasizes the development of a memory of a speaker's voice generalized over several words, this memory may not contribute to the strength of the representations of the individual words spoken by that speaker.

The search for the appropriate experimental procedure as a means for testing memory models. The findings presented above not only raise several objections about current memory models, but they also question the procedures employed to test these models. The present investigation began with an attempt to derive from the available evidence a defined set of predictions for the *multiple-memory system* model of implicit memory for spoken words so that it could be adequately compared with the *transfer-appropriate processing* model. Experimental findings, however, were found to be closely intertwined with the procedures employed to obtain those findings so that the testing of the models was limited

by the procedural choices made by each investigator. An exemplification of this difficulty is the claim, made by the proponents of the *multiple-system* model that implicit memory for spoken words results from a memory system devoted to pattern recognition (Schacter & Church, 1992). Schacter and Church used word-level memory procedures and focused on encoding task manipulations as a means of investigating their claim without first assessing whether encoding manipulations were indeed indicative of a processing change at study and without acknowledging that word-level memory procedures prevent a test of their claim.

It was then clear that although modeling and subsequent testing is an interesting and valuable undertaking, such effort should always be preceded by a detailed and comprehensive task analysis. The present investigation attempted to overcome the problems encountered in Schacter and Church's study (1992) by selecting an experimental procedure that allowed for the test of the assumption that implicit memory for spoken words derives from pattern-recognition operations. Speech perception operations, both at study and at test were made sensitive to lexical factors, such as word frequency and lexical status, without constraining participants to process the stimuli as whole words. Under this experimental setting, the *multiple-memory system* and the *transfer-appropriate processing* models make contrasting predictions. The former hypothesizes that although speech perception may involve lexical processing, priming is based on pre-lexical representations. The latter hypothesizes that if speech perception operations entail lexical processing so will priming. The evidence obtained in Experiment 1 and Experiment 3 permitted an evaluation of these predictions and supported the *multiple-memory system* model.

Other models and evidence. The present findings, along with earlier research, question theories of information processing that have disregarded voice as a component of the activity traces constructed by speech perception. The present evidence also questions theories that have not embodied in their models a pre-lexical stage of processing (Cutler, Mehler, Norris, & Segui, 1986; Mehler, Dommergues, Frauenfelder, & Segui, 1981; Segui, Dupoux, & Mehler, 1990; Mehler, Dupoux, & Segui, 1990), which appears to be necessary to account for the pattern recognition operations involved in implicit memory of spoken words. For an idea of how this neglect of voice information and of a pre-lexical stage of processing has been expressed in the modeling of speech perception operations that entail lexical activation, a brief summary of some of the more relevant models will be presented. Note that the majority of the theories of word identification presented below do not account for the processes that lead from an acoustic stimulus to phonemes.

Morton's logogen model (1969, 1979, 1982) is unable to account for the encoding of voice information. This model assumes that "logogens" are monitoring devices that contain the phonemic, orthographic, syntactic and semantic makeup of a word. Logogens are activated by incoming stimuli. A word is identified when sufficient evidence is collected for the incoming stimulus, so that a logogen crosses its activation threshold. Because the information contained in the logogen is abstract in the sense that it is a generalization over many encounters with a word, there is no way to account for instance-specific effects such as those observed with study-to-test voice changes. An isolated pre-lexical level of processing is also not accommodated by this model.

A similar argument applies to Forster's search model (1976, 1979). Although this

model embodies a modular approach in which word identification occurs hierarchically, it assumes abstract representations in which perceptual details are not taken into account. Although the model postulates a modular arrangement of processing stages, it does not entail a separate module for pre-lexical information in word recognition. In fact, an incoming stimulus is first processed by a peripheral system in which a phonemic representation for that word is obtained through a word frequency-ordered search and matched with a given preexisting representation.

The cohort model (Marslen-Wilson, 1975, 1987, 1989, 1990; Marslen-Wilson & Tyler, 1975, 1980; Marslen-Wilson & Welsh, 1978) assumes that phonemic information activates a set of preexisting memory representations of whole words in a left-to-right fashion so that the sequential nature of speech stimuli is acknowledged. Once the initial cohort of possible word candidates is activated the model assumes that a word will be "identified" when sufficient information regarding phonemic sequences and higher-order information isolates that word from all the possible alternative candidates in memory.

The cohort model does not account for the encoding of voice information and its facilitatory effect on word identification. It is potentially able, however, to account for pre-lexical pattern recognition processes, if the buildup of information prior to lexical access is isolated as an intermediate and separate stage of processing. In fact, the model postulates two stages for word recognition, the first of which, driven by pre-lexical information, allows for the activation of a series of words candidates (e.g., what Marslen-Wilson calls "access"). This stage could be broken down into two separate substages. The first would process subword patterns produced by the left-to-right incoming signal and by

the changes that each phoneme undergoes following coarticulation. These subword patterns will then be used to select an initial set of lexical candidates in the second stage of word identification.

The Neighborhood Activation Model (NAM) assumes that a spoken word activates a set of acoustic-phonetic patterns, which, in turn, activates a set of word candidates (Luce, Pisoni, Goldinger, 1990). These word candidates then monitor the pattern of activation at the phonetic level. Word recognition occurs when a word is selected to be the best match for the stimulus word. As in previous models, word units in NAM do not contain voice information. This model is potentially able to account for voice effects by modifying its assumptions relative to the acoustic-phonetic patterns activated by an incoming stimulus so that voice information can be encoded and stored in memory. Similarly, acoustic-phonetic patterns can be re-conceptualized as defining an intermediate pre-lexical state of processing.

TRACE (Elman & McClelland, 1986; McClelland & Elman, 1986) comprises three levels of representations: phonetic features, phonemes, phonemic sequences at the word level. The model is a network comprised of nodes for each of these levels. Incoming information is represented as a set of activated phonemic features. As the connections between levels are bidirectional except for the connection between phonetic features and phonemes, word identification occurs with the aid of higher-order word information. The model, however, does not have a mechanism to account for the encoding of voice information. Because of the distributed nature of the representational assumptions, however, TRACE is potentially able to deal with voice characteristics. A change in the

representational assumptions of the first level of processing units would solve the problem. The first two levels of processing can be thought of as representing pattern recognition operations. To account for equal implicit memory for words and pseudowords, a change in the representational assumptions between the second and third stage of processing would be required so that subword patterns would precede word activation as an isolated stage of processing.

LAFS (Klatt, 1979) entails an acoustic "lexicon" of speaker-specific spectral templates, which are matched with the spectral representation of an incoming stimulus. No phonemic representations are necessary for word recognition. The spectral templates are similar to "Wickelphones" (Wickelgren, 1969) that represent subword units. Although this model of word recognition is able to account for voice information by encoding acoustic variability in the "lexicon", it does not entail the generation of episodic representations in which voice is a relevant component. In fact, preexisting speaker-specific representations are activated by the match with an incoming stimulus. However, the use of sub-lexical rather than lexical representations is consistent with the idea of pattern recognition processes as devoid of any word-level information.

The apparent disregard that previous models have shown for the encoding of voice characteristics is not always found in current memory models. Some of these models are examined below in light of two critical issues that the present findings suggest: (1) the accuracy of human memory in maintaining detailed structural traces, and (2) the possible modular apparatus that subserves implicit memory for spoken words.

A prototypical example of a memory model that has acknowledged the first issue

is MINERVA 2 (Hintzman, 1986, 1988; Hintzman, Nozawa, & Irmscher, 1982). MINERVA 2 assumes that each stimulus and each repetition of that stimulus generates an episodic trace, which contains perceptual details along with other types of information, including meaning. Although Hintzman's model distinguishes between primary memories and secondary memories (James, 1983), which, in the modular model introduced here, correspond to the activity traces of speech perception and to the structural traces in long-term memory, respectively, his model does not account for the modular nature of the structural traces that this investigation has found to support priming. MINERVA 2 also does not account for equal study-to-test voice changes for pseudowords and words. In fact, it predicts that the uniqueness of a repeated stimulus (e.g., a pseudoword relative to a word), will determine the extent to which the stimulus will activate its detailed trace, rather than a collection of traces. MINERVA 2 can thus account for the episodic speaker-specific nature of implicit memory for spoken words, but it falls short of accounting for the modular nature of this memory.

The general properties of McClelland and Rumelhart's (1985) distributed memory model can also account for the episodic stimulus-specific nature of the representations that this investigation has found to support priming. McClelland and Rumelhart's model, however, encounters the problem of explaining how episodic representations can be preserved after long delays of time and several intervening stimuli. In fact, the network is an adaptable structure that adjusts to incoming stimuli so that the unique pattern of activation for each stimulus is lost as time goes by. However, although not clearly specified, McClelland and Rumelhart's conceptualization of processing modules is

consistent with the evidence obtained in the present investigation regarding pattern-recognition processes.

The present findings, however, are different from those predicted by the hybrid model proposed by Feustel, Shiffrin, and Salasoo (1983) and Salasoo, Shiffrin, and Feustel (1985), which attempted to account for the effects of repetition on words and pseudowords. Feustel, et al. (1983), and Salasoo, et al. (1985) found an advantage of words over pseudowords following one repetition and claimed that this advantage was to be attributed to the fact that words benefit from larger contributions of preexisting memory traces (e.g., lexical information). No similar finding was obtained in the present investigation where a word-level procedure was not employed. The differential effect of one repetition on words and pseudowords described by Feustel, et al. (1983) and Salasoo, et al. (1985) can be accounted for by word-level biases and thus does not violate the pattern-recognition assumption of the modular model. The idea of levels of word activation, from physical features to word, however, negates the modular pre-lexical nature of the processes assumed to be responsible for priming and thus contradicts the findings of the present investigation. Feustel, et al. (1983), and Salasoo, et al. (1985) also suggested that episodic traces contribute to the effect of repetition. These episodic traces, however, are assumed to contain primarily abstract-symbolic characteristics regarding the prior occurrences of a stimulus in the experimental context and to contain only minimal surface information. The modular model predicts that surface information, such as voice, is a defining characteristic of episodic memory representations.

In summary, although the modular model is not always consistent with previous

models of word recognition and memory for spoken words, it represents a modest attempt at identifying the pattern-recognition nature of implicit memory for spoken words. The model presented and tested here, however, is consistent with previous findings of speaker-specific long-term memory representations of spoken words (see Goldinger, Pisoni, & Luce, 1995). Additional investigations, currently in progress, will attempt (1) to specify the role of voice familiarity in implicit memory, (2) to assess whether the modular nature of implicit memory for spoken words can be generalized to visually-presented words, and (3) to see whether the current findings can be replicated under different study conditions.

Bibliography

Allard, F., & Henderson, L. (1975). Physical and name codes in auditory memory: The pursuit of an analogy. *Quarterly Journal of Experimental Psychology, 28*, 475-482.

Assman, P. F., Nearey, T. M., & Hogan, J. T. (1982). Vowel identification: Orthographic, perceptual, and acoustic aspects. *Journal of the Acoustical Society of America, 71*, 975-989.

Bargh, J. A. (1992). The ecology of automaticity: Toward establishing the conditions needed to produce automatic processing effects. *American Journal of Psychology, 105*, 181-199.

Bartholomeus, B. (1974a). Dichotic singer and speaker recognition. *Bulletin of the Psychonomic Society, 2*, 407-408.

Bartholomeus, B. (1974b). Effects of task requirements on ear superiority for sung speech. *Cortex, 10*, 215-223.

Bassili, J. N., Smith, M. C., MacLeod, C. M. (1989). Auditory and visual word stem completion: Separating data-driven and conceptually-driven processes. *Quarterly Journal of Experimental Psychology, 41A*, 439-445.

Bowers, J. S. (1994). Does implicit memory extend to legal and illegal nonwords? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 534-549.

Bradlow, A. R., & Nygaard, L. C. (1996, May). The effect of talker rate and amplitude variation on memory representation of spoken words, *Poster session presented at the annual meeting of the Acoustical Society of America*, Indianapolis, IN.

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.

Brown, J. M., & Fowler, C. A., & Rueckl, J. C. (1996, May). Voice effect in implicit memory tasks, *Poster session presented at the annual meeting of the Acoustical Society of America*, Indianapolis, IN.

Brown, J. S., & Carr, T. H. (1993). Limits on perceptual abstraction in reading: Asymmetric transfer between surface forms differing in typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 1277-1296.

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.

Carr, T. H., & Pollatsek, A. (1985). Recognizing printed words: A look at current models. In K. G. W. D. Besner, & E. M. MacKinnon (Ed.), *Reading research: Advances in theory and practice, Vol. 5* (pp. 1-81). Toronto: Academic Press.

Carterette, E. C., & Barnebey, A. (1975). Recognition memory for voices. In A. C. S. G. Nooteboom (Ed.), *Structure and process in speech perception*. New York: Springer-Verlag.

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

Clarke, F. R., & Becker, R. W. (1969). Comparison of techniques for discriminating among talkers. *Journal of Speech and Hearing Research, 12*, 747-761.

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

Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General, 120*, 235-253.

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

Cooper, W. E., & Sorenson, J. (1981). *Fundamental frequency in sentence production*. New York: Springer.

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

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

Cutler, A., Mehler, J., Norris, D. & Segui J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language, 25*, 385-400.

Dorfman, J. (1994). Sublexical components in implicit memory for novel words. *Journal of Experimental Psychology: Learning , Memory and Cognition, 20*, 1108-1125.

Durso, F. T., & Johnson, M. K. (1979). Facilitation in naming and categorizing repeated pictures and words. *Journal of Experimental Psychology: Human Learning and Memory, 5*, 449-459.

Eimas, P. D., Hornstein, S. B., & Payton, P. (1990). Attention and the role of dual codes in phoneme monitoring. *Journal of Memory and Language, 29*, 160-180.

Eimas, P. D., & Nygaard, L. C. (1992). Contextual coherence and attention in phoneme monitoring. *Journal of Memory and Language, 31*, 375-395.

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

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

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: Bradford/MIT Press.

Forster, K. I. (1976). Accessing the mental lexicon. In R. J. Wales, & E. C. T. Walker (Eds.), *New Approaches to language mechanisms* (pp. 257-287). Amsterdam: North-Holland.

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

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 680-698.

Foss, D. J., & Swinney, D. (1973). On the psychological reality of the phoneme: Perception, identification, and consciousness. *Journal of Verbal Learning and Verbal Behavior, 12*, 246-257

Francis, W., & Kucera, M. (1982). *Frequency Analysis of English Usage: Lexicon and Grammar*. Boston: Houghton Mifflin.

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

Garner, W. R., & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology, 1*, 225-241.

Gazzaniga, M. S. (1989). Organization of the human brain. *Science, 245*, 947-952.

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

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

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

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

Gipson, P. (1986). The production of phonology and auditory priming. *British Journal of Psychology, 77*, 359-375.

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

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

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

Goldinger, S. D., Pisoni, D. B., & Luce, P. A. (1995). Speech perception and spoken word recognition: Research and theory. In N.J. Lass (Ed.), *Principles of experimental phonetics* (pp. 277-327). New York: Mosby.

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.

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., & 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 and Psychophysics, 50,* 524-536.

Greenberg, J. H., & Jenkins, J. J. (1964). Studies in the psychological correlates of the sound system of American English. *Word, 20*, 157-177.

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

Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley.

Hebb, D. O. (1961). Distinctive features of learning in the higher animal. In R.W.G. A. Fessard, & Konorski (Ed.), *Brain mechanisms and learning* (pp. 37-51). Oxford: Blackwell.

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

Holmgren, G. L. (1967). Physical and Psychological correlates of speaker recognition. *Journal of Speech and Hearing Research*, *10*, 57-66.

Horri, Y., House, A. S., & Hughes, G. W. (1971). A masking noise with speech envelope characteristics for studying intelligibility. *Journal of the Acoustical Society of America, 49*, 1849-1856.

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

Humphreys, M. S. (1976). Relational information and the context effect in recognition memory. *Memory and Cognition, 4*, 221-232.

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

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

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., & 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., Toth, J. P., & Yonelinas, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring Recollection. *Journal of Experimental Psychology: General, 122*, 139-154.

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

James, W. (1983). *The principles of psychology*. Cambridge: MA: Harvard University Press.

Joordens, S., & Merikle, P. M. (1995). Independence or redundancy? Two models of conscious and unconscious influences. *Journal of Experimental Psychology: General, 122*, 462-467.

Johnson, K. (1990). The role of perceived speaker identity in F_0 normalization of vowels. *Journal of the Acoustical Society of America, 88*, 642-654.

Jusczyk, P. W., Pisoni, D. B., & Mullenix, J. W. (1992). Some consequences of stimulus variability on speech processing by 2-month old infants. *Cognition, 43*, 253-291.

Kahn, M. (1975). Arabic emphatics: The evidence for cultural determinants of phonetic sex-typing. *Phonetica, 31*, 38-50.

Kempler, D., & Van Lancker, D. (1987). The right turn of phrase, *Psychology Today* (pp. 20-22).

Kirk, R. E. (1982). *Experimental design: Procedures for the behavioral sciences*. Belmont: CA: Brooks/Cole.

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

Kolers, P. A., (1975). Specificity of operations in sentence recognition. *Cognitive Psychology, 1*, 289-306.

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

Kreiman, J., & Van Lancker, D. (1988). Hemispheric specialization for voice recognition: Evidence from dichotic listening. *Brain and Language, 34*, 246-252.

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 and Psychophysics, 31*, 279-292.

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

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

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

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

Levy, B. A., & Begin, J. (1984). Proofreading familiar text: Allocating resources to perceptual and conceptual processes. *Memory and Cognition, 12*, 621-632.

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.

Levy, B. A., DiPersio, R., & Hollingshead, A. (1992). Fluent rereading: Repetition, automaticity, and discrepancy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 407-417.

Lewandowsky, S., Dunn, J. C., & Kirsner, K. (1989). *Implicit memory: Theoretical issues*. New Jersey: Lawrence Erlbaum.

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

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

Lively, S.E. (1994). Preserving the perceptual record: Retention of talker-specific information in long-term memory. *Research on Spoken Language Processing Technical Report No. 9*. Bloomington, IN: Indiana University Press.

Lively, S. E., & Pisoni, D. B. (1990). Some lexical effects in phoneme categorization: A first report. *Research on Speech Perception Progress Report No. 16*. Bloomington, IN: Indiana University Press.

Lockhart, R. S., Craik, F. I. M., & Jacoby, L. L. (1975). Depth of processing in recognition and recall: Some aspects of a general memory system. In J. Brown (Ed.), *Recognition and recall* (pp. 75-102). London: Wiley.

Logan, J., & Pisoni, D. B. (1987). Talker variability and recall of spoken word lists: A replication and extension. *Research on Speech Perception Progress Report No. 13*. Bloomington, IN: Indiana University Press.

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

Madden, D. J., & Bastian, J. (1977). Probing echoic memory with different voices. *Memory and Cognition*, 331-334.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87*, 252-271.

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

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 into meaning. In W. D. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 3-24). Cambridge, MA: MIT Press.

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

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., Mullenix, 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.

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.

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

McClelland, A. G. R., Pring, L. (1991). An investigation of cross-modality effects in implicit and explicit memory. *Quarterly Journal of Experimental Psychology, 43A*, 19-33.

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.

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

Mehler, J. Dommergues, J. Y., Frauenfelder, U., & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior, 20*, 298-305.

Mehler, J., Dupoux, E., & Segui, J. (1990). Constraining models of lexical access: The onset of word recognition. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 236-262). Cambridge, MA: MIT Press.

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

Miller, R. R., Kasprow, W. J., & Schachtman, T. R. (1986). Retrieval variability: Sources and consequences. *American Journal of Psychology, 99*, 145-218.

Morton, J. (1969). The interaction of information in word recognition. *Psychological Review, 76*, 340-354.

Morton, J. (1979). Facilitation in word recognition: Experiments causing change in the logogen model. In P. A. Kolers, M. E. Wrolstad, & Bouma (Eds.), *Processing models of visible language* (pp. 259-268). New York: Plenum.

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.

Moscovitch, M. (1992). Memory and working-with-memory: A component process model based on modules and central systems. *Journal of Cognitive Neuroscience, 4*, 257-267.

Mullenix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception and Psychophysics, 47*, 379-390.

Mullenix, J. W., Pisoni, D. B., & Martin, C. S. (1989). The effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America, 85*, 365-378.

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

Norman, D. A., & Shallice, T. (1986). Attention to action. Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz & D. Shapiro (Eds.), *Consciousness and self-regulation* (pp. 1-118). New York: Plenum Press.

Nygaard, L. C., & Pisoni, D. B. (1996). Talker-specific perceptual learning in spoken word recognition: Preliminary findings and theoretical implications. *Research on Spoken Language Processing Progress Report No. 20.* Bloomington, IN: Indiana University Press.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. (1994). Speech perception as a talker-contingent process. *Psychological Science, 5*, 42-46.

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 309-328.

Palmeri, T. J., Lively, S. E., & Pisoni, D. B. (1991). Implicit learning of auditory sequences: A first report. *Research on Speech perception Progress Report No. 17.* Bloomington, IN: Indiana University Press.

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

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.

Peterson, D. K., & Pitz, G. F. (1988). Confidence, uncertainty, and the use of information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 85-92.

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* (pp. 1399-1407). Kobe: Japan.

Pisoni, D. B. (1996). Some thoughts on "normalization" in speech perception. In K. Johnson & J. W. Mullenix (Eds.), *Talker variability in speech processing* (pp.1-48). San Diego: Academic Press.

Pisoni, D. B., & Luce, P. A. (1986). Speech perception: Research, theory, and the principal issues. In E. C. Schwab & H. C. Nusbaum (Eds), *Perception of speech and visual form: Theoretical issues, models, and research* (pp. 1-50). New York: Academic Press.

Pisoni, D. B., Nusbaum, H. C., Luce, P.A., & Slowiaczek, M. L. (1985). Speech perception, word recognition and the structure of the lexicon. *Speech Communication, 4*, 75-95.

Pisoni, D. B., & Sawusch, J. R. (1975). Some stages of processing in speech perception. In A. Cohen & S. G. Nooteboom (Eds.), *Structure and process in speech perception* (pp.16-35). New York: Springer-Verlag.

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

Posner, M. I. (1993). Seeing the mind. *Science, 262*, 673-678.

Posner, M. I., Boies, S. J., Eichelman, H. W., & Taylor, R. L. (1969). Retention of visual and name codes of single letters. *Journal of Experimental Psychology Monographs, 79*, 1-16.

Posner, M. I., & Keele, S. W. (1967). Decay of visual information from a single letter. *Science, 158*, 137-139.

Posner, M. I., & Mitchell, R. F. (1967). Chronometric analysis of classification. *Psychological Review, 74*, 392-409.

Posner, M. I., & Rogers, M. (1978). Chronometric analysis of abstraction and recognition. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes, Vol. 5: Human information processing* (pp. 143-188). Hillsdale: NJ: Erlbaum.

Rajaram, S., & Roediger, H. L. (1993). Direct comparison of four implicit memory tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 765-776.

Ratcliff, R., McKoon, G., & Verwoerd, M. (1989). A bias interpretation of facilitation in perceptual identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 378-387.

Reingold, E. M., & Merikle, P. M. (1990). On the inter-relatedness of theory and measurement in the study of unconscious processes. *Mind and Language, 5*, 9-28.

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

Roediger, H. L. III. (1990). Implicit memory: Retention without remembering. *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 and Cognition, 15*, 379-388.

Roediger, H. L., III, Srinivas, K., & Weldon, M. S. (1989). Dissociations between implicit measures of retention. In S. Lewandowsky, J. C. Dunn & K. Kirsner (Eds.), *Implicit memory: Theoretical issues* (pp. 67-84).

Rubin, P., Turvey, M. T., & Van Gelder, P. (1976). Initial phonemes are detected faster in spoken words than in spoken nonwords. *Perception and Psychophysics, 19*, 394-398.

Rugg, M.D., & Nagy, M. E. (1987). Lexical contribution to non-word repetition effects: Evidence from event-related potentials. *Memory and Cognition, 15*, 473-481.

Salasoo, A., & Pisoni, D. B. (1985). Interaction of knowledge sources in spoken word identification. *Journal of Memory and Language, 24*, 210-231.

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.

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. (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 system and implicit memory: Toward a resolution of the multiple memory debate. In A. Diamond (Ed.), *Annals of the New York Academy of Sciences: The development and neural bases of higher cognitive functions, Vol. 608* (pp. 543-571). New York: New York Academy of Sciences.

Schacter, D. L. (1992a). Understanding implicit memory. *American Psychologist, 47*, 559-569.

Schacter, D. L. (1992b). Understanding implicit memory: A cognitive neuroscience approach. *American Psychologist, 47*, 559-569.

Schacter, D. L., Chiu, C. Y. P., & Ochsner, K. N. (1993). Implicit memory: A selective review. *Annual Review of Neuroscience, 16*, 159-182.

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

Schacter, D. L., Rapcsak, S. Z., Rubens, A. B., Tharan, M., Laguna, J. M. (1990). Priming effects in a letter-by-letter reader depend on access to the word form system. *Neuropsychologia, 28*, 1079-1094.

Schneider, W., Dumais, S. T., & Shiffrin, R. M. (1984). Automatic and control processing and attention. In R. Parasuraman & R. Davies (Eds.), *Varieties of attention* (pp. 1-27). New York: Academic Press.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search and attention. *Psychological Review, 84*, 1-66.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*, 127-190.

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry, 20*, 11-21.

Segui, J., Dupoux, E., Mehler, J. (1990). The role of the syllable in speech segmentation, phoneme identification, and lexical access. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 263-280). Cambridge, MA: MIT Press.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*, 523-568.

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.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*, 127-190.

Squire, L. R. (1987). *Memory and brain*. New York: Oxford University Press.

Squire, L. R., & Zola-Morgan (1991). The medial temporal lobe memory system. *Science, 253*, 1380-1386.

Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review, 99*, 195-231.

Squire, L. R., & Knowlton, B. J. (1995). The organization of memory. In J. H. Morowitz & Singer (Ed.), *SFI studies in the sciences of complexity, Vol. XXII* (pp. 63-97). New York: Addison-Wesley.

Squire, L. R., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science*, *253*, 1380-1386.

Summerfield, Q. H., & Haggard, M. P. (1973). Vocal tract normalization as demonstrated by reaction times. *Report of Speech Perception Research in Progress No. 3*. Belfast: The Queen's University of Belfast.

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

Thompson, C. P. (1985). Voice identification: Attempted recovery from a biased procedure. *Human Learning, 4*, 213-224.

Thorne, B., Kramearae, C., & Henley, B. (1983). Language, gender and society. Rowley, MA: Newbury House.

Tulving, E. (1985). How many memory systems are there? *American Psychologist, 40*, 385-398.

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.

Turner, C. W., & Pisoni, D. B. (1993-1994). Some comments on implicit effects of voices on spoken word recognition. *Research on Spoken Language Processing Progress Report No. 19.* Bloomington, IN: Indiana University Press.

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

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, 36*, 198-212.

Voiers, W. D. (1964). Perceptual bases of speaker identity. *Journal of the Acoustical Society of America, 36*, 1065-1073.

Warrington, E. K., & Weiskrantz, L. (1970). Amnesic syndrome: Consolidation or retrieval. *Nature, 228*, 628-630.

Warrington, E. K., & Weiskrantz, L. (1974). The effect of prior learning on subsequent retention in amnesic patients. *Neuropsychologia, 12*, 419-428.

Weldon, M. S. (1991). Mechanisms underlying priming on perceptual tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 526-541.

Wickelgren, W. A. (1969). Context-sensitive coding, associative memory, and serial order in (speech behavior). *Psychological Review, 76*, 1-5.

Willinghan, D. B., & Preuss, L. (1995). The death of implicit memory. *An Interdisciplinary Journal of Research on Consciousness* [On-line], 2. Available: E-mail: psyche-l@iris.rfmh.org.

Woltz, D. J. (1990). Repetition of semantic comparisons: Temporary and persistent

priming effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 392-403.