RESEARCH ON SPOKEN LANGUAGE PROCESSING

Technical Report No. 9

September 1, 1994

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

Supported by:

Department of Health and Human Services
U.S. Public Health Service
National Institutes of Health
Research Grant No. DC-00111-17
and Training Grant No. DC-00012-15

PRESERVING THE PERCEPTUAL RECORD: RETENTION OF TALKER-SPECIFIC INFORMATION IN LONG-TERM MEMORY

Scott Eric Lively

Submitted to the faculty of the University Graduate School

in partial fulfillment of the requirements

for the degree

Doctor of Philosophy

in the Department of Psychology

Indiana University

August, 1994

Acknowledgments

When I came to Bloomington in 1985, I had no idea how many people would have a such a significant impact on my life. Now that I am finishing my dissertation, I am lucky enough to get a chance to thank those people in a public forum. First and foremost, I want to thank my wife, Juli Brown, for her unwavering support and patience throughout my undergraduate and graduate careers. I also owe a huge debt of gratitude to my parents, Barry and Ellie, and my sister, Kristin. Without their constant encouragement, I am sure that I would not have finished my degree. I have been very lucky to be associated with David Pisoni and his lab. I will always be grateful for David's support. He is a tremendous leader of a world-class laboratory. I would also like to thank the other members of my committee, Rich Shiffrin, Bob Peterson, and Bob Port, for their suggestions and encouragement during all phases of this project. Finally, I would like to thank a number of the SRLers who have supported, challenged, and encouraged me during my stay. John Logan and Steve Goldinger have been wonderful friends and collaborators. Tom Palmeri, Lynne Nygaard, Luis Hernandez, Ann Bradlow, Steve Chin, Matt Peuquet, Lisa Burgin, Carolyn Trilling, Carl Turner, John Karl, and Clovis Lark all deserve special thanks for friendship and support. I am very lucky to be surrounded by such a great group of people.

This dissertation is dedicated to the memory of R. L. Lively, whose voice I will always remember. Work in this dissertation was sponsored by Department of Health and Human Services U. S. Public Health Service Reseach Grant DC-00111-18 and Training Grant No. DC-00012-15 to Indiana University, Bloomington, Indiana.

Abstract

Memory for the surface forms of spoken words was investigated in a series conceptually-driven and data-driven test of implicit and explicit memory. Explicit recognition memory performance was facilitated when the voice of the talker was preserved between the study and test sessions. Access to the characteristics of a talker's voice was observed even after a delay of twenty-four hours. In contrast, the tests of implicit memory were insensitive to changes in the surface forms of spoken words between the study and test sessions: The magnitude of repetition priming effects were not influenced by changes in voice. The results of the present investigation suggest that information related to a talker's voice is automatically encoded into memory. However, this information may only be accessed and used to facilitate recognition processes under some data-driven conditions. The observation that talker-specific information is encoded and retained in long-term memory calls into question the traditional assumption that spoken language processing operates on a set of abstract, canonical linguistic units. Rather, the present findings suggest that fine perceptual details are retained in memory and these details are used to facilitate processing.

Preserving the Perceptual Record: Retention of Talker-Specific Information in Long-Term Memory

Table of Contents

Preserving the Perceptual Record: Retention of Talker-Specific Information in Long-Term Memory

CHAPTER I: Introduction

In 1959, Chomsky broke the strangle-hold that behaviorists held over experimental psychology with his review of Skinner's book *Verbal Behavior*. According to behaviorist dogma, all human behavior, including language, could be accounted for by principles that related external stimuli to observable responses. Chomsky's review dealt the death blow to behaviorism by pointing out that explanations of the infinite creativity and generativity of language are not amenable to simple stimulus-response pairings. Rather, what is needed is a complex mental architecture that supports representations which are operated on by a set of rules or a grammar. This grammar, when combined with the appropriate lexicon, produces all or the legal strings within a language and none of the illegal ones. Chomsky's critique of the behaviorist version of linguistic skill and acquisition put language back in the mind of the user, rather than in the environment between the user and the objects of his or her "mands" (Skinner, 1957).

While Chomsky's review contributed greatly to the fall of behaviorism and to the rise of the cognitive revolution, it also played a foundational role in determining the issues that psycholinguists and researchers in speech perception would pursue for the next forty years. Theoretical linguists set about the business of describing the abstract nature of language and trying to uncover linguistic universals. Psycholinguists spent a great deal of research effort attempting to demonstrate the psychological reality of the constructs that theoretical linguists described (Miller, 1962). Researchers who were interested in speech perception devoted much of their effort to demonstrating the psychological reality and utility of phonemes, syllables, and other units of linguistic analysis (Cole & Scott, 1974; Massaro, 1972; Oden & Massaro, 1978; Savin & Bever, 1970; Wicklegren, 1969).

For example, much of the speech research done in the 1950's was conducted with the goal of finding a set of first-order invariants that map the acoustic waveform onto a set of phonemic units. Even though it quickly became apparent that such a set of invariants might not exist, considerable effort was still expended on determining which linguistic units are the primary ones used by the neural mechanisms responsible for processing speech (Cooper, Delattre, Liberman, Borst, & Gerstman, 1952). Arguments for a variety of different types of segments have been made throughout the years: These elements have included units as small as phonetic features and phonemes to units as large as the syllable and possibly even larger units (Cutler & Norris, 1988; Marslen-Wilson, 1987; McNeill & Lindig, 1973).

This close association between the study of speech perception and theoretical linguistics has also been combined with information processing models drawn from cognitive psychology (Studdert-Kennedy, 1976). The benefits of combining the units of analysis from linguistic theory with the computational machinery of an information processing model are appealing. Parsimonious accounts can be given for how the cognitive architecture operates on a set of idealized, abstract symbols in order to combine them together. However, one consequence of this approach to combining linguistic theory with cognitive psychology has been that the nature of the speech signal is often completely disregarded. For example, most theories of speech perception deal strictly with how phonetic information is extracted from a "sanitized" signal that has been stripped of its paralinguistic or indexical information (Liberman, 1970; Liberman & Mattingly, 1985). Sources of variability are discarded as noisy and uninformative in favor of abstract symbols derived through transformation and recoding of the original input.

In addition to forming tight bonds with linguistic theory and borrowing from information processing theories, researchers in the field of speech perception have also borrowed heavily from findings in auditory psychophysics (Kuhl, 1992; Macmillan, Caplan, & Creelman, 1977). The common assumption has been that

basic pscychoacoustic principles might be used to explain general auditory functioning as well as the more specialized domain of speech perception (Cutting & Rosner, 1974; Miller, Wier, Pastore, & Kelly, 1976; Pisoni, 1977). One consequence of this relationship has been that a great deal of effort has been devoted to determining the psychophysical limits of the auditory system with regard to speech sounds (Kewley-Port & Watson, 1994). This style of research represents another way in which the true nature of speech signals has been ignored. Rather than treating speech as a highly complex, time-varying signal, the psychophysical approach to speech perception attempts to set the lower bounds on the functioning of the auditory system.

Another consequence of combining speech perception with traditional linguistic theory and psychophysical inquiry has been that researchers working on issues in spoken language processing have tended to ignore developments in research on memory and categorization (Pisoni & Lively, 1994). Whereas research on speech perception has tended to emphasize minimality, economy and information reduction during processing and storage, recent theorizing on issues in long-term memory and categorization has emphasized the storage of specific, highly detailed information. Many current models stress the idea that memory is composed of individual traces that are brought to bear on the recognition, recall, identification and categorization of perceptual information (Eich, 1982; Gillund & Shiffrin, 1984; Hintzman, 1986; Medin & Schaeffer, 1978; Nosofsky, 1986, 1987). Although many of the models differ greatly from each other in terms of their storage and retrieval assumptions, the basic tenet of each model is that experience is composed of individual events and that stored events are relevant to the perception of new events.

The general purpose of the present investigation is to examine the contribution that recent theorizing on long-term memory can make to the study of speech perception and spoken language processing. More specifically, the experiments detailed in this report describe conditions under which listeners encode and use talker-specific information during spoken language processing. The guiding assumption, borrowed from recent models of categorization and memory, is that memory for specific experiences or instances plays an active role in the processing of incoming information. In the sections that follow, a more detailed examination is given to the traditional assumptions that have driven research spoken language processing. Following this review, the contributions that recent models of long-term memory and categorization to speech perception and spoken word recognition are considered.

A. Reductionist Assumptions in Speech Perception and Spoken Word Recognition

Research on spoken language processing has been dominated by at least two reductionist assumptions. The first important assumption is that many issues in spoken word recognition can be addressed by making the appropriate changes to models of visual word recognition. While this may be true to a certain extent, one of the goals of this thesis is to demonstrate that issues in spoken word recognition can also be addressed by considering the relationship between spoken language processing and memory processes, specifically the differences between implicit and explicit memory. By broadening the scope of evidence considered relevant to spoken word recognition, we may begin to get a firmer grasp on a number of long-standing issues concerning how the acoustic speech signal is translated into a neural representation that listeners can act upon.

The second assumption that has guided research on spoken language processing is that variability in the acoustic signal is noise (Elman & McClelland, 1986). Sources of variability must be reduced or eliminated in order to facilitate the translation of the physical signal into a mental, symbolic representation. Recently, however, a number of studies have shown that variability in the speech signal may not be noise (Goldinger, 1992; Nygaard, Sommers, & Pisoni, 1994). Rather, different sources of variability in the acoustic waveform may serve as useful information to the listener that can be used to enhance recognition processes. A second goal of this thesis is to add to the growing body of literature that demonstrates listeners can use information about a speaker's voice to facilitate the processing of spoken words. In particular, this thesis outlines some of the task

demands in implicit and explicit memory paradigms that require listeners to access and encode the surface forms of previously stored spoken words.

A.1. The Relationship between Visual Word Recognition and Spoken Word Recognition

Before describing how research on implicit memory might be beneficial to the study of spoken word recognition, it is important to consider in more detail the two reductionist assumptions listed above that have guided spoken language processing in the past. The first assumption was that models of visual word recognition could be adapted to solve problems in spoken word recognition (Forster, 1976, 1979; Morton, 1969, 1970). While the types of mechanisms employed by models in the two modalities may be similar, the translation of the inputs from visual characters on a page to auditory signals has not always been smooth. Four important issues concerning the nature of spoken words need to be addressed when translating models of visual word recognition into models of spoken word recognition. The first important difference between spoken word recognition and visual word recognition concerns the constancy of the input signal. In visual word recognition, under normal viewing conditions, the same signal is available for rescanning when ambiguities or misinterpretations are encountered. Spoken language, in contrast, is transient in nature, unfolds over time, and is only available to the listener once. Furthermore, the phonemes or segments that compose spoken language differ in duration from each other. As a consequence, some segments, such as vowels, are available to listeners for longer periods of time than others. This difference in temporal distribution between written and spoken language has not always been respected in the development of models of spoken word recognition (Klatt, 1989; McClelland & Elman, 1986).

A second important difference between written language and spoken language concerns the invariance of the signal. Written language displays a certain degree of physical invariance: Each time a typist strikes a particular key on a typewriter, the same form is produced. In contrast, spoken language does not display such invariance (Chomsky & Miller, 1963). Variation in the acoustic waveform arises due to both within-speaker and between-speaker differences. For example, variability within a speaker may occur as a function of speaking rate, place of articulation, phonetic context, or perceived emotional stress (Liberman, Delattre, Cooper, & Gerstman, 1954; Lisker & Abramson, 1970; Lively, Pisoni, Summers, & Bernacki, 1993; Streeter, MacDonald, Apple, Krauss, & Galotti, 1983; Summerfield, 1981). Variation among speakers may arise because of physical differences, such as differences in head size or vocal tract length, or cultural differences, such differences in dialects and accents (Fant, 1973; Joos, 1948; Ladefoged, 1967; Ladefoged & Broadbent, 1957; Ladefoged, 1980; Peterson & Barney, 1952).

The importance of the difference in constancy between written language and spoken language is typically ignored (see however, Klatt, 1979, 1980). Both models of visual word recognition and spoken word recognition typically assume that the inputs are matched to abstract static templates stored in the lexicon in long-term memory. This strategy raises two problems for models of spoken word recognition that are not typically addressed in an explicit manner. First, most theorists have not explicitly stated how variability in the signal is removed from the signal when a match to a template in the lexicon is made. Second, given that spoken language is temporally extended, researchers have not adequately described how a static template model can account for the speed and accuracy of spoken word recognition (Marslen-Wilson, 1985, 1987).

A third important contrast between written and spoken language concerns the linearity of the input signal. Characters on a page are clearly segregated so that successive letters are separated from each other by a fixed distance. Each successive letter or cluster of letters represents a successive sound within the word. Spoken language, in contrast, does not demonstrate such orderly linearity. In contrast, the symbols that are used to transcribe a spoken utterance do not correspond uniquely to a given portion of the speech waveform (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Studdert-Kennedy, 1983). Thus, there is no one-to-one correspondence between discrete portions of the acoustic signal and the abstract symbols used to

transcribe it (Chomsky & Miller, 1963; Fant, 1973). This overlap blurs the boundaries in the acoustic signal between discrete phonemes. In natural speech, phonemes are coarticulated so that they overlap in time. By coarticulating phonemes, the information transmission rate of speech can exceed ten phonemes per second (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Studdert-Kennedy, 1980). This increases the efficiency of human acoustic communication far beyond levels obtained with tone-based cipher alphabets (Liberman et al., 1967, Liberman, 1991).

The failure of speech signals to demonstrate linearity has had two important consequences for models of speech perception and spoken word recognition. One effect of the failure of linearity is that it is difficult to find invariant acoustic features that demonstrate a one-to-one correspondence with perceived phonemes (Kewley-Port, 1982, 1983; Stevens & Blumstein, 1978, 1981). A second consequence of the failure of linearity concerns how the acoustic signal is used to access the mental lexicon. Klatt (1989) argues that if the acoustic signal is mapped onto a phonemic representation prior to activating candidates in the lexicon and if word recognition is accomplished through the serial activation of successive phonemes, then important coarticulatory information about successive phonemes may be lost and may be unrecoverable. Thus, early commitment to an intermediate phonemic representation entails a loss of potentially useful information that may be informative to the perceiver. Errors caused by an early commitment to a particular symbolic representation may be difficult to correct because important coarticulatory information in the acoustic signal may be unavailable for further analysis.

A final difference between the inputs used to drive written language processing and spoken language processing deals with segmentation. Spatial boundaries separate each letter within a printed word and larger boundaries separate words on a printed page. In spoken language, temporal boundaries do not segregate successive sounds or words. Rather, the speech signal appears to flow as a continuous acoustic stream. It is important to note that the lack of acoustic-phonetic segmentation is very important to maintaining the efficiency of spoken language. As Liberman (1991) points out, if discrete segmentation were available in the speech signal, speaking would be tantamount to spelling and the high information transmission rate of spoken language would be lost. In spite of the fact that speech signals fail to demonstrate orderly segmentation, listeners appear to effortlessly parse spoken utterances by relying on a variety of cues such as stress, intonation and context-sensitive cues to phonemes (Cutler, 1976; Cutler & Darwin, 1981; Nakatani & Schaffer, 1978; Wicklegren, 1969).

Hockett (1955) neatly summarizes the problems of invariance, linearity and segmentation by comparing spoken language to a series of colored Easter eggs:

> Imagine a row of Easter eggs carried along a moving belt; the eggs are of various sizes, and variously colored, but not boiled. At a certain point, the belt carries the row of eggs between the two rollers of a wringer, which quite effectively smash them and rub them more or less into each other. The flow of eggs before the wringer represents the series of impulses from the phoneme source; the mess that emerges from the wringer represents the output of the speech transmitter. At a subsequent point, we have an inspector whose task it is to examine the passing mess and decide, on the basis of the broken and unbroken yolks, the variously spread out albumin, and the variously colored bits of shell, the nature of the flow of eggs which previously arrived at the wringer [p. 210].

The failures of invariance, linearity, and segmentation present serious challenges to the development of models of spoken word recognition that are simple adaptations of models of visual word recognition (Lively, Pisoni, & Goldinger, 1994). In particular, these problems point out the necessity for carefully specifying how the acoustic signal is gathered over time and converted into a neural representation that can be used to access the mental lexicon. These problems are particularly difficult for models of spoken word recognition that assume an

early commitment to a phonemic representation of the input and discard sources of variability that may be "lawful" and informative to the perceiver (McClelland & Elman, 1986). In a later section, evidence is reviewed which suggests that listeners may not discard some sources of variability, such as information about a speaker's voice, when recognizing spoken words. These studies demonstrate that preserving the surface forms of spoken words has a beneficial effect on the later processing of both old and new words.

A.2. The Traditional Role of Variability in Spoken Language Processing

The second important reductionist assumption that has guided spoken language processing for more than forty years concerns the type of information that is used to contact the mental lexicon and to derive meaning from the acoustic signal. Typically, research on spoken language processing has adopted the information processing framework in which the time-varying acoustic signal is recoded into a more durable, symbolic form. During the process of recoding the acoustic-phonetic input, perceptual information, such as the pitch of the speaker's voice or her speaking rate, is assumed to be lost and unavailable for further analysis. The following quote makes this point clear:

> "The study of speech perception ... has in recent years begun to adopt the aims, and often the methods, of the information processing models of cognitive psychology which have proven fruitful in the study of vision ... The underlying assumption is that perception has a time-course, during which information in the sensory array is 'transformed, reduced, elaborated (Neisser, 1967) and brought into contact with long-term memory (recognized)" (Studdert-Kennedy, 1976).

Two points are important to note about Studdert-Kennedy's perspective on spoken language processing. The first point concerns the fact that research on speech perception and spoken word recognition has often been guided by findings from visual perception and visual word recognition (Eimas & Corbit, 1973; Goldinger, 1992; Remez, 1987). However, as noted above, this translation between input modalities has not always been smooth and has generally not shown a concern for the difficult problems that spoken language presents for the perceptual system.

The second point concerns the nature of the perceptual operations that are assumed to be conducted on the acoustic signal. These operations include transformation from an auditory percept into a reduced, idealized, abstract phonemic form. This assumption concerning information reduction may allow the perceptual system to efficiently encode and decode the message carried by the speech waveform. However, this efficiency comes at a cost. First, by reducing the amount of information in the input and committing to an early phonemic representation, the system is made brittle. Error recovery is difficult because the original signal is not retained for possible further analysis (Klatt, 1979, 1980). Second, the assumptions concerning information reduction also tacitly deny the possibility that variability in the acoustic signal could be informative to the perceiver. Indeed, the following quotes from two recent theoretical papers on speech perception address very clearly the assumed role that sources of variability, such as information about a speaker's voice, might play in decoding the speech signal:

> "Like most ecological events ..., one in which linguistic communication takes place is highly structured and complex. Accordingly, it can be decomposed for study in may different ways. One way in which it is almost invariably decomposed by psycholinguists and linguists is into the linguistic utterance itself on the one hand and everything else on the other. In ordinary settings in which communication takes place, this is almost certainly not a natural partitioning because it leaves out several aspects of the setting that contribute interactively with the linguistic utterance itself. These include the talker's gestures (McNeill, 1985), aspects of the

environment that allow the talker to point rather than to refer verbally, and the audience, whose shared experiences with the talker affect his or her speaking style." (Fowler, 1986, p. 3)

"... as we have noted, speech perception uses all of the information in the stimulus that is relevant to phonetic structures: every potential cue proves to be an actual cue ... In contrast, irrelevant information in the stimulus set is *not* used .... The exclusion of the irrelevant extends of course to stimulus information about voice quality, which helps to identify the speaker (perhaps by virtue of some other module) but has no phonetic importance..." (Liberman & Mattingly, 1985, p. 28)

The quotes above from Fowler (1986) and Liberman and Mattingly (1985) point out the contrasting way in which sources of variability are treated by researchers in the field of spoken language processing. On the one hand, Fowler points out that the division between the linguistic properties of speech and "everything else" is probably incorrect. However, because speech perception is typically cast in terms of this dichotomy the easiest way to make progress on traditional issues in the field is to continue to operate under the same set fundamental assumptions. To her credit, Fowler does treat speech perception as larger problem in ecological psychology in which the job of the researcher is to find the information that affords the perceiver recovery of the speaker's linguistic intent. In contrast, Liberman and Mattingly (1985, 1989) explicitly disavow any role that voice information and variability could play in speech perception and argue that the phonetic content of the acoustic signal is privileged (Liberman, 1982). As a consequence, they subject the field to a continued search for invariance at some level of representation.

A.3. On Perceptual Normalization

If information about a talker's voice is assumed to be noise that is uninformative to the perceiver, then some mechanism must be provided to remove this noise source from the speech waveform. In general two different approaches to normalization and the recovery of phonemic invariants have been taken. First, some models of speech perception have attempted to ignore or minimize the importance of the problem of variability in the signal. For example, the revised Motor Theory (Liberman & Mattingly, 1985) has avoided the problem by suggesting that a special purpose "phonetic module" recovers the intended articulatory gestures of the talker in order to arrive at a phonetic percept. According to this perspective, speech is an auditory signal that is perceived by a neurologically specialized module (Fodor, 1983; Liberman, 1982; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Mattingly & Liberman, 1985). A related model, Fowler's (1986, 1990) direct realist approach, also suggests that articulatory gestures are the objects of perception. However, in contrast to the motor theory, the direct realist approach maintains that speech perception is not mediated by a special module. Rather, listeners directly recover articulatory gestures. Fowler (1986) argues that the problem of invariance is more imagined than real and that the burden of explaining speech perception rests on developing more accurate descriptions of the complex processes that generate articulatory gestures (Kelso, Salzman, & Tuller, 1986). Both the revised motor theory and the direct realist approach down-play the significance of the role of variability in speech perception.

The second approach to dealing with the problem of invariance has been to propose a normalization mechanism that strips away variability from the acoustic signal in order to sanitize it and make it usable by a simple pattern matching operation (Joos, 1948). The traditional assumption concerning normalization has been that natural speech signals are noisy and need to be rescaled in order to be accurately perceived. For example, without rescaling, accurate vowel spaces for different talkers cannot be mapped out because vowel categories overlap across speakers (Ladefoged & Broadbent, 1957; Nearey, 1989; Peterson & Barney, 1952). The normalization mechanism is assumed to remove information about the talker's voice, such as glottal source characteristics, speaking rate, information about dialect, and a number of other factors that can be used to identify the speaker from the acoustic waveform. Once this information has been discarded, the listener is

assumed to be left with a signal that can be matched to an abstract, canonical linguistic representation that is stored in long-term memory.

Two basic types of normalization mechanisms have been proposed for dealing with acoustic-phonetic variation among talkers. These processes vary in the size of the unit they take as the reference for normalization. Intrinsic normalization mechanisms take the smallest unit of speech as their input. In this case, a single syllable or even a single phonetic segment is assumed to provide the basis for normalization. According to intrinsic normalization hypotheses, listeners compensate for differences between talkers by using information about a speaker's fundamental frequency (F0), the third formant (F3) of their vowels or a combination of the two sources of information to provide reference points for normalization (Ainsworth, 1975; R. L. Miller, 1953; J. D. Miller, 1989; Peterson, 1961; Rakerd & Verbrugge, 1987; Syrdal, 1984; Syrdal & Gopal, 1986; Verbrugge & Rakerd, 1986; Verbrugge, Strange, Shankweiler & Edman, 1976). The rational for using F0 and/or F3 is that these sources of information are relatively invariant across vowels and phonetic contexts. Thus, they provide stable reference points for rescaling the vowel spaces of different talkers. Nusbaum and Morin (1991) refer to intrinsic normalization mechanisms as "self-normalizing" or "structural estimation" processes because all of the information necessary for estimating the size and shape of the vocal tract is assumed to be contained within a single segment.

Intrinsic normalization mechanisms can be contrasted with extrinsic normalization mechanisms (Ainsworth, 1975; Nearey, 1989). Whereas intrinsic systems work within a single vowel or syllable, extrinsic mechanisms require information distributed across the vowel systems of different talkers. According to assumptions about extrinsic normalization mechanisms, differences among talkers are determined by sampling a range of vowels from each speaker in order to derive estimates of vocal tract size or vowel formant ranges (Gerstman, 1968; Joos, 1948; Ladefoged; 1967; Ladefoged & Broadbent, 1957; Nearey, 1978). Thus, extrinsic mechanisms require more than a single segment or syllable in order to establish a stable phonetic space for a set of talkers. In contrast to the intrinsic mechanisms, which were self normalizing, extrinsic mechanisms require a process of perceptual learning or "contextual tuning" (Nusbaum & Morin, 1991).

The goal of each of the procedures described above (direct or mediated recovery of articulatory gestures, intrinsic normalization, extrinsic normalization) is to reduce variability in the acoustic waveform in order to provide an unambiguous signal that can be matched against a linguistic representation stored in long-term memory. The benefit of a normalization procedure to the perceptual system is that it may allow for efficient processing of phonetic information. The drawback to a normalization mechanism is that potentially important sources of information may be discarded. This information may be useful for resolving ambiguity in the signal or for filling in lost information due to degradation of the speech waveform.

As noted above, the purpose of the normalization mechanism is to convert the acoustic signal into an idealized, symbolic code. In the next section, the fate of the information discarded by the linguistically-based normalization process is considered. In particular, attention is given to the processing and identification of speakers' voices. Following the review of research on the perception of voices, the impact of talker variability on spoken language processing is addressed.

A.4. On Voice Recognition

Although researchers who work on theoretical issues in speech perception and spoken word recognition have largely ignored questions concerning the processing and retention of properties related to a talker's voice, an active research campaign has been carried out to examine listeners' abilities to recognize and identify familiar and unfamiliar voices (Bricker & Pruzansky, 1976; Hecker, 1971; Pollack, Pickett, & Sumby, 1954). Much of the research on voice identification has been carried out in collaboration with the legal community. The goal has been to establish the reliability of "ear witness" testimony, the auditory analog of "eye witness"

testimony (Clifford, 1983; Tosi, 1979). In addition to considering the legal ramifications of listeners' abilities to identify voices, the neurological basis of voice identification has also been investigated extensively (Van Lancker, Cummings, Kreiman, & Dobkin, 1988). This work has focused largely on showing that familiar voices are processed by different neural mechanisms than unfamiliar voices.

McGehee (1937, 1944) conducted some of the earliest investigations of listeners' abilities to identify voices. Her research was motivated by the trial of United States v. Hauptmann. Hauptmann was convicted of kidnapping Charles Lindbergh's son largely on evidence that Lindbergh claimed to be able to identify Hauptmann's voice several years after the kidnapping occurred. In McGehee's experiments, listeners were asked to recognize a target voice presented after delays of one day, one week, two weeks, one month, and five months. Recognition accuracy decreased rapidly as a function of delay: accuracy after 24 hours was 83%. However, after five months, recognition accuracy decreased to only 13%. Although McGehee's findings suggest that listeners' abilities to identify voices decreases rapidly over time, her findings are difficult to interpret because target voices were presented in the context of different distracter voices across subject groups.

Forty years after McGehee conducted her study, Saslove and Yarmey (1980) investigated listeners' abilities to identify unfamiliar voices. In their experiment, subjects were tested immediately after hearing a target passage or were tested 24 hours later. Some listeners were informed that they would be asked to identify a particular voice at a later time, whereas other listeners were uninformed. At the time of test, some subjects heard exactly the same passage that they had heard earlier. Other subjects were tested with the same passage, however, the tone of the voice of the talker changed between the first presentation and the test presentation. Several aspects of Saslove and Yarmey's (1980) findings are interesting to note. First, subjects who were informed that they would be tested in a voice recognition task performed more accurately than uninformed subjects. This suggests that explicit recognition of indexical information may be poor if subjects are not directed explicitly to attend to this information. Second, Saslove and Yarmey reported that a delay of 24 hours had little effect on subjects' recognition accuracy. Finally, they reported that changing the tone of the voice from the first presentation to the second presentation reduced subjects' accuracy to chance levels. This finding suggests that listeners' accuracy in recognizing voices is highly sensitive to contextual factors related to the talker's voice (Roediger, 1990).

Legge, Grosman, and Pieper (1984) argued that much of the research conducted on memory for voices has focused on memory for a small set of voices and that this may have encouraged subjects to use special encoding strategies that are resistant to forgetting (Carterette & Barneby, 1975; Clifford, Rathborn, & Bull, 1981; McGehee, 1937; Saslove & Yarmey, 1980). Thus, reports of highly accurate memory for unfamiliar voices may be accounted for by assuming that subjects engage a complex set of encoding strategies. In their study, Legge et al. manipulated the number of talkers that were presented during a study phase and then tested recognition after a variable delay. They found that recognition accuracy did not decrease significantly from a test given 15 minutes after the encoding phase to a test given 10 days later (Clifford et al., 1981; McGehee, 1937). However, Legge et al. also reported that recognition accuracy decreased significantly as the number of voices to be remembered was increased from 5 to 20. These results suggest that subjects' abilities to uniquely preserve unfamiliar voices in memory is poor and that accessibility decreases as subjects are introduced to more unfamiliar voices.

Much of the early work conducted on voice recognition differentiated between familiar and unfamiliar voices. Typically, only unfamiliar voices were used in these experiments. Moreover, all unfamiliar voices were treated as equal. No distinction was made as to whether some voices might be easy or difficult to perceive or remember. However, more recent studies have made a distinction between different types of unfamiliar voices. For example, Papcun, Kreiman, and Davis (1989) recorded ten males with the same dialect and had subjects rate the memorability of each of the voices. A voice that was rated as easy to remember, a voice that was rated as hard to remember, and an intermediate voice were selected as target voices for a recognition memory

experiment. Subjects then listened to a passage during study produced by the easy-to-remember talker, the hard-to-remember talker or the intermediate talker. The test phase was conducted after delays of one, two or four weeks. Across delays, hit rates did not decrease for the three types of voices. However, false alarm rates increased more for the difficult to remember voice than for the easy to remember voice or the intermediate voice. Papcun et al. interpreted these findings in the context of a prototype model which suggested that some voices are difficult to remember because they lie closer to a modal value or a schema of what voices with a particular dialect should sound like (Evans & Arnoult, 1967). Thus, voices that do not deviate substantially from the prototype are assimilated, while other more distinctive voices are not assimilated.

While Papcun et al.'s (1989) results suggest that familiarity with a dialect may affect explicit voice identification, Goggin, Thompson, Strube, and Simental (1991) and Thompson (1987) have addressed the issue of familiarity with the speaker's language more directly. The researchers examined English listeners' abilities to identify unfamiliar voices speaking English or German and German listeners' abilities to identify unfamiliar voices speaking English or German. A strong effect of language familiarity was obtained: English listeners were more accurate at identifying English voices, while German listeners were more accurate at identifying German voices. Interestingly, in another condition, Goggin et al. tested English-Spanish bilinguals and found no familiarity effect: Bilinguals were equally accurate at identifying voices speaking Spanish or English. Finally, they reported that English listeners' abilities to identify talkers' voices decreased as the linguistic structure of the stimulus materials deviated from grammatical English. Performance decreased dramatically as normal text was altered to reversed text. Goggin et al.'s results suggest that familiarity with a language can provide useful information to voice identification. We will return to this point below when we consider the relationship between speech perception and voice identification.

Linguistic familiarity alone, however, cannot account for listeners' abilities to distinguish between familiar and unfamiliar voices. DeCasper and Fifer (1980) showed that 3-day old infants are capable of distinguishing their mothers' voices from other voices. Furthermore, Mann, Diamond, and Carey (1979) demonstrated that young children do not differ from adults in terms of their ability to identify familiar voices. These results indicate that voice recognition precedes the acquisition of language and may play an important social role in development. For example, the ability to recognize caregivers from noncaregivers based on voice cues is displayed very early in development.

Much of the research described above was concerned with the processing and identification of unfamiliar voices. Van Lancker and her colleagues, however, have investigated extensively the identification of familiar voices (Van Lancker, Kreiman, & Emmorey, 1985; Van Lancker, Kreiman, & Wickens, 1985). Their studies have focused on the critical cues that are necessary to successfully recognize a talker's voice. In general, no single cue or small set of cues has be determined to be sufficient to support accurate identification of all voices. For example, Van Lancker, Kreiman, and Emmorey (1985) found that some familiar voices were more affected than others by reversing the speech waveform. Under these conditions, the phonetic structure of speech is lost, but cues about pitch and the range of pitch are maintained. In another study, Van Lancker, Kreiman, and Wickens (1985) found that accuracy of identification of famous voices varied idiosyncratically with changes in speaking rate. Moreover, voices that were affected by changes in speaking rate were not necessarily affected by reversing the waveform. On the basis of these results, Van Lancker and her colleagues argued that voice identification is not based on an invariant set of cues. Rather, voice identification relies on "a loosely structured constellation of cues, any of which or any combination of which can evince recognition" (Van Lancker, Kreiman, & Emmorey, 1985, p.33).

Each of the experiments described above has been concerned with a cognitive measure of voice identification accuracy. However, an extensive body of research has also been conducted to localize the neuroanatomical structures necessary for voice identification. Several sources of evidence suggest that the right hemisphere is responsible for the identification of familiar voices. Kreiman and Van Lancker (1988) showed a

relative left ear advantage for the identification of famous voices: Subjects were more accurate at identifying the voices of celebrities when they were presented to the left ear in a dichotic listening task, relative to when the famous voices were presented to the right ear. Furthermore, the authors reported a right ear advantage for word identification. Their findings suggest that the two types of information were processed in different neuroanatomical locations: Linguistic information was processed by the left hemisphere, while voice information was processed by the right hemisphere.

The processing of voice information by the right hemisphere can be further localized by examining patients with brain damage. Van Lancker and her colleagues have examined voice identification and voice discrimination in a large number of patients with left hemisphere brain damage, right hemisphere damage and bilateral damage (Van Lancker, Cummings, Kreiman, & Dobkin, 1988; see Van Lancker, 1991 for a review). In general, patients with right parietal lobe damage identify the voices of celebrities very poorly (Van Lancker & Canter, 1982). In contrast, patients with left hemisphere damage are able to identify famous voices, although they may have trouble processing the linguistic content of the utterances due to aphasia. It is interesting to note, however, that while voice identification is spared in patients with left hemisphere damage, discrimination among voices is impaired in subjects with either left or right temporal lobe damage. These findings suggest that voice identification and voice discrimination are served by different neuroanatomical structures and that the two processes are independent of each other (Van Lancker & Kreiman, 1987; Van Lancker, Kreiman, & Cummings, 1989).

Taken together, the results of the studies described above suggest that listeners are very accurate at recognizing familiar and unfamiliar voices. However, the ability to explicitly recognize unfamiliar voices decreases rapidly with the passage of time (McGehee, 1937, 1944). The capacity to recognize familiar voices appears at a very early stage developmentally and may be spared by some forms of brain damage (DeCasper & Fifer, 1980; Van Lancker & Kreiman, 1987). Finally, it appears that listeners do not use an invariant set of acoustic or prosodic cues to recognize familiar voices. Rather, listeners rely on idiosyncratic elements of an individual talker's voice to support identification and recognition.

A.5. On the Effects of Talker Variability

A.5.a. Vowels and Consonants

As outlined in the section on perceptual normalization, researchers in speech perception have tended to ignore or minimize the role that information about speaker's voice might play in spoken language processing. Similarly, many of the findings on voice and speaker identification have been generated independently of any concern for the effects that a talker's voice might have on a listener's ability to encode, identify and recognize spoken syllables and words. However, another line of research has emerged recently which suggests that voice information has a profound influence on the identification, recognition and recall of spoken words. The findings reviewed below suggest that the processing of linguistic content and characteristics of a talker's voice may not be independent of each other.

Many of the early studies on the influence of talker variability on speech perception were conducted to test hypotheses about the need for normalization mechanisms in speech perception. For example, Summerfield and Haggard (1973; Summerfield, 1975) found that listeners were slower to categorize synthetic vowels when successive test items were perceived to be produced by different voices. They argued that the increases in response times were due to an automatic normalization mechanism that is engaged whenever the voice of the speaker changes. According to Summerfield and Haggard (1973), the speech perception mechanism needs to be retuned to the vocal tract of the perceived new voice and that over time, the efficiency of this tuning process increases (Summerfield, 1975).

While Summerfield and Haggard (1973) showed that response latencies were affected by talker variability in the acoustic test set, Verbrugge, Strange, Shankweiler and Edman (1976) found that the accuracy of vowel identification was also affected. They reported that identification accuracy was higher when listeners were presented with vowels produced by only a single talker, compared to a condition in which listeners heard vowels produced by multiple talkers. The perceptual advantage for vowels produced by a single voice was retained even when listeners were exposed to a subset of a speaker's vowels in a multiple talker condition prior to identifying the critical test vowel (see Strange, Verbrugge, Shankweiler, & Edman, 1976 for a conflicting result). Based on these results Verbrugge et al. argued that knowledge of a speaker's vowel space did not provide sufficient information to overcome the effects of talker variability in the stimulus set.

The studies by Summerfield and Haggard (1973) and Verbrugge et al. (1976) show that vowel identification is adversely affected by the presence of talker variability in the stimulus set (see also Assmann, Nearey, & Hogan, 1982). Fourcin (1968) and Rand (1971) have also shown that the perception of stop consonants is affected when the voice of the talker is changed from trial to trial.

Each of the studies described above has shown that perception of a single stimulus is affected by trial-to-trial variation in the voice of the speaker. Allard and Henderson (1975) and Cole, Coltheart and Allard (1974) have shown that the effects of talker variability extend to the same-different task. Both studies compared conditions in which listeners had to decide whether two auditory stimuli were examples of the same word or not. Listeners made faster "same" judgments when both tokens were produced by the same voice, compared to the condition in which both token on a particular trial were produced by two different voices. The advantage for "same" judgments was preserved even when the two items in the stimulus pair were separated by up to eight seconds.

Taken together, the studies described in this section indicate that trial-to-trial changes in the voice of the talker affect the translation of the acoustic signal into a phonetic code. These results raise the question of whether phonetic information can be processed independently of voice information. Mullennix and Pisoni (1990) addressed the issue of the perceptual independence of voice and phonetic information using a Garner (1974) speeded classification paradigm. They reasoned that if processing a talker's voice and resolving the acoustic-phonetic input were perceptually independent, then listeners should be able to selectively attend either to the voice of the talker or to the word that they were speaking. However, if voice and phonetic information are not perceptually independent, subjects should display evidence of interference from the unattended dimension. Mullennix and Pisoni found that the voice and phonetic dimensions interfered with each other in an asymmetric manner: Interference was significantly greater when listeners were asked to selectively attend to categorizing the initial phoneme of each word while ignoring any changes along the voice dimension. Mullennix and Pisoni's (1990) results suggest that the perceptual system does not treat the two sources of information as independent. Rather, the resolution of phonetic information is contingent on processing the voice of the talker. Similarly, classifying the voice of the talker is dependent upon processing the acoustic-phonetic input. These findings contradict the traditional assumption that characteristics related to a talker's voice are processed independently of information about the linguistic content of the message.

A.5.b. Isolated Words

In many of the experiments described above, listeners identified or classified individual vowels or consonants. Indeed, much of the research on the effects of talker variability has been conducted using isolated phonetic segments or nonsense syllables. However, several studies have also examined the influence of talker variability on the perception of real words. Creelman (1957) found that isolated words produced by multiple talkers were identified less accurately than the same words produced by a single voice. Mullennix, Pisoni, and Martin (1989) extended Creelman's results using a perceptual identification task to a larger set of words. In addition, Mullennix et al. also reported that naming latencies were longer for words produced by multiple talkers. They argued that the presence of multiple talkers in the stimulus set incurs a processing cost on listeners.

This processing cost may occur early on in the translation of the speech signal into a phonetic code or may be due the to the encoding of talker-specific details. The results from perceptual identification and naming do not provide a definitive statement about the precise locus of talker variability effects in spoken word recognition.

Taken together, the findings from the experiments described above suggest that listeners incur a processing cost when multiple talkers are included in the stimulus set. Identification and categorization of individual phonetic segments is slower and less accurate when the items are produced by multiple talkers. A similar result is obtained when real words are used. Furthermore, the costs incurred due to talker variability appear to extend beyond identification tasks to the same-different task and are apparent even at long interstimulus intervals. Finally, Mullennix and Pisoni's (1990) results obtained in a Garner speeded classification paradigm suggest that the recognition of spoken words is contingent upon resolving information about the talker's voice. These findings suggest that voice information may play a much larger role in speech perception and spoken word recognition than has been traditionally been assumed in the past.

A.5.c. Talker Variability Effects in Recall and Recognition

Each of the investigations reviewed above suggested that talker variability placed a processing load on listeners and that this load adversely affected the speed and accuracy of identification and classification. However, several experiments on memory suggest that voice information may be used as a cue to facilitate recall and recognition of spoken words. For example, Martin, Mullennix, Pisoni, and Summers (1989) compared free recall of word lists produced by multiple talkers to recall of word lists produced by a single talker. The principle finding from their investigation was that words from the multiple-talker lists were recalled more poorly in the primacy portion of the serial position curve than words from the single-talker lists. They argued that listeners in the multiple talker condition had to dedicate a portion of their limited capacity of processing resources to compensating for the stimulus variability in the test lists (Baddeley & Hitch, 1974). As a consequence, fewer resources were available for rehearsal and for transferring items into long-memory (Atkinson & Shiffrin, 1968; Waugh & Norman, 1965).

Martin et al.'s results suggest that variability in the stimulus set may be detrimental to free recall accuracy. However, Goldinger, Pisoni, and Logan (1991) demonstrated that talker variability could be used to improve recall under the appropriate experimental conditions. Goldinger et al. manipulated the interstimulus interval among items in lists of words produced by a single voice or by multiple voices. At short ISI's they reported that words from the single talker lists were recalled more accurately in the primacy portion of the serial position curve than words from the multiple talker lists. However, as the ISI was increased, Goldinger et al. observed a cross-over interaction in which words from the multiple-talker lists were recalled more accurately than words from the single-talker lists in the primacy portion of the serial position curve. They argued that while variability in the stimulus set may affect the initial encoding of the test items, information about the voices producing the stimuli on the list may be used as retrieved cues to facilitate recall.

The Martin et al. (1989) and Goldinger et al. (1991) investigations focused on the recall of spoken words produced by single and multiple talkers. A number of investigators have also examined recognition memory processes. Craik and Kirsner (1974) conducted one of the earliest investigations on the effects of changes in voice on recognition memory using a continuous recognition memory paradigm (Shepard & Teghtsoonian, 1961). They found that listeners were more accurate at recognizing old items when they were repeated in the same voice than when the items were repeated in a different voice. Craik and Kirsner also reported that subjects could accurately distinguish between trials in which old words were repeated in the same voice and trials in which old items were presented in a new voice. Recently, Palmeri, Goldinger, and Pisoni (1993) have extended Craik and Kirsner's (1974) results to a larger set of voices and to longer delays between repeated words. Taken together, these two studies suggest that voice information is preserved in memory for at least a few minutes and that this information can be brought to bear on the recognition of spoken words.

Geiselman and his colleagues offered one hypothesis on the nature of the representation that listeners form for a speaker's voice (Geiselman, 1979; Geiselman & Bellezza, 1976, 1977; Geiselman & Crawley, 1983). According to the "voice connotation hypothesis," different semantic interpretations are assigned to words according to the gender of the speaker. For example, the connotation of "horse" produced by a male voice may be that of a stallion, whereas the connotation of "horse" produced by a female voice may be that of a pony (Geiselman & Bellezza, 1976, 1977). One prediction of the voice connotation hypothesis is that a change in the gender of the voice should have a detrimental effect on the recognition of a repeated word, whereas a change in voice within a gender should not have such an effect. Palmeri et al. (1993) falsified the voice connotation hypothesis by demonstrating that a change of voice within a gender was just as detrimental to recognition as was a change of voices across genders. Results such as these have led a number of researchers to suggest that detailed information about the characteristics of a talker's voice are encoded into memory, rather than simply information about the gender of the speaker (Carterette & Barneby, 1975; Craik & Kirsner, 1974; Palmeri et al., 1993).

Taken together, the results of the memory experiments described above suggest that talker variability is not necessarily associated with a processing cost or deficit in the perceptual system. Rather, under the appropriate experimental conditions, voice information can be used to facilitate recall and recognition. Furthermore, characteristics related to a talker's voice are retained in long-term memory for a period of at least several minutes. This storage and use of fine perceptual details is contrary to the traditional assumption of spoken language processing that acoustic, nonlinguistic information fades rapidly from the auditory perceptual field (Crowder & Morton, 1969).

A.5.d. Effects of Talker Variability on Implicit Memory

The investigations described in the previous section were all concerned with the effects of a change in voice on an explicit measure of memory. Subjects were asked to consciously access memory in order to recall or recognize an event that may have occurred during the experimental session. Conscious or explicit measures of memory can be contrasted with implicit measures of memory (Graf & Schacter, 1985; Richardson-Klavehn & Bjork, 1988; Roediger & McDermott, 1993). In an implicit task, subjects are not asked to intentionally access a prior event. Instead, subjects respond by identifying a stimulus item masked by noise, completing the item from partial information or by making some other judgment about the stimulus token (Ellis, 1982; Franks, Plybon, & Auble, 1982; Gabrieli, Miberg, Keane, & Corkin, 1990; McClelland & Pring, 1991; Shimamura, 1986). The dependent measure in an implicit memory test is a priming score, which is derived by comparing performance on items that have been presented previously during the experimental session to items that the subjects have not seen or heard before (Schacter, 1987). The degree to which performance is better on the old items represents the amount of priming within a given condition. A positive priming score indicates facilitation of processing due to previously encoded perceptual episodes.

Several investigations have examined whether changes in a speaker's voice between a study session and a later test session have an influence on the amount of priming observed in different types of implicit memory tasks. Jackson and Morton (1984) investigated cross-modal and unimodal repetition priming using a perceptual identification task. During the encoding phase of their experiment, subjects read or listened to a set of nouns and were asked to judge whether each item was living or nonliving. During the test phase, subjects identified spoken words that were presented against a background of white noise. The results which bear on the issue of voice effects in implicit memory come from the subjects who listened to the stimuli during the study and test sessions. For these subjects, half of the items presented during the test session were produced by the same voice that spoke them during the study session. The remaining items were produced by a speaker of a different gender. Jackson and Morton reported significant priming effects: Old items were identified significantly more accurately than new items. However, the magnitude of the priming effect was not affected by a change in voice from the study session to the test session. Jackson and Morton argued that these findings were consistent with a model of word recognition that assumes abstract, context-free lexical representations.

Jackson and Morton's (1984) conclusion that changes in voice between study and test sessions do not affect word recognition may be premature. For example, Schacter and Church (1992; see also Church & Schacter, 1994) examined how repetition priming was affected by changes in voice between study and test in two tasks. First, they examined repetition effects in perceptual identification in noise. Across a variety of encoding conditions, changes in the voice of the speaker between the study and test sessions did not influence the magnitude of the observed priming effect. However, when the white noise was eliminated and the task given during the test phase was changed to auditory stem completion, Schacter and Church reported larger priming effects when the voice of the talker was preserved between the study and test sessions than when it changed. They argued that noise added to stimulus during the perceptual identification task degraded fine perceptual details related to a talker's voice and that this may have eliminated any voice effects during the test condition. In contrast, important surface details were preserved in the auditory stem completion task and this led to increased priming when the voice was held constant between study and test conditions.

In a more recent investigation, Church and Schacter (1994) reported that changes in some characteristics of a voice had an impact on repetition priming, whereas others did not. For example, in a series of implicit memory experiments using perceptual identification and auditory stem completion tasks, the authors found that repetition priming was reduced when the fundamental frequency of the talker's voice was changed between the study and test session. In contrast, changes in the amplitude had little effect on the magnitude of the repetition priming effect. Based on these results, Church and Schacter argued that some aspects of a talker's voice, such as fundamental frequency, are encoded into memory and are used to facilitate perceptual identification of repeated words.

Jackson and Morton (1984) and Schacter and Church (1992) also examined the effect of changes in voice on the magnitude of repetition priming obtained in perceptual identification. Both of these studies focused on the accuracy of identification under degraded listening conditions or conditions of incomplete information. Jacoby, Allan, Collins, and Larwill (1988) examined a different aspect of repetition priming. In addition to measuring the accuracy of identification, Jacoby et al. also measured the "fluency" of perception by having subjects make judgments about the relative signal-to-noise ratio of repeated items. Subjects were asked to transcribe sentences during a test phase and to judge how loud the signal was, relative to the background noise. Jacoby et al. reported that repeated items were perceived to be louder than the new items presented at the same signal-to-noise ratio. They argued that reinstantiating an old item increased the fluency of perceptual processing and that memory for the old items was a "tool" that could be used to have an impact on present experience.

The findings described above were obtained using many of the standard methods and comparisons for investigating implicit memory: Performance on old items was compared to performance on new items and a priming score was obtained. Furthermore, when voice characteristics were manipulated between the study and test conditions, they were manipulated in a binary manner: Voices were either the same or different between study and test. Goldinger (1992) examined the effects of changes in voice on implicit memory for spoken words in a more fine-grained manner by relating the size of repetition priming effects to the perceptual similarity of voices heard during study and test sessions. He found that the magnitude of repetition effects in perceptual identification was inversely proportional to the perceptual similarity of the voices used in the test set: Larger repetition effects were obtained when the change in voice between the study and test session was perceptually small than when the change in voice was large. Goldinger's observation that perceptual similarity influences memory performance fits well with a number of recent theoretical developments in research on long-term memory and categorization that stress the importance of similarity relationships among stimulus items (Gillund & Shiffrin, 1984; Hintzman, 1986; Nosofsky, 1986, 1987).

Taken together, the findings described in this section suggest that information about a speaker's voice plays an important role in repetition priming. Results such as those reported by Schacter and Church (1992;

Church & Schacter, 1994) and Goldinger (1992) indicate that the identification of spoken words is facilitated by reinstantiating previously experienced perceptual details. Jacoby et al. (1988) account for these increases in perceptual fluency by assuming that memory is used as an active tool, rather than a passive retention device. The assumption that memory has an active influence on perception runs counter to the traditional dogma of spoken word recognition which says that the lexicon is composed of a collection of abstract, idealized passive entries that need to be searched or activated in order for words to be recognized (Morton, 1970). Rather, these findings suggest that previous experience plays an active role in guiding spoken word recognition.

A.5.e. Effects of Training on Talker Variability

Each of the investigations described above has been concerned with processing costs and benefits associated with listening to unfamiliar voices. Several recent studies have examined changes in processing fluency as a function of gaining familiarity or experience with a set of voices. Lightfoot (1989) trained listeners to associate names with a set of unknown voices in a closed set identification task. After a single day of training, listeners' performance in identifying the voices was significantly above chance. By the end of 10 training sessions, subjects correctly identified the voice of the talker on approximately 80% of the trials. At the completion of training, subjects were given a serial recall task in which items were produced by single or multiple talkers. Subjects who were trained to identify the voices used in the memory test recalled more words than subjects who were not trained to identify the voices, particularly in the primacy portion of the serial recall curve. Lightfoot's (1989) results suggest that information about the voices of the talkers used during training was encoded in long-term memory and that this information was brought to bear on an explicit test of memory (Goldinger et al., 1992).

Nygaard, Sommers, and Pisoni (1994) conducted a similar training experiment in which listeners were trained to identify a set of 10 speakers over a ten day period. In contrast to Lightfoot (1989), however, Nygaard et al. examined implicit memory for the voices of the talkers presented during training. After training was complete, subjects identified new words spoken by the talkers used in training. A trained group of subjects in a control condition identified a matched set of words produced by a new set of talkers. Perceptual identification accuracy was higher for subjects who heard familiar voices during the test phase of the experiment. Nygaard et al.'s results indicate that listeners retain information about a talker's voice acquired during the training phase and that this information may be used to facilitate the identification of new spoken words.

A final example comes from an investigation conducted by Lively, Pisoni, Yamada, Tohkura, and Yamada (in press). They trained Japanese listeners to identify the English /r/ and /l/. After fifteen days of identification training, subjects identified new words produced by a familiar and unfamiliar voice. Tokens produced by the familiar voice were identified significantly more accurately than items produced by the unfamiliar voice. This difference in identification accuracy was obtained even when subjects were retested after six months without training. Results obtained from an untrained group of subjects showed that the differences in performance with the talkers could not be accounted for by baseline differences in intelligibility. Rather, familiarity with a voice appears to have facilitated the recognition of words produced by that talker.

Taken together, the training studies described above suggest that familiarity with a voice increases perceptual fluency. Words produced by a familiar voice are identified more accurately and recalled more easily than words produced by unfamiliar voices. Contrary to traditional assumptions about spoken language processing and normalization, familiarity with a voice does appear to facilitate spoken language processing.

A.5.f. Summary of Research on Talker Variability

The studies described in the sections above demonstrate that speech perception and spoken language processing may be served by processes other than those that are strictly used to recover the linguistic content of a speaker's utterance. Talker variability was shown to have an effect during the earliest stages of encoding and was shown to influence higher level processes such as recall, recognition and perceptual identification. In some

cases, trial-to-trial variation in the voice of the talker was shown to have a detrimental effect on processing (e.g., Mullennix et al., 1988). However, under other conditions, such as recall, recognition, or repetition priming, information about the voice of the speaker was shown to have a positive influence on performance. For example, words spoken by multiple talkers were recalled more accurately (Goldinger et al., 1991) and were recognized more often when they were repeated in the same voice (Palmeri et al., 1993). Taken together, these findings indicate that indexical information about a talker is not discarded during the earliest stages of processing. Rather, this information appears to be preserved in long-term memory and can be used to facilitate performance in a variety of perceptual and memory tasks. These findings are contrary to the traditional abstractionist assumptions made in speech perception that sources of variability, such as features of a talker's voice, speaking rate or dialect, are discarded very early in spoken language processing.

B. Contributions of Nonanalytic Cognition to Spoken Language Processing

In the preceding sections, two traditional reductionist assumptions concerning spoken language processing were outlined. The first assumption was that many of the problems in spoken word recognition could be solved by simply adapting models of visual word recognition. The second assumption was that sources of variability in the acoustic signal were noise that had to be filtered out in order to arrive at an abstract symbolic representation. Some observations about how spoken language differs from written language were reviewed in order to demonstrate that the first assumption is oversimplified. Next, a number of findings on the role of talker variability in spoken language processing were reviewed. These findings suggested that a reconceptualization of the processes involved in spoken word recognition may be necessary. In this section, we consider the influences that nonanalytic cognition and implicit memory have had on new developments in speech perception.

Jacoby and Brooks (1984) contrast nonanalytic cognition with the classical, analytical or abstractionist view of cognition. According to their analysis, the classical view of cognition holds that generalization to new items occurs by abstraction: Multiple instances of the same item are examined for common features and dimensions. This abstracted information is assumed to be relevant to future identification and categorization and is assumed to be stored in a type of stable semantic memory. Irrelevant, idiosyncratic details are assumed to be discarded during the process of perceptual analysis. Within this framework, memory serves as a passive storage device that does not play a guiding role in perception. This view conforms closely to the traditional view of speech perception and spoken language processing (Liberman & Mattingly, 1985; Studdert-Kennedy, 1976, 1980).

Whereas the analytic view of cognition relied on generalizations made across examples, the nonanalytic view emphasizes the use of the instances themselves. Information about the item itself and its surrounding context are assumed to be stored in memory (Tulving & Thompson, 1973). This contextual information extends to surface features, such as the font of written words or the voice associated with spoken words. This view emphasizes that perception, identification and categorization may operate by referring to previously stored, highly specific processing operations. According to the nonanalytic view, the processing operations that were originally used to encode a stimulus are reinstantiated when a similar object is encountered (Kolers, 1979).

Three points are important to note about the nonanalytic view. First, nonanalytic cognition emphasizes the encoding and storage of both general and contextual information. Jacoby and Brooks (1984) state "a nonanalytic procedure depends on tightly integrated combinations of definitionally relevant and irrelevant (including adventitiously correlated) information." They argue that activities, such as identification, recognition and categorization, may not rely exclusively on common features that characterize objects. Rather, each of these functions uses information about the object of perception itself and the context in which it occurs.

Second, the manner in which stable, "robust" representations are developed in a nonanalytic framework differs from the analytic perspective. Jacoby and Brooks argue that stability within an analytic framework arises

due to abstraction. Characteristic features or dimensions are recruited across examples and these abstracted characteristics form the basis for stable semantic representations in long-term memory. In contrast, stability in a nonanalytic system is achieved through the storage of multiple examples that preserve contextual information. Jacoby and Brooks argue that stability arises because perceivers tend to treat similar objects and situations in a similar manner (see also Brooks, 1987). According to their argument, problems of identification and categorization are resolved by analogy to other examples, rather than through an analysis of component features.

Third, Jacoby and Brooks (1984) emphasize the importance of the operations used to encode the stimulus, rather than the memory trace of the stimulus itself (Dunn & Kirsner, 1989; Kolers, 1976, 1979). By conceptualizing memory in terms of processing operations rather than in terms of static instances, they stress the interactive and dynamic nature of perception and memory. Jacoby and Brooks suggest that "memory for episodes is not something that can only be searched after perception of a test item but, rather, memory for episodes contributes to the perceptual identification of the test item; perception and memory are not discrete acts." According to this nonanalytic framework for cognition, perception involves recruiting stored processing operations from memory and applying these operations to the perception and identification of new items.

The nonanalytic framework outlined above stressed interactive nature of memory and perception and the need to store specific processing operations, rather than abstractions. Brooks (1978) has outlined a number of factors that might encourage subjects to engage in a nonanalytic processing strategy. Under these conditions, listeners may store individual instances of spoken words instead of context-free abstractions. The criteria used for the nonanalytic processing of visual objects can be directly applied to issues in spoken language processing (Pisoni & Lively, 1994). First, Brooks argues that subjects are likely to store individual instances when a category is composed of highly variable members. In the case of speech, physical variation from token to token can be very large due to a variety of factors related to speaker characteristics, microphone characteristics, background noise and reverberation (Klatt, 1986). Within the framework described by Jacoby and Brooks, token-to-token variation is information that is encoded along with the linguistic form of the item. Furthermore, these fine perceptual details may be actively engaged in order to identify other speech signals.

Second, Brooks (1978) argues that the storage of individual instances is useful when subjects have incomplete information. Under laboratory conditions, subjects listen to spoken language under ideal conditions: Signal-to-noise ratios are extremely high and competing signals are kept to a minimum. However, in natural listening environments, conditions are typically less than ideal. Indeed, signal-to-noise ratios may be very low and listeners' attention may be divided among several simultaneous auditory and visual signals. As a consequence, listeners may only have access to incomplete information when they attempt to encode and decode the speech signal. Early commitment to a phonemic representation, which entails a loss of stimulus information during perceptual processing, may lead to costly errors. Relevant missing features may not be able to be recovered for an abstract representation (Klatt, 1980). By encoding specific instances and their surrounding contexts, detailed information is preserved that may be used to guide processing under highly degraded listening conditions.

Third, Brooks (1978) claims that listeners encode particular instances when category membership is difficult to determine through traditional analytical procedures. Phonemes, the sounds that make up spoken language, meet this criterion for nonanalytic concepts. Speech signals encode complex category relationships whereby many potential acoustic cues are engaged by listeners to facilitate perceptual identification. Furthermore, these cues can be added, deleted, and traded with each other (Repp, 1982). Because speech perception is highly automatized and because the relevant cues may not always be obvious in the signal, the category structure of speech is not amenable to hypothesis testing. It has been extremely difficult to formalize a set of explicit rules that can successfully map speech cues onto discrete phoneme categories (Pisoni & Lively, 1994).

In summary, speech signals and phonetic categories demonstrate many characteristics of concepts that are assumed to be represented in a nonanalytic manner. These characteristics include the high variability of speech signals, their complex category structures and the failure to describe a set of explicit rules that map acoustic cues onto phonetic categories. When these characteristics are combined with some of the results described above on talker variability, it becomes clear that recent theoretical developments in nonanalytic cognition and implicit memory may be applicable to issues in speech perception and spoken language processing.

C. Transfer Appropriate Processing

The nonanalytic framework described by Jacoby and Brooks (1984) provides an alternative perspective for reconceptualizing some of the reductionist assumptions in speech perception and spoken language processing. The framework is provided to demonstrate the potential contribution that instance-based memory representations can make to our understanding of perception, identification, and categorization (Jacoby, Marriott, & Collins, 1989). However, the criteria that Brooks (1978) describes for instances of nonanalytic cognition refer mainly to stimulus properties, rather than processing operations. As such, the nonanalytic framework outlined above does not adequately address the processing demands that require subjects to encode instance-specific information. For example, it is not clear under what conditions information about a voice will be used to facilitate the perception of spoken words and when it may be irrelevant.

In contrast to the criteria described by Brooks (1978), the transfer appropriate processing framework outlined by Roediger and his colleagues is much more explicit about the types of processing operations that will be influenced by perceptual characteristics of the stimuli, such as changes in the voice of the talker between a study and test phase (Roediger, 1990; Roediger & Blaxton, 1987a, b; Roediger & Srinivas, 1993; Roediger, Weldon, & Challis, 1989). Transfer appropriate processing is based on the concept of stimulus generalization: The amount of priming or positive transfer observed should be proportional to the degree that the perceptual operations required at the time of testing reinstantiate those operations needed during the study period (Kolers & Roediger, 1984; Morris, Bransford, & Franks, 1977; Roediger, 1990; Roediger & Srinivas, 1993; Tulving, 1983).

The transfer appropriate processing framework is defined by four assumptions. First, Roediger and his colleagues assume that performance on implicit and explicit tests of memory is related to the similarity between the study and test conditions. Performance is expected to increase as the test conditions overlap with the training conditions. As noted above, the idea of similarity or overlap in processing operations plays an important role in theorizing about implicit memory, nonanalytic cognition, and explicit memory (Jacoby & Brooks, 1984; Roediger, 1990; Tulving & Thompson, 1973).

The second assumption of the transfer appropriate processing framework is that implicit or indirect tests of memory typically require different types of processing operations than explicit or direct tests. This assumption is expanded upon in the remaining two assumptions. Third, performance on explicit tests of memory is typically improved by requiring subjects to engage in elaborative or generative encoding or some type of semantic encoding. There is now a great deal of evidence which suggests that recall and recognition performance is facilitated when subjects are required to process stimulus for meaning, rather than for their surface characteristics, during an encoding session (Craik & Lockhart, 1972; Craik & Tulving, 1975; Eysenck & Eysenck, 1980; Slamecka & Graf, 1978).

Fourth, performance on tests of implicit memory is typically facilitated by preserving the surface forms of stimuli between the study and test conditions. For example, Jacoby and Dallas (1981; see also Jackson & Morton, 1984; Roediger & Blaxton, 1987b) obtained larger priming effects when the stimulus items were presented in the same modality during the study and test sessions than when the modality of presentation was changed. In addition, they also found that perceptual priming was relatively unaffected by manipulations related

to levels of processing or depth of encoding. As described above, Schacter and Church (1992; Church & Schacter, 1994) found that changes in the voice of the talker between the study and test sessions affected the magnitude of the repetition priming effect. These results suggest that repetition priming in some types of implicit memory test is influenced by changes in the surface characteristics of the items presented during the study and test phases.

The third and fourth assumptions of the transfer appropriate processing framework are closely related to the distinction that Jacoby (1983) draws between "conceptually driven" processing and "data-driven" processing (Roediger, 1990). Conceptually driven processing predominates when subjects have to derive a response from a set of semantically or conceptually related cues. For example, subjects might be asked to give a category name in response to a number of cues that are examples of that category. Jacoby (1983) argues that subjects engage in conceptually driven processing when contextual cues encourage them to generate expectations about potential stimulus items. These expectations minimize the role that perceptual details might play in recognizing or identifying a test item. Roediger (1990) and Jacoby (1983) argue that typical tests of explicit memory are guided by conceptually-driven processing.

Conceptually-driven processing can be contrasted with data-driven processing. Data-driven processing occurs when subjects must rely on perceptual analysis of the stimulus for successful identification or recognition. Contextual cues are removed in order to minimize expectations about the occurrence of a particular stimulus (Jacoby, 1983). Data-driven processing strategies are assumed to be engaged during implicit memory tasks such as perceptual identification or stem completion. In these tasks, subjects are typically given a degraded or incomplete test pattern and no conceptual cues.

The third and fourth assumptions of the transfer appropriate processing framework suggest that conceptually-driven processing typically falls within the domain of explicit memory tasks. In contrast, data-driven processing is engaged by implicit memory tasks. However, Blaxton (1989) has described experimental conditions that suggest certain tests of explicit memory may be sensitive to changes in surface forms, while other tests of implicit memory are influenced by conceptual processing. For example, she reported that subjects recalled words most accurately in a graphemic cued recall test, a test of explicit memory, when the cues matched the font of the studied items. This result represents an example of an explicit memory task that is sensitive to data-driven components of processing. Similarly, Blaxton found that answering general knowledge questions was facilitated by having subjects generate mental images of the correct responses during a prior study session. This finding suggests that the elaborative, conceptual processing that is typically discussed in relation to tests of explicit memory may also have an influence on tests of implicit memory.

In summary, the basic assumptions of the transfer appropriate processing framework emphasize the encoding and overlap of perceptual operations. Whereas Brooks (1978) stressed the importance of stimulus-related properties, Roediger and his colleagues emphasize the importance of cognitive procedures. The transfer appropriate processing framework and the analysis offered by Blaxton (1989) provides an important new perspective for reconsidering the role that information about a speaker's voice might play in the processing of spoken language. In the next section, the arguments offered by Blaxton are extended to form the basis for the experiments conducted in the present investigation.

D. Outline for the Present Investigation

According to Blaxton's (1989) analysis, the surface forms of spoken words should influence spoken language processing under some conditions, but not under others. More specifically, conditions that require data-driven processing or require listeners to carefully attend to the surface forms of spoken words should be influenced by changes in voice between a study condition and a test condition. However, when listeners engage

in more conceptual processing, changes in the voice of the talker between a study session and a test session should not have as strong an impact on the processing of spoken words.

In order to examine this dissociation in the use of information about a speaker's voice, a series of experiments was conducted in which subjects were tested in explicit and implicit memory tasks that were thought to tap data-driven or conceptual sources of information, respectively. The explicit memory tasks were surprise recognition tests given either immediately after an incidental encoding task or after a 24 hour delay. Based on previous research (Craik & Kirsner, 1974; Goldinger, 1992; Palmeri et al., 1993), subjects were expected to engage in data-driven processing which would be sensitive to changes in voice between the initial encoding phase and the test phase.

Three types of implicit memory tests were employed in the present investigation. The lexical decision task was selected as an example of an implicit task that should be sensitive to conceptual information, rather than data-driven processing. Because of the focus on conceptually derived information, the lexical decision task would not be expected to be sensitive to changes in surface forms of the stimuli between the study and test conditions. In contrast to the lexical decision task, the gating task and an auditory stem completion task were selected as examples of implicit memory tests that should be influenced by changes in the surface forms of spoken words between the study and test conditions. As noted above, Schacter and Church (1992; Church & Schacter, 1994) have shown that the auditory stem completion task is sensitive to changes in the voice of the talker. However, the effects of changes in voice have not been investigated in previous experiments using the gating task, which permits us to measure the amount of signal duration required for word recognition.

In addition to assessing the effects of changes in the voice of the talker in a variety of implicit and explicit memory tasks, we were also interested in examining the influence of similarity among voices on repetition effects. To achieve this goal, a similarity space was derived for the voices used in the present experiments. Based on findings obtained by Goldinger (1992), it was anticipated that under conditions of data-driven processing, repetition effects would be proportional to the degree of similarity between the two voices. Thus, a larger repetition effect would be expected when items were repeated in a voice that was perceptually similar to an old voice than when items were repeated by a talker who was perceptually dissimilar to an old talker.

Results indicating that changes in voice would influence performance on some tasks, but not on others, would be consistent with the transfer appropriate processing framework and Blaxton's (1989) analysis. More importantly, however, evidence suggesting the automatic encoding and use of talker-specific information would contradict traditional claims concerning perceptual normalization and the abstract nature of spoken language processing. The anticipated findings from the present investigation, taken together with results from a number of other recent studies (Church & Schacter, 1994; Goldinger, 1992; Nygaard et al., 1994; Schacter & Church, 1992), suggest that theorizing on spoken language processing needs to be reconceptualized. Rather than focusing on the abstract, context-free nature of speech perception and spoken word recognition, attention in human speech perception and spoken word recognition needs to be given to the encoding and use of specific prior processing operations during perception.

CHAPTER II: Multidimensional Scaling of Voices

Introduction

As discussed in the previous chapter, perceptual similarity has been shown to play an important role in research on implicit and explicit memory. For example, the transfer appropriate processing framework stresses the importance of the overlap between processing operations carried out during a study phase and the processing operations conducted during a test phase (Roediger, 1990). Similarity is also critically important to models of explicit memory and categorization (Gillund & Shiffrin, 1984; Hintzman, 1986; Nosofsky, 1986, 1987; Medin & Schaffer, 1978). However, in spite of the emphasis placed on similarity, much of the research conducted on distinctions between implicit and explicit memory has failed to consider similarity in a direct manner. For example, during the typical implicit memory test, performance on "old" or "same" items is often compared to performance on "new" or "different" items (see, however, Goldinger, 1992).

Because one of the goals of the present investigation was to examine the importance of perceptual similarity among voices across several implicit and explicit memory tasks, a necessary preliminary step was to obtain measures of similarity among the voices used in the stimulus set. Similarity data were collected using an AX same-different speeded classification task and were submitted to a multidimensional scaling program, KYST (Davison, 1992; Kruskal, Young, & Seery, 1973; Kruskal & Wish, 1978; Shepard, 1980;). The AX task has been used extensively as a tool to obtain similarity data in both the auditory and visual domains (Nickerson 1967, 1970; Weiner & Singh, 1974). The basic assumption with regard to perceptual similarity has been that response times to "same" trials will vary directly with the degree of perceived similarity between the two items in a stimulus ensemble: Results obtained with the speeded AX task have been shown to be highly similar to results obtained in a nonspeeded task and measures of direct similarity estimation (Sergent & Takane, 1987).

The present experiment was modeled very closely after one conducted recently by Goldinger (1992). Subjects made speeded same-different decisions to pairs of words or nonwords. Words and nonwords within a trial were always produced by different male talkers. Within a trial, words were always paired with other words, while nonwords were always paired with other nonwords. On half of the trials, two voices spoke the same item. On the remaining half of the trials, two voices produced two different items. Over the course of the experiment, subjects heard nine different male voices. Subjects were told to ignore the change in voices and to respond "same" whenever the two voices produced the same items. Latencies to correct responses during "same" trials were used as the input to several multidimensional scaling analyses.

Three issues were addressed using the results obtained from the AX procedure. First, the overall similarity among the voices in the stimulus set was assessed. These data were used to assign voices to the study and test conditions of each of the memory experiments in the following chapters. Based on the results of the multidimensional scaling analysis, talkers in the study condition were all selected to be highly similar to each other. Voices in the test conditions were selected to be highly similar or dissimilar to one of the old voices.

The second goal of collecting the similarity data and conducting multidimensional scaling was to determine if the similarity spaces for voices varied as a function of lexical status. Previous investigations that have examined the perceptual similarity among voices have used isolated words, vowels or sentences as their stimulus materials (Matsumoto, Hiki, Sone, & Nimura, 1973; Murry & Singh, 1978; Singh & Murry, 1980; Walden et al., 1978). In the present investigation, separate analyses were conducted on the scaling data collected with words and with nonwords. If similarity among voices is determined by acoustic dimensions related to talkers' voices, then the multidimensional scaling solution derived from words and nonwords should be quite similar. However, if subjects rely on lexical factors, in addition to the psychological correlates of acoustic dimensions, then the multidimensional scaling solution derived for words and nonwords may differ considerably. To our knowledge, this is an important issue that has not been studied in the past.

The third goal in conducting the multidimensional scaling analysis was to provide an interpretation of the psychological dimensions of variation among the voices. Previous investigations have revealed several physical variables that appear to be correlated with the psychological dimensions used to distinguish among voices. Singh and Murry (1978), Murry and Singh (1980) and Goldinger (1992) all reported that pitch or fundamental frequency (F0) was a critical dimension that was used to distinguish between male and female voices. Similarly, Walden, Montgomery, Gibeily, Prosek and Schwartz (1978) found that pitch and duration were the major factors used to differentiate among male voices (see also Clarke & Becker, 1969; Holmgren, 1967). Dimensions such as voice quality, age, clarity, roughness, amplitude and animation have also been cited in various studies (Walden et al., 1978; Voiers, 1964).

EXPERIMENT

Method

Subjects

Subjects in the present investigation were 220 undergraduates enrolled in an introductory psychology course at Indiana University. All subjects received partial course credit for their participation. Listeners reported no history of speech or hearing disorders at the time of testing.

Materials

Stimuli for the AX experiment were 54 words and 54 nonwords spoken be nine talkers each. All words and nonwords were monosyllabic. The words were divided into 27 high frequency words and 27 low frequency words. The mean frequency of the high frequency words was 84 occurrences per one million printed words, while the mean frequency of the low frequency items was 3 occurrences per one million words (Kucera & Francis, 1967). Nonwords words were selected from a previously generated database (Luce, 1986).

Nine male talkers with midwestern dialects recorded the stimulus items in a sound attenuated chamber (IAC Model 401A). All stimuli were digitally recorded into separate files on a VAX-3500 workstation using an Electrovoice condenser microphone (Model C090) mounted on a headset. Words and nonwords were recorded at a resolution of 16 bits and a sampling rate of 10 KHz. All stimuli were low-pass filtered at 4.8 KHz. Words and nonwords were collected in separate blocks of trials conducted during the same day. Speakers were shown an orthographic transcription of each word on a VT100 terminal and were asked to produce each item in a clear voice at a normal speaking rate. To insure that all nonwords were produced in a uniform manner, subjects were shown an orthographic transcription of the nonword and a rhyming word (BISH rhymes with DISH). Simultaneously, talkers also heard a reference voice producing the nonwords over the headset. The experimenter monitored all productions on-line and repeated any trials in which the talker made a production error.

After all of the stimulus items were recorded, silent portions preceding and following the speech signal were edited out using a digital waveform editor (Sawusch, 1991). Stimuli were equated for RMS amplitude and were downsampled to a resolution of 12 bits for playback on a PDP 11/34 computer. The intelligibility of all stimulus items was tested using native speakers of English. Each group of five listeners identified either words or nonwords produced by one of the nine talkers. Words were identified by typing responses into a computer monitor. Nonwords were identified with a naming response. Any word or nonword that was misidentified by more than one listener was rerecorded.

Procedure

Subjects in the present experiment were tested in groups of six or fewer. Stimulus presentation and response collection were controlled by a PDP 11/34 laboratory computer. All items were converted to analog form by a 12 bit digital-to-audio converter and were low-pass filtered at 4.8 KHz. Subjects listened to the test materials at a fixed level of 75 dB SPL over TDH-39 matched and calibrated headphones. Listeners sat in sound attenuated booths equipped with a two-button response box and a video monitor.

Listeners heard a total of nine voices over the course of the experiment. During every trial, two different voices were presented. Because the same voice never produced both items presented during a trial, there were 36 possible pairs of voices. Each pair of voices was presented an equal number of times during an experimental session. Every pair of voices produced three "same" trials and three "different" trials. Furthermore, the order of presentation within each pair of talkers was counterbalanced across trials. For example, given the pair of talkers T1 and T2, T1 produced the first item on three trials and T2 produced the first item on the remaining three trials.

Across experimental sessions, the same pair of talkers produced different words and nonwords. By changing the words and nonwords on a session-by-session basis, all pairs of voices produced every word or nonword an equal number of times. Consequently, the resulting similarity matrix consisted of cells that contained average response times generated over the same set of words and nonwords. This counterbalancing of items across the cells of the similarity matrix allowed for the voices to be scaled in a manner that was unconfounded by differences in items.

During "same" trials, two different talkers produced the same word or nonword. "Different" trials were generated by creating pairs of words and nonwords produced by two talkers. Words were always paired with other words and nonwords were always paired with other nonwords ("wide-folk"; "nong-cham"). Furthermore, high frequency words were always combined with other high frequency words, while low frequency words were always combined with other low frequency words. Pairs of different items were held constant throughout the experiment. Thus, "wide" was always paired with "folk," regardless of the pair of voices producing the items.

Each trial of the experiment began when a "Get Ready" prompt appeared on the monitor in front of the subjects for 500 ms. After the prompt disappeared, the first item of the AX trial was presented. After a 500 ms ISI, the second item of the pair was presented. Subjects were instructed to decide as quickly as possible if the two items were the same word or nonword, regardless of the change in voices. "Same" responses were made with the left button of the response box. "Different" responses were made with the right button. Listeners were given a total of 5000 ms to make a decision. If no response was made during this interval, the computer scored an error for the subject for that trial. After all subjects made a response or the response interval expired, a feedback light over the correct button was illuminated for 500 ms. The next trial began immediately after the feedback light was turned off. Subjects heard a total of 216 pairs of stimuli. Half of the trials were "same" trials and half of the trials were "different" trials. Subjects were given a short break after 108 trials. Each experimental session lasted approximately one hour.

Results

The response times from correct "same" responses were combined across experimental sessions to yield three similarity matrices; one for words and nonwords combined, one for words only, and one for nonwords only. Each cell in the three matrices shows the mean response time for subjects to respond "same" for a given pair of talkers. Corresponding cells above and below the diagonal of each matrix were combined to yield half-matrices which were used as the input to the multidimensional scaling routine KYST (Kruskal, Young, & Seery, 1973). Each set of data was analyzed separately.

Figure 2.1 shows the similarity matrix for the nine talkers in the condition in which words and nonwords were combined. The labels T1 through T9 along the axes of the figure correspond to each of the talkers in the experiment. The mean response time in each cell of the matrix is based upon the total number of correct responses for that pair of talkers. A total of 660 correct responses were possible for each cell of the matrix, given that there were 220 subjects in the experiment and each subject contributed three observations to each cell. The overall error rate was 2.63%.

Insert Figure 2.1 about here.

Figure 2.2 shows the corresponding output from the multidimensional scaling analysis. The labeled circles correspond to each of the voices in the stimulus set. The best fitting solution occurred with two dimensions. The stress value or badness-of-fit statistic was reduced to 0.124 for the two-dimensional scaling solution. Possible stress values range from 1.0 to 0.0. High values of stress indicate a poor fit of the scaling solution to the data. It should be noted that a higher dimensional solution to these data would be inappropriate because only nine voices were being scaled. Kruskal and Wish (1978) recommend that the number of objects to be scaled should be more than four times as large as the number of dimensions to be used. Interpretation of the dimensions of the solution will be delayed until the separate analyses of the words and the nonwords have been presented.

Insert Figure 2.2 about here.

Figure 2.3 shows the similarity matrix of response times derived from the trials in which only words were presented. Figure 2.4 shows the two-dimensional scaling solution from KYST. The labeled circles represent each of the talkers used in the experiment. Again, the best fit to the data was provided by a two-dimensional solution. The stress value for the analysis conducted on the words was 0.140. While this stress value is somewhat higher than the one reported for the overall solution, it should be noted that the solution derived for the words is based on only half as many observations as the combined solution.

Insert Figures 2.3 and 2.4 about here.

Finally, Figure 2.5 represents the similarity matrix for the talkers derived from the trials in which nonwords were presented. Figure 2.5 shows the resultant two-dimensional output from KYST. The final stress value for the solution was 0.126.

Insert Figures 2.5 and 2.6 about here.

Similarity Matrix:
Words and Nonwords

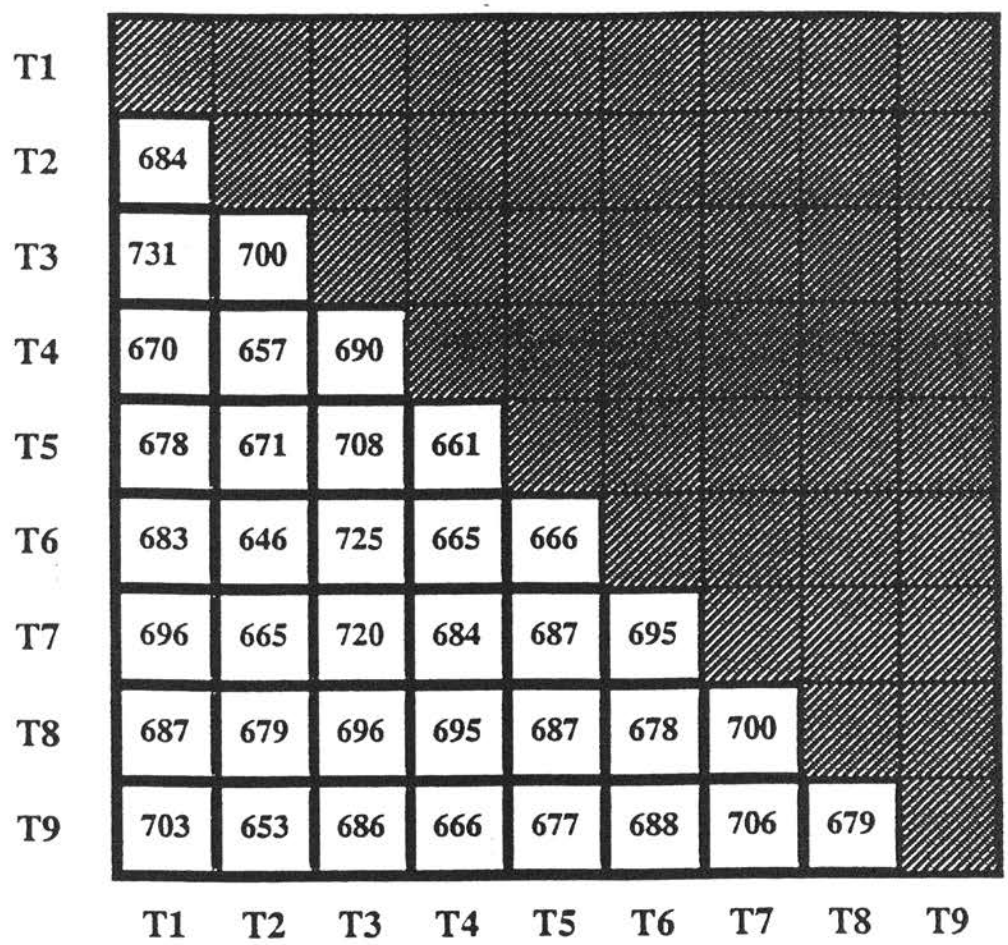

Figure 2.1. The figure displays mean response times for "same" judgements for each pair of talkers in the stimulus set.

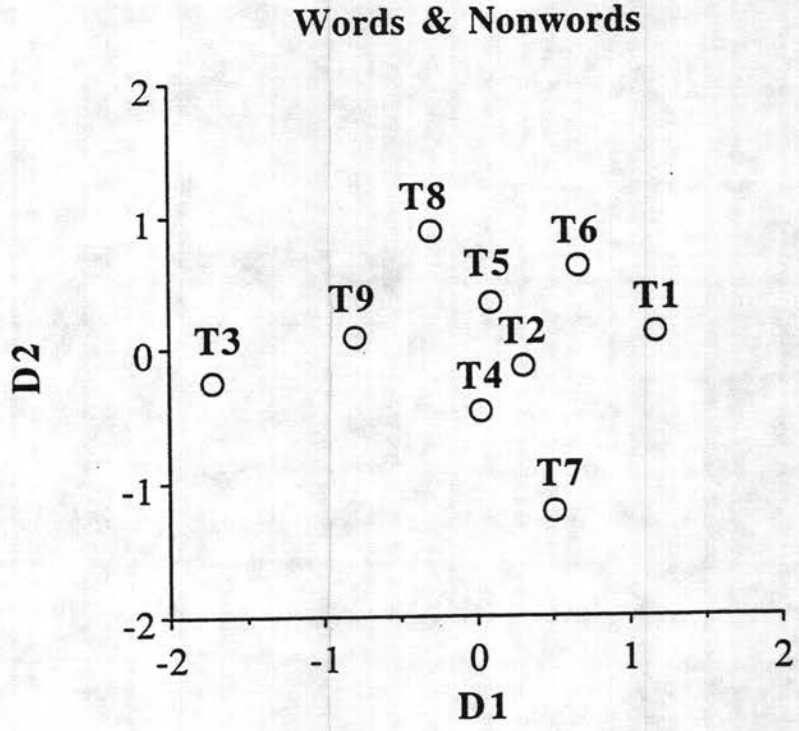

Figure 2.2. The figure displays the multidimensional scaling solution derived from the similarity matrix shown in Figure 2.1.

Similarity Matrix:
Words Only

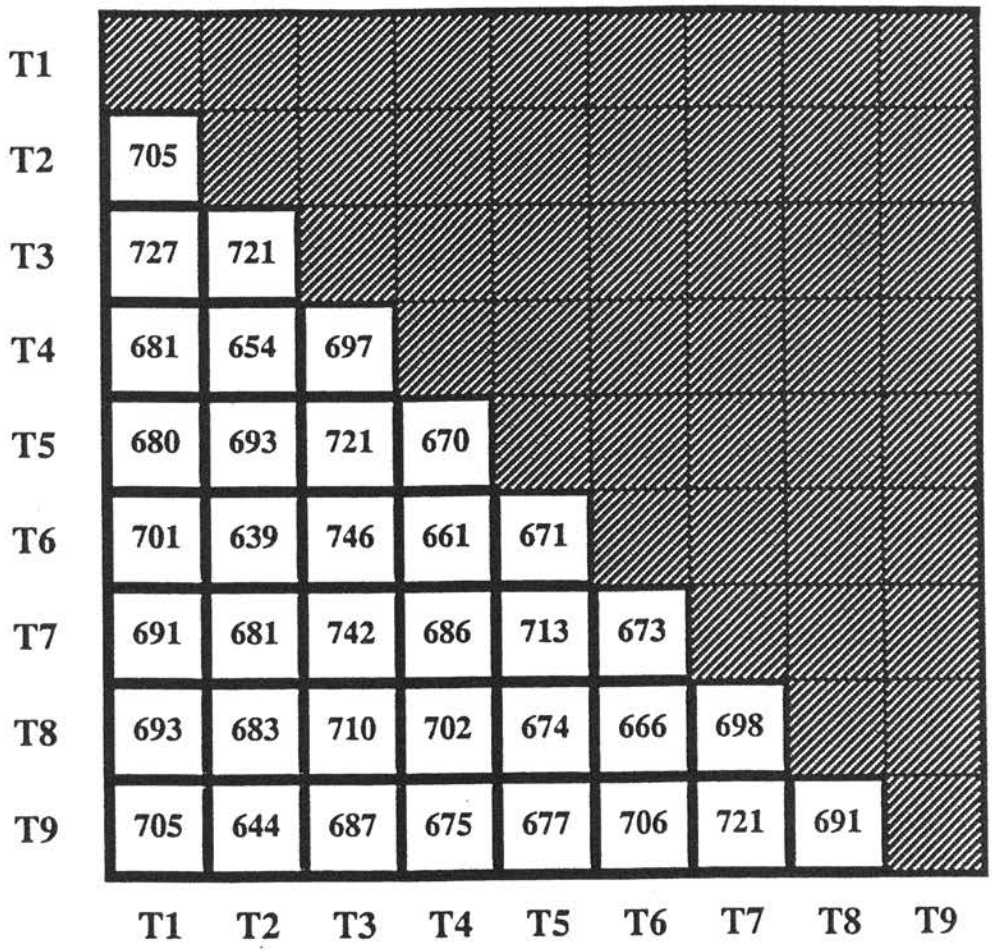

Figure 2.3. The figure shows mean response times for "same" judgements for each pair of talkers in the stimulus set. The similarity matrix was derived from trials in which only words were presented.

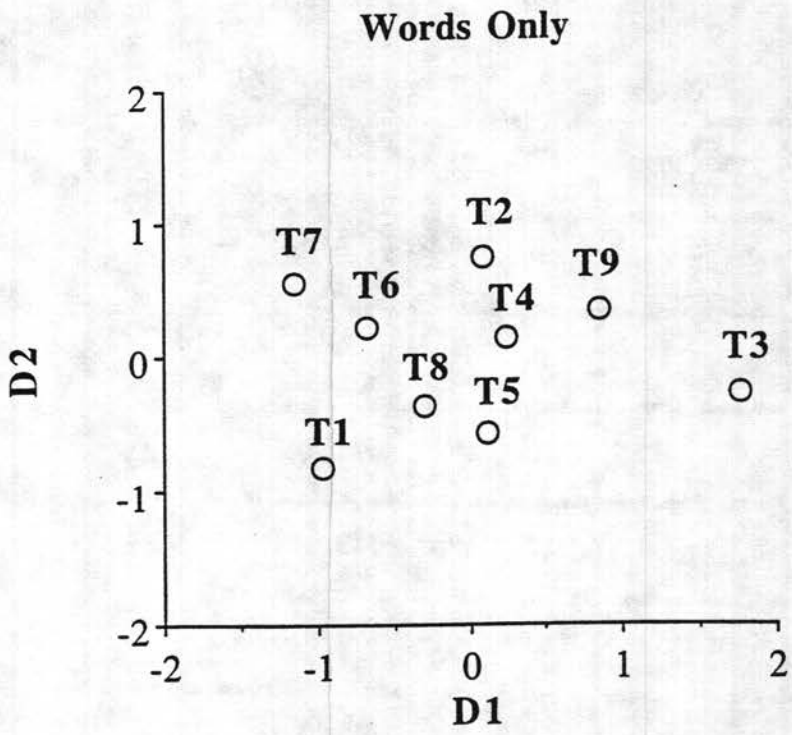

Figure 2.4. The figure shows the multidimensional scaling solution derived from the similarity matrix shown in Figure 2.3.

Similarity Matrix:
Nonwords Only

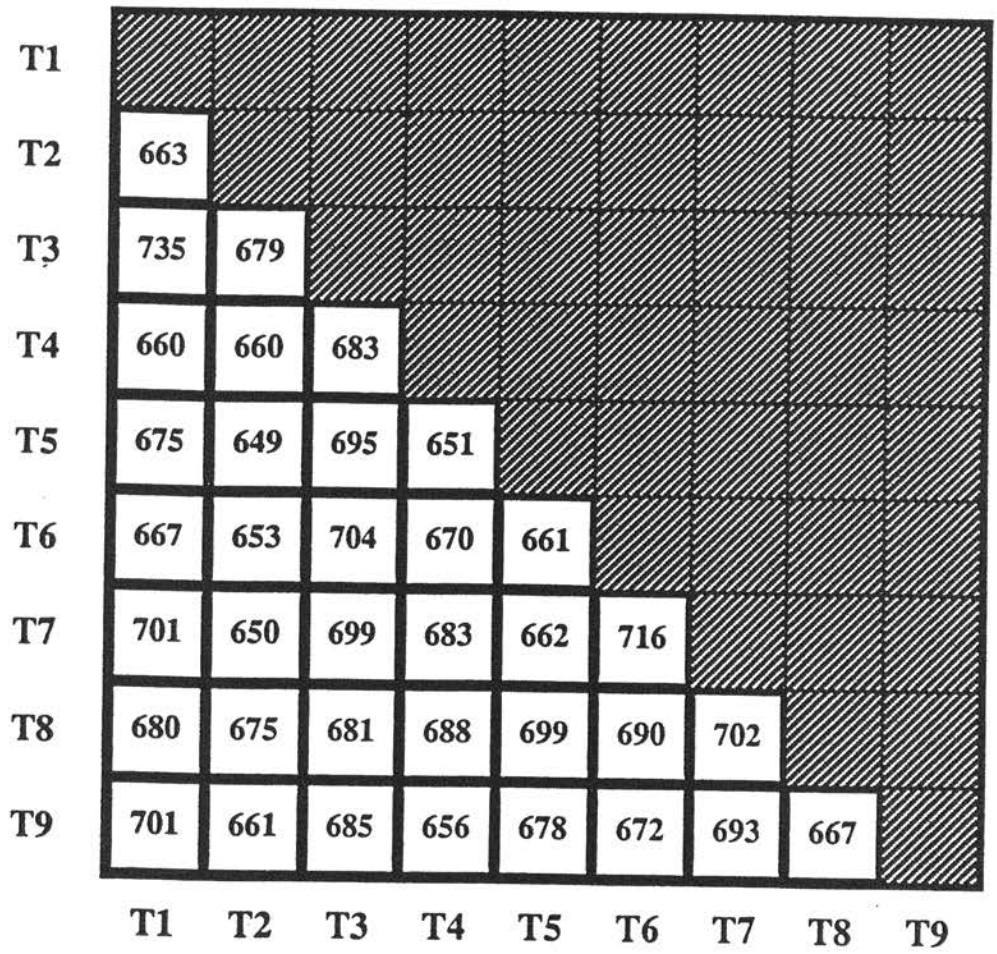

Figure 2.5. The figure shows mean response times for "same" judgements for each pair of talkers in the stimulus set. The similarity matrix was derived from trials in which only nonwords were presented.

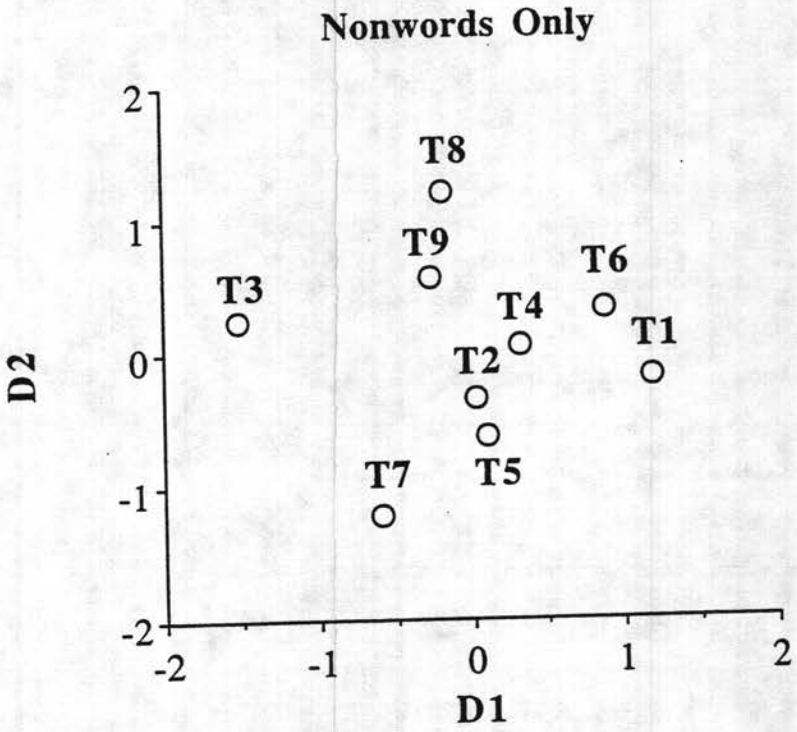

Figure 2.6. The figure shows the multidimensional scaling solution derived from the similarity matrix shown in Figure 2.5.

Discussion

The present experiment was conducted to assess the degree of perceptual similarity among a group on nine male native speakers of English with midwestern dialects. Similarity data were collected using response time measures in a "same-different" task and were analyzed using multidimensional scaling techniques. The scaling solution derived from the similarity data was used to achieve three goals. The first concerned the use of the scaling solution as a means for selecting voices to be used in the study and test phases of the remaining experiments of this report. The scaling data allows for voices to be selected in such a way that all of the voices used during the study sessions are similar to each other. Furthermore, voices for the test sessions can also be selected so that they are either similar or dissimilar to voices heard during the study sessions.

The second issue addressed by the present experiment concerned whether the similarity spaces for voices were similar when subjects listened to words or nonwords. Examination of Figures 2.4 and 2.6 suggest that the same underlying perceptual dimensions were employed when subjects hear words and nonwords. Euclidean distances between pairs of voices in each figure were moderately correlated with each other (r=.429). In general, the two figures are mirror images of each other. For example, T1 and T6 lie fairly close together in both figures, while T3 tends be an outlier on Dimension 1 in both figures. In addition, talkers T2,T4,T5, and T6 tend to cluster together along Dimension 2. The similarity of the two solutions suggests that listeners are using the same voice information when responding to words and nonwords. Thus, lexical contributions do not appear to influence the perceptual similarity among voices.

The final issue addressed by the present scaling experiment concerned the psychological dimensions listeners used to differentiate among the voices. In order to gain some insight into this issue, several gross acoustic analyses were conducted on the stimulus items. First, the durations and variability in durations across items were obtained for each talker. Neither measure appeared to account systematically for either of the two dimensions. This failure was not unexpected: All of the words and nonwords were monosyllables and each of the talkers was instructed to speak at a normal speaking rate. The mean duration of the fastest talker's utterances was 470 ms, while the mean duration of the slowest talker's utterances was 538 ms. Previous investigations that have found duration to be a critical dimension have tended to use fewer items, more voices and longer utterances, such as sentences (Clarke & Becker, 1969; Walden et al., 1978).

A second analysis was conducted to determine the average fundamental frequency (F0) and range of pitch for each talker. This analysis was conducted using the ILS signal processing package running on a VAX-3500 workstation. Again, neither measure mapped onto the dimensions of the multidimensional scaling solution. The failure of pitch to account for one of the dimensions of the scaling solution is somewhat surprising. Fundamental frequency is the characteristic dimension that differentiates male and female voices (Goldinger, 1992; Murry & Singh, 1980). Pitch has even been found to be a critical factor that is used to differentiate among male voices (Walden et al., 1978). In the present investigation, however, the differences in mean fundamental frequency among talkers were quite small. The highest mean fundamental frequency was 130 Hz, while the lowest mean fundamental frequency was 109 Hz. The range of variation in fundamental frequency was also very similar across talkers.

Loudness represents a third possible perceptual dimension for the present scaling solution. However, this attribute can be effectively ruled out, *a priori*, because all of the tokens were equated for RMS amplitude prior to the experiment. Thus, the peak loudness of talker and token was the same. Previous studies by Voiers (1965) and Holmgren (1967) have cited amplitude as a critical dimension of variation for voices.

Given that the three simplest acoustic dimensions fail to map onto the observed solution, several more esoteric and less easily quantified dimensions need to be considered. Phenomenologically, voices lying along Dimension 1 appear to be varying in their breathiness/hoarseness. Talkers T1 and T6 represent the least breathy

or hoarse voices, while T3 and T9 represent the most breathy or hoarse voices. Murry and Singh (1980) reported that breathiness and hoarseness were two dimensions that were important for discriminating among male and female voices and for making discriminations among voices of the same gender. Although the appropriate acoustic analyses to support the claim that dimension 1 is related to breathiness or hoarseness are beyond the scope of the present investigation, several previous studies have described increases in the periodicity of the fundamental frequency that may be related to the perception of breathy or hoarse voices (Muta & Baer, 1988; Sander & Ripich, 1983; Scherer, Titze, Rapael, Wood, Ramig, & Blager, 1987; Yanigahara, 1967).

While Dimension 1 may be related to the perceived breathiness of the voices, Dimension 2 may be related to the precision of articulation of the speaker (Walden et al., 1978). T8 appears to have articulated the stimulus items the most carefully and clearly, while T7 appears to have produce the items the least clearly. This interpretation is supported by the finding that during the initial testing of intelligibility, T8 was required to rerecord the fewest number of items, while T7 was required to rerecord the most. One problem with interpreting Dimension 2 in terms of precision of articulation is that it is unclear what set of acoustic attributes should be measured in order to reinforce this interpretation (Walden et al., 19878; Voiers, 1964). Analyses that examine factors such as phonological reduction, deletion and the release of final stop consonants may provide some insight on this issue (Byrd, 1992, 1993). However, such analyses are beyond the scope of the present investigation (see Bradlow & Pisoni, 1994).

Taken together, the results from the present investigation provide a principled means for assigning voices to the study and tests conditions of the following experiments. Furthermore, we found that listeners appear to be using the same dimensions to differentiate among voices when listening to words and nonwords. Finally, although the dimensions of the multidimensional scaling solution may be difficult to interpret, it should be pointed out that a definitive interpretation is not required for the purposes of the following experiments.

CHAPTER III: Perceptual Similarity in Implicit and Explicit Memory: Lexical Decision vs. Recognition Memory

Introduction

The primary purpose of the experiments reported in this chapter was to investigate a dissociation in the use of talker-specific information across implicit and explicit memory tasks. Subjects participated in a common study phase and were then given either an implicit or explicit memory test. During the study condition, subjects performed a lexical decision task. Stimulus materials were either multiple repetitions of words and nonwords produced by a single talker or words and nonwords produced by seven different talkers. The talker manipulation was included in the study phase to assess whether the amount of variability in the stimulus set would have any effect on the magnitude of the repetition effects observed during the test condition. During the test phase, subjects either made lexical decisions as an implicit test of memory or participated in an recognition memory task as an explicit test of memory. Test materials were produced by one of the voices heard during the study session or two new voices. One of the new voices was perceptually similar to the old voice, based on the multidimensional scaling solution derived from the previous experiment. The other new voice was perceptually dissimilar to the old voice. This direct manipulation of perceptual similarity allowed a closer examination of the effects of changes in voice on repetition effects in lexical decision and recognition memory.

According to the transfer appropriate processing framework and findings reported by Blaxton (1989), lexical decision and recognition memory tasks require subjects to access memory using different forms of information in order to successfully complete the task. If subjects access memory in different ways, changes in the voice of the talker between the study and test conditions may have differential effects on subjects' speed and accuracy in performing the two tasks. Consider the role of voice information in repetition effects during lexical decision. According to the theoretical framework described by Blaxton (1989), the lexical decision task stresses conceptual processing. As a consequence, listeners should be relatively insensitive to changes in voices between the study and test sessions.

The absence of voice effects in lexical decision can be contrasted with predictions concerning effects of changes in voice during the recognition memory test. As discussed in the introduction, a number of studies have found that old items are recognized more accurately when they are repeated in the same voice (Craik & Kirsner, 1974; Palmeri et al., 1992). This sensitivity to the surface forms of the stimuli suggests that, under some conditions, subjects may rely on data-driven processing to facilitate recognition memory performance. Under these conditions, surface information, such as acoustic features of a talker's voice, may serve as additional cues for the retrieval of information in long-term memory.

Because the voices used in the test phase of the present experiment were selected on the basis of their similarity to each other, a more fine-grained analysis can be conducted on the effects of changing the voice of the talker during the recognition memory test. Goldinger (1992) found that recognition memory performance was moderately correlated with the perceptual similarity among a set of male and female voices. It was of interest in the present experiment to examine recognition memory as a direct function of the similarity of the voices presented during the study and test sessions. Based on Goldinger's results, we would expect that the speed and accuracy of the recognition process should be graded according to how similar the voices heard during the test condition are to voices heard during the study session. For example, words produced by an old talker should be recognized most accurately. However, test items produced by a voice that is a similar to an old talker should be recognized more accurately than stimuli produced by a voice that is dissimilar to the old talker.

In addition to considering a dissociation in the use of voice information across an implicit and explicit memory task, several additional predictions within each task were also addressed. A number of previous investigations have examined repetition effects in lexical decision (McKoon & Ratcliff, 1979; Scarborough,

Cortese, & Scarborough, 1977). One consistent finding has been a reduction in the word frequency effect for repeated high and low frequency words. High frequency words benefit from a small savings in response times during repetition. In contrast, low frequency words receive much more facilitation from repetition. This greater benefit for repeated low frequency words leads to a reduced word frequency effect. In the present investigation, it was anticipated that the reduction in word frequency effects would be obtained for repeated test words.

Repetition effects for nonwords in lexical decision are more problematic. Scarborough et al. reported facilitation for repeated nonwords: Lexical decision response times to repeated nonwords were faster than to nonwords that had not been presented before. Based on results from both lexical decision and naming, Scarborough et al. argued that the facilitation was due to increased efficiency in encoding the nonwords. In contrast to the finding reported by Scarborough et al., Forbach, Stanners, and Hochaus (1974) found no repetition effect for nonwords and McKoon and Ratcliff (1979) reported inhibition in lexical decision for repeated nonwords. McKoon and Ratcliff (1979) and Feustel, Shiffrin, and Salasoo (1983) argued that the inhibition they observed was due to a confounding between task and stimulus factors: The first time a nonword is presented, no lexical representation for the item can be accessed from memory and the appropriate response is "nonword." However, if information about the nonword, such as its orthographic or phonological form, is encoded during the first presentation, an episodic representation of the stimulus can be accessed during a subsequent repetition. Under these conditions, the nonword has taken on some "word-like" properties and the appropriate response is ambiguous (Feustel et al., 1983). Under these conditions, inhibition in lexical decision response latencies would be expected. Given these mixed results, it is not clear what predictions can be developed for repetition effects in nonwords under the present testing conditions using spoken words.

The effects of word frequency on recognition memory have been investigated quite extensively over the years (Glanzer & Adams, 1985, 1990; Glanzer, Adams, & Iverson, 1991; Glanzer & Bowles, 1976; Gregg, 1976; Hintzman, Caulton, & Curran, 1992). The typical finding is that recognition memory for old low frequency words is better than recognition memory for high frequency words. Furthermore, the false alarm rate for new low frequency words is lower than for new high frequency words. One explanation of the advantage for low frequency words has been to assume that the distributions of old and new items are more widely separated for low frequency words than for high frequency words (Gillund & Shiffrin, 1984). In terms of a signal detection framework, low frequency words are more discriminable than high frequency words.

Recognition memory for nonwords has not been investigated as extensively as recognition memory for high and low frequency words. Some findings suggest that the recognition of nonwords may be similar to the recognition of very low frequency words or words that the subjects may not know (Gillund & Shiffrin, 1984; Mandler, Goodman, & Wilkens-Gibbs, 1982; Rao, 1983; Rao & Proctor, 1983; Zechmeister, Curt, & Sebastian, 1978). Unfamiliar words or nonwords are recognized more poorly than low frequency words that the subjects are familiar with and may even be recognized more poorly than high frequency words. Given these findings from the literature, we predicted that recognition memory for nonwords encoded during the lexical decision task in the study phase would be very poor.

In summary, the present experiments assessed several critical predictions concerning the role of information related to a speaker's voice in tests of implicit and explicit memory. First, changes in the voice of the talker should not have any effect on lexical decision response times for old items because the task stresses the use of conceptual information. Second, in contrast to the lexical decision results, recognition memory should be affected by changes in the surface forms of spoken words between the study and test conditions because information about a talkers voice may serve as a reliable retrieval cue to facilitate recognition. Furthermore, the speed and accuracy of recognition may be scaled according to the familiarity and similarity of the voices used in the test set. This dissociation in the use of voice information between implicit and explicit memory tasks would be consistent with the transfer appropriate processing framework and the theoretical analysis proposed by Blaxton (1989). Third, a larger repetition effect should be obtained in lexical decision for words than for

nonwords. Finally, an advantage for low frequency words over high frequency words and nonwords should be observed in the recognition memory data.

EXPERIMENT 1A: Lexical Decision

Method

Subjects

Three sets of subjects were recruited from the volunteer subject pool at Indiana University for participation in Experiment 1A. All subjects were native speakers of English who were enrolled in an introductory psychology course. Listeners reported no history of a speech or hearing impairment at the time of testing. Thirty-five subjects participated in the control condition. Thirty-four subjects were assigned to the multiple-talker condition. An additional 35 subjects were assigned to the single-talker condition.

Materials

The stimuli for the present experiment were 48 nonwords and 96 words produced by each of the nine male speakers used in the previous experiment. The nonwords were a subset of those used in the scaling experiment. Fifty-four of the words were also drawn from the previous experiment. An additional 42 words were collected from each of the talkers to increase the total number of words to 96. Twenty-one of the new items were high frequency words; the remaining 21 new stimuli were low frequency words. These tokens were collected at the same time and under the same conditions as the original stimuli. The same signal processing techniques were applied to the new test materials.

In addition to the test materials provided by the nine talkers described above, another seven repetitions of the 96 test items were produced by talker T2. These items were recorded and processed under the same recording conditions described in the preceding experiment. The intelligibility of these tokens was also assessed by several groups of native speakers of English. Unintelligible tokens were replaced with new tokens. The set of items repeated by T2 provided the materials for the single-talker condition.

Talker Assignment

The nine talkers were divided into a set of voices to be used during the study phase and a set of voices to be presented during the test phase. Talkers T1, T2, T5, T6, T7, T8 and T9 were assigned to the study session in the multiple-talker condition based on their clustering in the multidimensional scaling analysis derived in the previous investigation. Talkers T2, T3, and T4 were selected for the test sessions. Talker T2 was selected to overlap between the study and test lists because his voice was near the center of the perceptual space. Talker T4 was selected for the test list because his voice was perceptually similar to T2's voice. Finally, talker T3 was used in the test list because of his dissimilarity to T2, the old voice from the study session. Thus, seven talkers were assigned to the study condition and three talkers were assigned to the test condition.

Procedure

Subjects were randomly assigned to one of three experimental conditions. Approximately one-third of the subjects participated in the "multiple-talker study" condition, one-third of the subjects were assigned to the "single-talker study" condition and the remaining subjects were assigned to the control or "test only" condition. The control condition was included to insure that any differences among talkers observed in the two experimental conditions could not be accounted for by *a priori* differences among the voices, such as differences in durations. Listeners participated in groups of six or fewer and sat in sound attenuated cubicles equipped with

a two-button response box and a pair of TDH-39 matched and calibrated headphones. A PDP 11/34 laboratory computer controlled stimulus presentation and response collection on-line. Subjects in the experimental conditions participated in both a study-phase and a test-phase. The study phase consisted of 504 trials of a lexical decision task. As noted above, subjects were assigned to either a single-talker study condition or a multiple-talker study condition. The test phase was composed of 144 additional trials of lexical decision. Listeners in the control condition only participated in the test phase.

Study Condition

The list of words and nonwords was divided into two halves. One set of items was presented during the study session and was treated as the "old" stimuli during the test session. The remaining words and nonwords were only presented during the test condition and were treated as "new" items. The two sets of materials were counterbalanced across groups of subjects so that every stimulus served as an old item and as a new item for an equal number of subjects.

During the study phase, subjects made a lexical decision response after hearing each word or nonword. All test materials were produced by each of the talkers assigned to the study condition. Thus, subjects heard seven repetitions of each item during the study phase of the experiment. Each experimental trial of the study phase began when a cue light on the response box was illuminated for 250 ms. Subjects heard one of the words or nonwords assigned to the study list and were then asked to decide as quickly and accurately as possible if the stimulus was a word or a nonword. The timing of responses was started at the onset of the stimulus. "Word" responses were made with the left hand and "nonword" responses were made with the right hand. Subjects had a maximum of 4 sec to make a response before the computer scored an error for the trial. After responses were collected for all subjects, a feedback light over the correct response button was illuminated for 250 ms. The next trial began once the cue light was extinguished. The study session consisted of 504 trials. Subjects were given a short break after 252 trials. Subjects were given a longer break between the conclusion of the study session and the beginning of the test phase.

Test Condition

The test phase of the experiment was identical for listeners in all of the study conditions (single-talker, multiple-talkers, control). Subjects performed lexical decision during the test phase of the implicit memory condition. Stimulus materials were either "old" or "new." Old items were words and nonwords that had been presented during the study phase of the experiment. New items were test stimuli that had not been heard before during the experiment. Old and new words and nonwords were produced in one of four talker conditions. Half of the test materials were produced by T2, one of the voices heard during the study phase. These stimuli were divided into two conditions. Half of the items were exact repetitions of words and nonwords heard during the study phase or their matching new items that had not been heard during the study phase. This was the "exact repetition" condition. The other half of the stimuli produced by the old talker were new productions of words that were either heard during the study and test sessions or only during the test session alone. This was the "familiar repetition" condition. One-fourth of the remaining items were produced by T4, the voice that was perceptually similar to T2. This was the "similar repetition" condition. The remaining stimuli were produced by T3, the voice that was dissimilar to T2. This was the "dissimilar repetition" condition. Stimulus words and nonwords were counterbalanced across subject groups so that every word and nonword was produced by each voice in the test phase an equal number of times.

The procedure for the test phase of the experiment was identical to the study phase. Listeners participated in 144 trials and each test session lasted approximately 15 minutes.

Results

The critical data for the present experiment come from the test conditions. The major results concern the influence of changes in voice on repetition effects during lexical decision: Based on theoretical analyses offered by Blaxton (1989), it was anticipated that changes in voice between the study and test session would not have an impact on the magnitude of the repetition effect. In addition to examining the effects of changes in voice, word frequency effects were also investigated in the present study.

Data from the control condition, single-talker study condition and multiple-talker study condition were analyzed separately using an analysis of variance. Separate analyses were conducted on the mean error rates and mean response times for correct responses for each subject in each condition. In each analysis, repetition status (old vs. new), lexical status (high frequency vs. low frequency vs. nonword), and talker (exact vs. familiar vs. similar vs. dissimilar) were treated as within-subjects variables. All post-hoc comparisons were conducted using Tukey's HSD tests.

Control Condition

Mean response times and response accuracy varied as a function of talker in the control condition [$F_{rt}(3,99)=15.45$, $p<.01$; $F_{err}(3,99)=13.37$ $p<.05$]. Although Tukey's tests failed to localize the differences among talkers in the analysis of the response times, subjects tended to be fastest when responding to the tokens produced by the old talker taken from the training session (T2=927.5 ms) and slowest when responding to the dissimilar talker (T3=981.5 ms). Error rates were lowest for the new talker who was similar to the old talker (T4=16%) and highest for the dissimilar talker (T3=20%).

Responses were significantly faster and more accurately to high frequency words (873 ms, 4% error) than to low frequency words (943 ms, 25% error) or nonwords (1030 ms, 23% error) [$F_{rt}(2,66)=86.18$, $p<.01$; $F_{err}(2,66)=106.5$, $p<.01$]. Lexical decisions to low frequency words were faster than to nonwords. However, error rates did not vary between responses to low frequency words and to nonwords.

Single -Talker Condition

Listeners were faster and more accurate when responding to old words and nonwords [old items: 944 ms, 11% error; new items: 1003 ms, 18% error; $F_{rt}(1,34)=35.16$, $p<.01$; $F_{err}(1,34)=55.50$, $p<.01$]. Response latency and accuracy also varied as a function of lexical status [$F_{rt}(2,68)=79.29$, $p<.01$; $F_{err}(2,68)=72.25$, $p<.01$]. Listeners were faster and more accurate in making lexical decisions to high frequency words than to nonwords (high frequency: 902 ms, 4% error; nonwords: 1064 ms, 18 % error). Responses to low frequency words were significantly faster than response to nonwords (low frequency: 956 ms). Finally, error rates for low frequency words were higher than for high frequency words (low frequency: 21% error)

Response times and error rates varied as a function of voice [$F_{rt}(3,102)=9.38$, $p<.01$; $F_{err}(3,102)=6.05$, $p<.01$]. Figure 3.1 shows response times and error rates as a function of talker in the test condition. The light hatched bars show performance from the subjects who participated in both the study and test conditions. The dark hatched bars show performance from the control condition only for purposes of comparison. Although Tukey's tests failed to localize the differences among talkers, responses were fastest when subjects heard exact repetitions of tokens presented during the study phase. Responses were slowest when subjects heard stimuli produced by the talker who was dissimilar to the voice heard during the study phase. Listeners also showed the most errors when listening to the dissimilar voice.

Insert Figure 3.1 about here.

Significant interactions between repetition status and lexical status were obtained in the analyses of the response time and error data [$F_{rt}(2,68)=24.60$, $p<.01$; $F_{err}(2,68)=20.32$, $p<.01$]. The upper panel of Figure 3.2 displays the interaction in the response time results. The lower panel shows the interaction in the error results. The light hatched bars show items repeated from the study phase. The dark hatched bars show items presented only during the test session. Post-hoc analyses of the response time results showed an attenuated word frequency effect. Old high frequency words were responded to faster than old nonwords. However, the difference between old high frequency words and old low frequency words did not approach significance. A similar pattern was obtained for the new words and nonwords. In addition, a significant repetition effect was obtained for the low frequency words: Old low frequency words were responded to faster than new low frequency words.

A similar pattern was obtained in the analyses of the error rates. Lexical decisions were most accurate to old and new high frequency words. Error rates to new low frequency words were significantly higher than to new nonwords or to old low frequency words.

Insert Figure 3.2 about here.

Finally, a significant interaction was obtained in the analysis of the error results between the talker and repetition status variables [$F(3,102)=4.15$, $p<.01$]. Listeners tended to have higher error rates for new words and nonwords when they were either exact repetition of items that were presented during the study phase and when the stimuli were produced by the unfamiliar talker.

Multiple-Talker Condition

As in the single-talker condition, listeners were faster and more accurate when they responded to old items than to new items [old: 973 ms, 13% errors; new: 1047 ms, 22% errors; $F_{rt}(1,33)=51.75$, $p<.01$; $F_{err}(1,33)=60.92$, $p<.01$]. Responses were faster to high and low frequency words than to nonwords (high frequency: 935 ms; low frequency: 979 ms; nonwords: 1117 ms; $F_{rt}(2,66)=62.82$, $p<.01$). Error rates were lower for high frequency words than for low frequency words or nonwords (high frequency: 8% errors; low frequency: 22%; nonwords: 22%; $F_{err}(2,66)=48.31$, $p<.01$).

Response latencies and error rates also varied as function of talker [$F_{rt}(3,99)=3.55$ $p<.01$; $F_{err}(3,99)=7.18$, $p<.01$]. The upper panel of Figure 3.3 shows response latencies as a function of voice in the multiple talker test condition. The lower panel shows error rates. The light hatched bars show results from the subjects who participated in the study phase and the test phase of the experiment. The dark hatched bars show results from the control condition. The results from the control condition are identical to those displayed in Figure 3.1. Latencies tended to be shortest for the talker heard during the study phase (T2) and longest for the dissimilar talker. Error rates were almost equivalent for the old talker and the similar talker and were slightly higher for the dissimilar talker.

Insert Figure 3.3 about here.

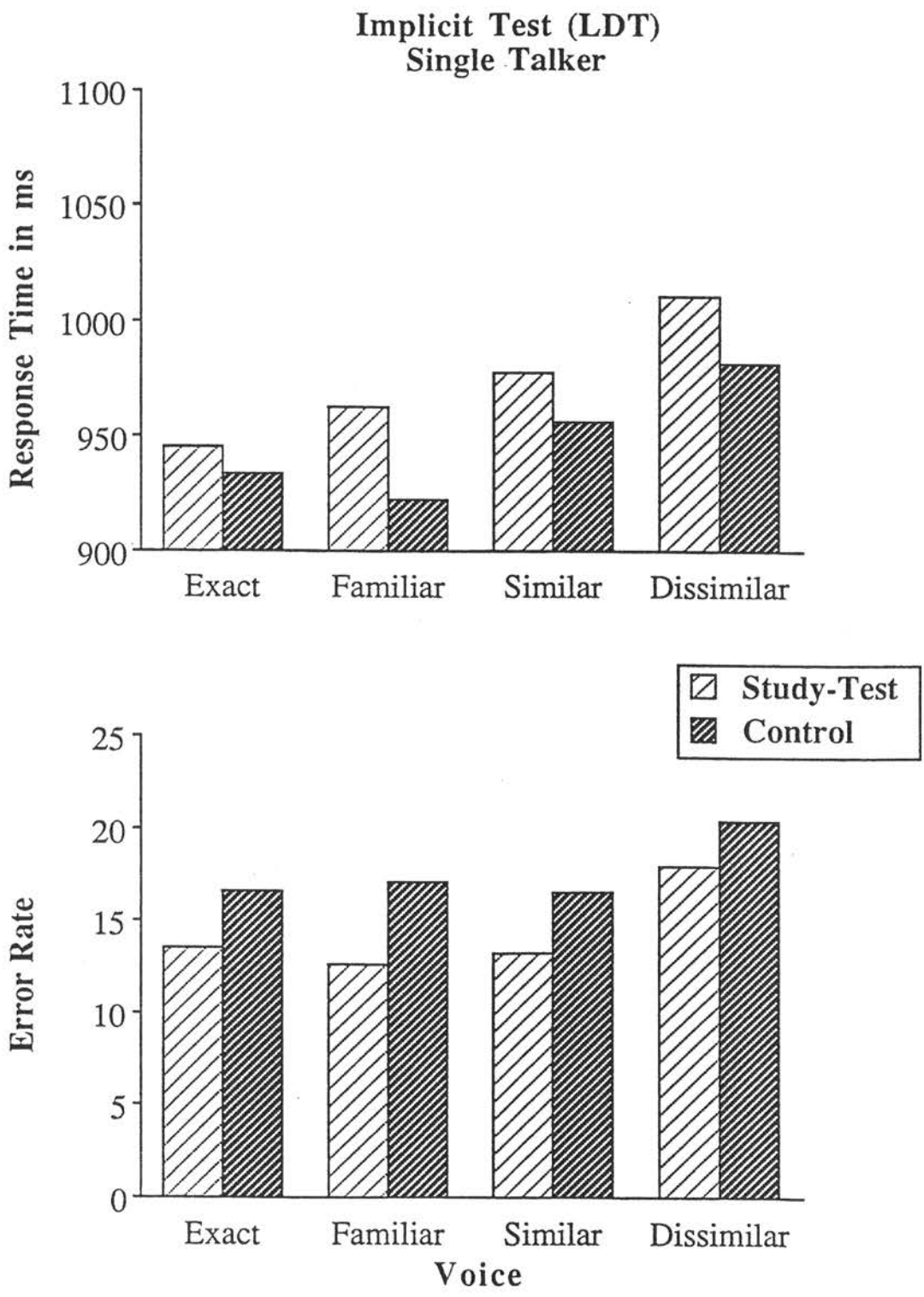

Figure 3.1. The upper panel shows mean response times to each voice in the lexical decision test in the single talker study condition. The lower panel shows error rates.

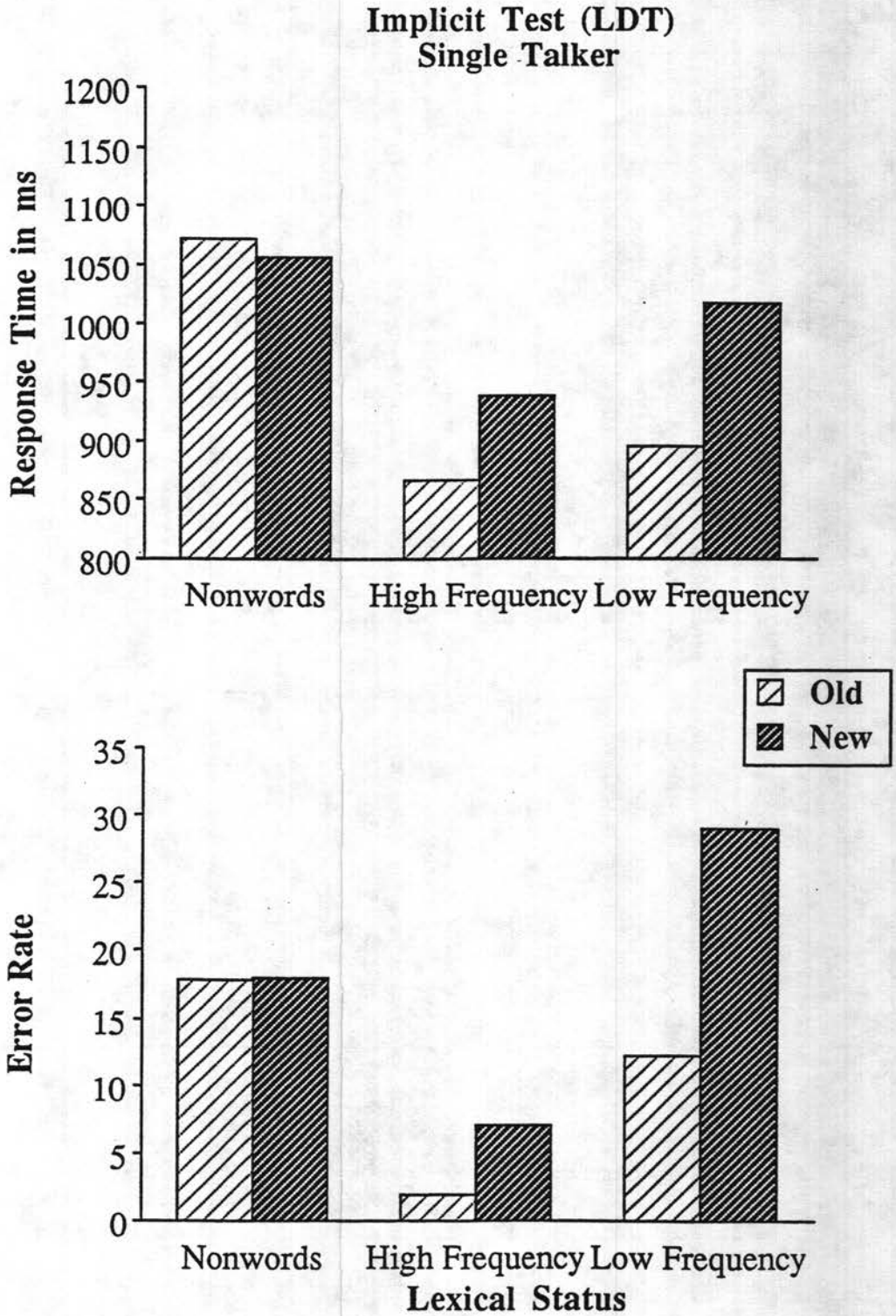

Figure 3.2. The upper panel shows mean response times to old and new high and low frequency words and nonwords in the lexical decision test in the single talker study condition. The lower panel shows error rates.

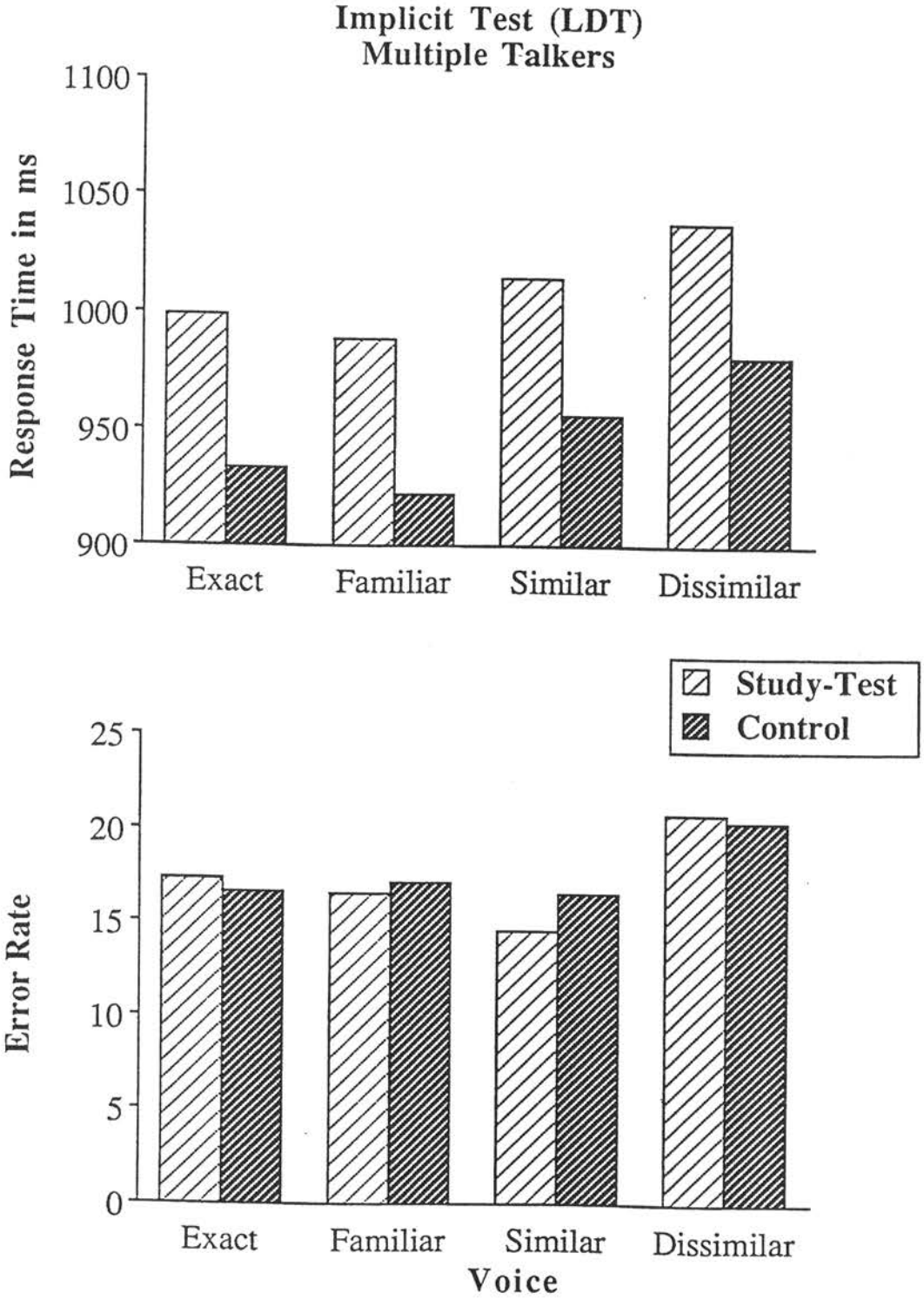

Figure 3.3. The upper panel shows mean response times to each talker in the lexical decision test in the multiple talker study condition. The lower panel shows error rates.

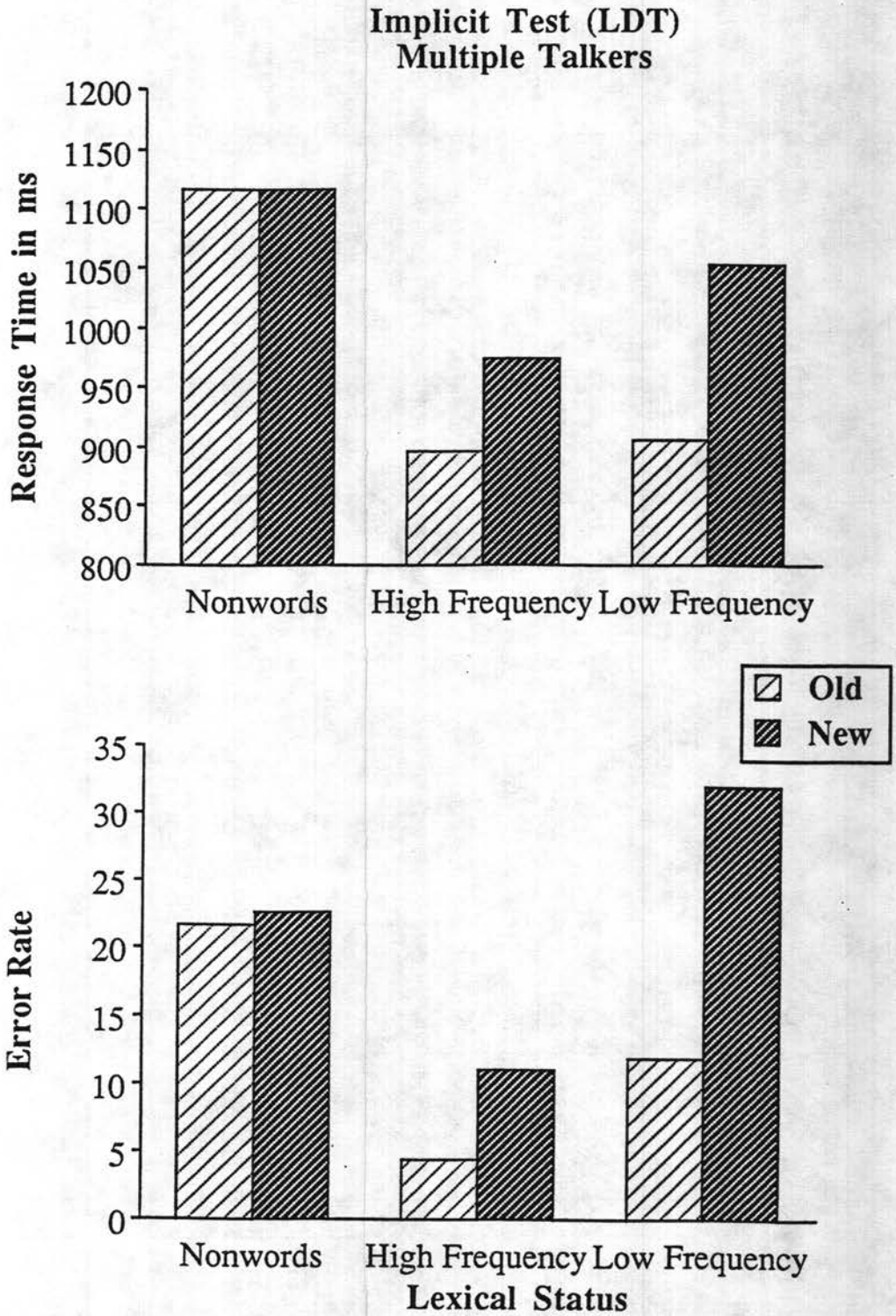

Figure 3.4. The upper panel shows mean response times to old and new high and low frequency words and nonwords in the lexical decision test in the multiple talkers study condition. The lower panel shows error rates.

The interaction between lexical status and repetition status obtained in the single talker condition was also obtained in the multiple talker condition in both the analysis of the response time results and the error results [$F_{rt}(2,66)=11.30$, $p<.01$; $F_{err}(2,66)=20.38$, $p<.01$]. The upper panel of Figure 3.4 shows the interaction in the response time results while the lower panel shows a similar interaction in the analysis of the error rates. Responses were faster to high frequency words than to nonwords for both old and new items. Latencies for old low frequency words were faster than for old nonwords. A significant repetition effect was also obtained for the low frequency words.

The results of the analysis of the error rates were similar to the response times. Old high and low frequency words were responded to more accurately than old nonwords. Error rates for new high frequency words were lower than for nonwords or low frequency words. Finally, error rates were significantly higher for new low frequency than for old low frequency words.

Insert Figure 3.4 about here.

Discussion

The primary purpose of this experiment was to assess repetition effects in lexical decision, an implicit test of memory. Repetition effects were examined both with respect to changes in voice and with respect to word frequency. Only the repetition effects related to word frequency will be discussed in this section. The discussion of the effects of voice on the accuracy and latency of lexical decision will be presented after the presentation of the recognition memory results in the next section. By reserving the discussion of the voice effects until after the presentation of the recognition memory results, the two sets of findings can be compared more directly.

Two issues related to repetition effects are important to highlight in this discussion. The first issue concerns the difference in the size of the repetition effect for high and low frequency words. A larger effect was obtained for low frequency words than for high frequency words. In both the single-talker and multiple-talker conditions, the word frequency effect, as measured by response times, was severely attenuated. However, a large frequency effect was obtained in the analysis of the error rates in both conditions. Error rates for the new low frequency words were extremely high, while error rates for the high frequency words and nonwords remained constant across conditions.

The response times and error rates to high and low frequency words can be easily accommodated within Balota and Chumbley's (1984) model for lexical decision. They argue that words and nonwords vary along a dimension of familiarity or meaningfulness, which generates two distributions. When subjects carry out a lexical decision, they set two criteria along the dimension of familiarity. If a stimulus exceeds the upper criterion and is, therefore, highly familiar, a fast "word" response can be made. Similarly, if a stimulus falls below the lower criterion, and is very unfamiliar, a fast "nonword" response can be made. If the stimulus falls somewhere between the two criteria, a careful analysis of the stimulus pattern has to be carried out and response times are lengthened.

The results from the present experiment suggest that subjects may have relaxed the upper criterion while simultaneously tightening the lower criterion. Loosening the top criterion would be particularly beneficial to old low frequency words. When these items were initially presented, subjects may have had to perform a careful analysis of the stimulus pattern in order to make a lexical decision. In contrast, the high frequency items may have exceeded the upper criterion on their initial presentation. However, when the upper criterion was relaxed,

the repeated low frequency words exceeded the criterion, thus allowing a fast decision to be made. This strategy can account for the attenuation of the word frequency effect as measured by the response times.

Tightening the lower criterion would help to account for the increase in error rates for the new low frequency words. When the lower criterion is raised, and the upper criterion is lowered, the region that requires careful analysis under the distribution for words is reduced accordingly. Thus, many of the low frequency words may fall below the lower criterion, and therefore receive errorful, fast nonword responses.

The second important issue addressed by the present results concerns repetition effects for nonwords. Recall that Scarborough et al. (1977) reported facilitation in lexical decision and naming for repeated nonwords. Forbach et al. (1974) found no repetition effect for nonwords whereas McKoon and Ratcliff (1979) found inhibition for repeated nonwords. In both the single-talker condition and the multiple-talker condition of the present investigation, no evidence of a repetition effect was obtained for the nonwords in either the response times results or the error rates.

One way to account for the absence of a repetition effect for nonwords is to consider again predictions derived from Balota and Chumbley's (1984) model of lexical decision. Stimulus patterns that did not exceed the lower criterion of familiarity are automatically rejected as nonwords. In the present experiment, all of the nonwords were pronounceable and this attribute may have pushed them above the lower criterion into the range requiring a more careful examination of the stimulus pattern. Thus, a more detailed analysis may have been conducted on the stimulus patterns at every presentation. Because both old and new items are assumed to fall between the two criteria, repetition effects would not be anticipated.

EXPERIMENT 1B: Recognition Memory

Method

During the test phase of Experiment 1A, subjects participated in a lexical decision task, which can be considered a test of implicit memory (Roediger et al., 1993). In Experiment 1B, subjects participated in the same study condition as subjects in Experiment 1A. At the time of test, however, subjects were given a surprise recognition memory test. The critical comparison between the two experiments deals with the effects of changing voices between the study and test conditions. According to the framework described by Blaxton (1989), changes in voice should not influence the repetition effect in the lexical decision task. However, if information about a speaker's voice is incidentally encoded during the study phase, old items produced by a familiar voice should be recognized faster and more accurately than words and nonwords produced by unfamiliar voices.

Subjects

Forty-two native speakers of English served as subjects in the "single talker" study condition. An additional 42 subjects participated in the "multiple talker" study condition. All listeners were enrolled in an introductory psychology class at Indiana University and received partial course credit for their participation. Subjects reported no speech or hearing problems at the time of testing.

Materials

The materials used during the study and test phase of the experiment were identical to those used in Experiment 1A.

Procedure

Study Condition

The study condition was identical to the procedure used in Experiment 1A.

Test Condition

The test condition for the present experiment was very similar to the test condition for Experiment 1A. However, in this experiment, subjects performed a surprise recognition memory task, rather than a lexical decision task. The counterbalancing of stimulus items across talkers and old versus new conditions was the same as in Experiment 1A. The presentation of stimuli and the timing of test trials and feedback was also identical to the previous experiment. Responses to "old" items were made with the left hand, while response to "new" items were made with the right hand.

Results

Separate analyses of variance were conducted on the data collected from subjects in the single-talker and the multiple-talker conditions. Analyses were performed on mean hit rates and response times and mean false alarm rates. Lexical status (high frequency words vs. low frequency words vs. nonwords) and talker similarity (exact repetition vs. familiar voice vs. similar voice vs. dissimilar voice) were treated as within-subjects variables in each analysis.

Single talker condition

The top panel of Figure 3.5 shows the percentage of hits for old items as a function of talker during recognition memory. A main effect for talker was obtained in the analysis [$F(3,123)=9.86$, $p<.01$]. Words and nonwords that were exact repetitions of old items and items that were new productions, but were repeated in the familiar voice from the study session, were recognized more accurately than stimuli produced by the dissimilar talker. Accuracy of recognition for the similar talker fell between the old talker and the dissimilar talker.

Insert Figure 3.5 about here.

The middle panel of Figure 3.5 shows mean response times for hits as a function of voice. Old items that were exact repetitions or were produced by the voice heard during the study session were recognized faster than words and nonwords produced by the dissimilar talker [$F(3,123)=16.88$, $p<.01$]. As in the analysis of the hits, response times to the similar talker fell between the old voice and the dissimilar voice.

The bottom panel of Figure 3.5 shows mean false alarm rates as a function of talker. False alarms for the old talker were collapsed over exact repetitions and new productions because all of the items in this condition were new to the subjects. Consequently, there is no distinction between an exact repetition and a new utterance by the same voice. False alarms varied significantly as a function of voice [$F(2,82)=3.58$, $p<.05$]. Although Tukey's tests failed to localize the differences among talkers, false alarms tended to be higher for the familiar voice than for the dissimilar talker.

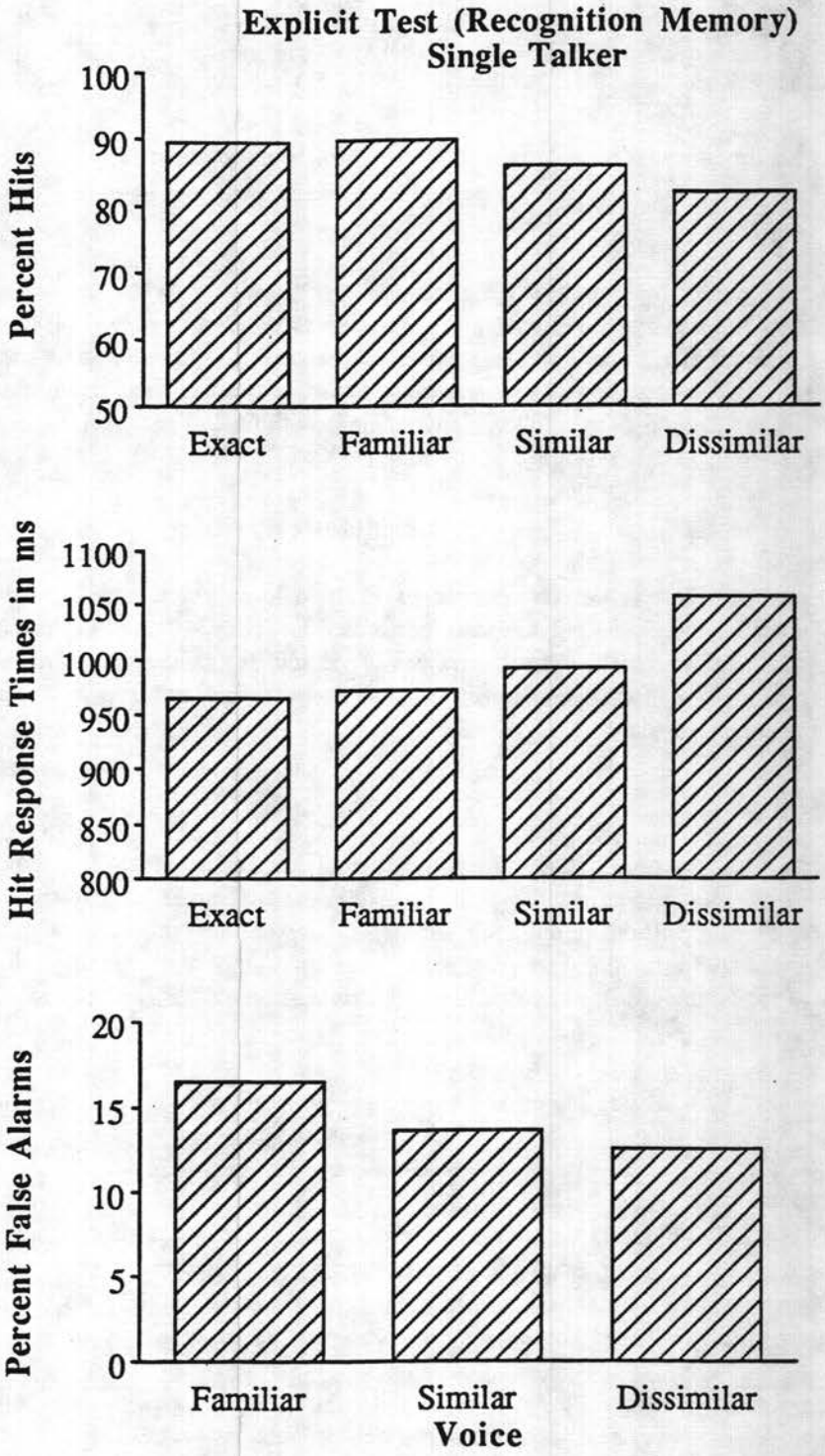

Figure 3.5. The upper panel shows mean hit rates as a function of voice in the recognition memory test of the single talker study condition. The middle panel shows mean response times for hits. The lower panel shows false alarm rates.

The top panel of Figure 3.6 shows the mean percentage of hits as a function of lexical status. The middle panel shows mean response times for hits and the lower panel shows false alarm rates. High and low frequency words were recognized faster and more accurately than nonwords [$F_{hit}(2,82)=65.64$, $p<.01$; $F_{rt\text{-}hit}(2,82)=48.18$, $p<.01$]. False alarm rates did not vary as a function of lexical status.

Insert Figure 3.6 about here.

Multiple-Talker Condition

The top panel of Figure 3.7 shows mean hit rates as a function of talker for subjects who participated in the multiple talker study condition. No main effect for voice was obtained [$F(3,123)<1$]. The middle panel shows mean response times for hits. A main effect for voice was observed in the analysis of the response times [$F(3,123)=5.65$, $p<.01$]. Although post-hoc tests failed to localize the significant differences, responses tended to be faster for exact repetitions and new instances of words and nonwords produced by the old voice (T2) than for the dissimilar voice. False alarm rates did not vary as a function of talker [$F(2,82)<1$].

Insert Figure 3.7 about here.

Figure 3.8 shows mean hit rates, mean response times for hits and mean false alarm rates as a function of lexical status. Listeners recognized high and low frequency words significantly more accurately and faster than nonwords [$F_{hit}(2,82)=29.77$, $p<.01$; $F_{hit\text{-}rt}(2,82)=77.11$, $p<.01$]. High frequency words and low frequency words did not differ significantly from each other in the hit rates or the response times for the hits. False alarm rates did not vary as a function of lexical status [$F(2,82)=2.58$, $p<.1$].

Insert Figure 3.8 about here.

Discussion

The purpose of the present experiment was to investigate the effects of changes in voice between the study condition and the test condition on recognition memory. Subjects studied the test items incidentally in an encoding task using a lexical decision response, and then were tested in a surprise recognition memory task. Listeners heard items produced by either a single talker or multiple talkers during the study phase. In both conditions, subjects heard words and nonwords during the test phase produced by an old talker, a similar voice and a dissimilar voice.

The critical results concern the effects of changes in voice between the study and test conditions. In both the single-talker study condition and the multiple-talker study condition, listeners were faster to respond to words and nonwords produced by the familiar talker than to a voice that was dissimilar to the familiar talker. The advantage for the old voice was also obtained in the analysis of the hits obtained from the single talker study condition. Taken together, these results suggest that changes in voice between the study and test conditions did produce effects on listeners' abilities to recognize spoken words and nonwords. Similar effects of changes in voice have been reported in a number of other studies using explicit procedures (see Craik & Kirsner, 1974; Palmeri et al., 1992).

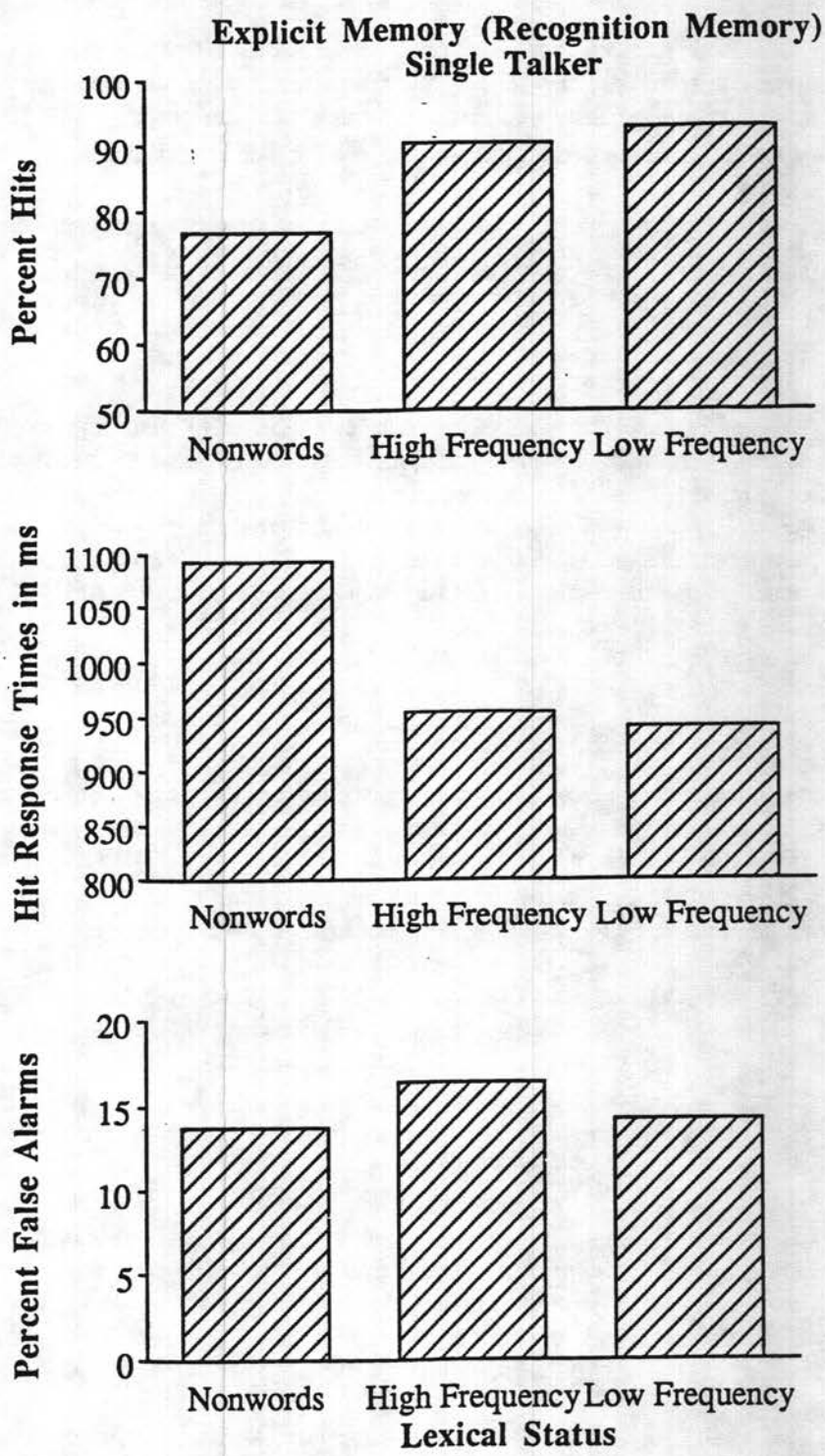

Figure 3.6. The upper panel shows mean hit rates as a function of lexical status in the recognition memory test of the single talker study condition. The middle panel shows mean response times for hits. The lower panel shows false alarm rates.

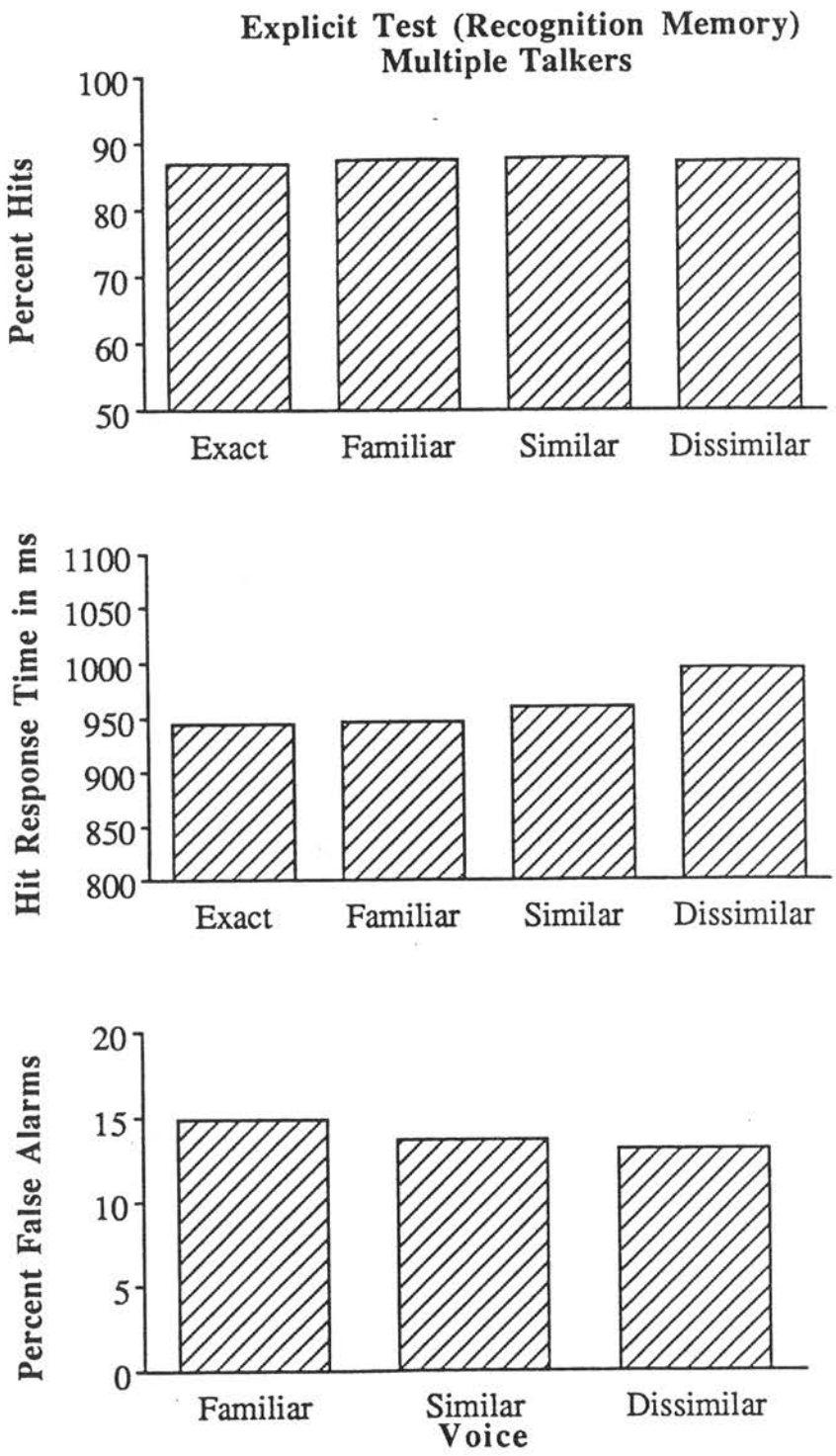

Figure 3.7. The upper panel shows mean hit rates as a function of voice in the recognition memory test of the multiple talker study condition. The middle panel shows mean response times for hits. The lower panel shows false alarm rates.

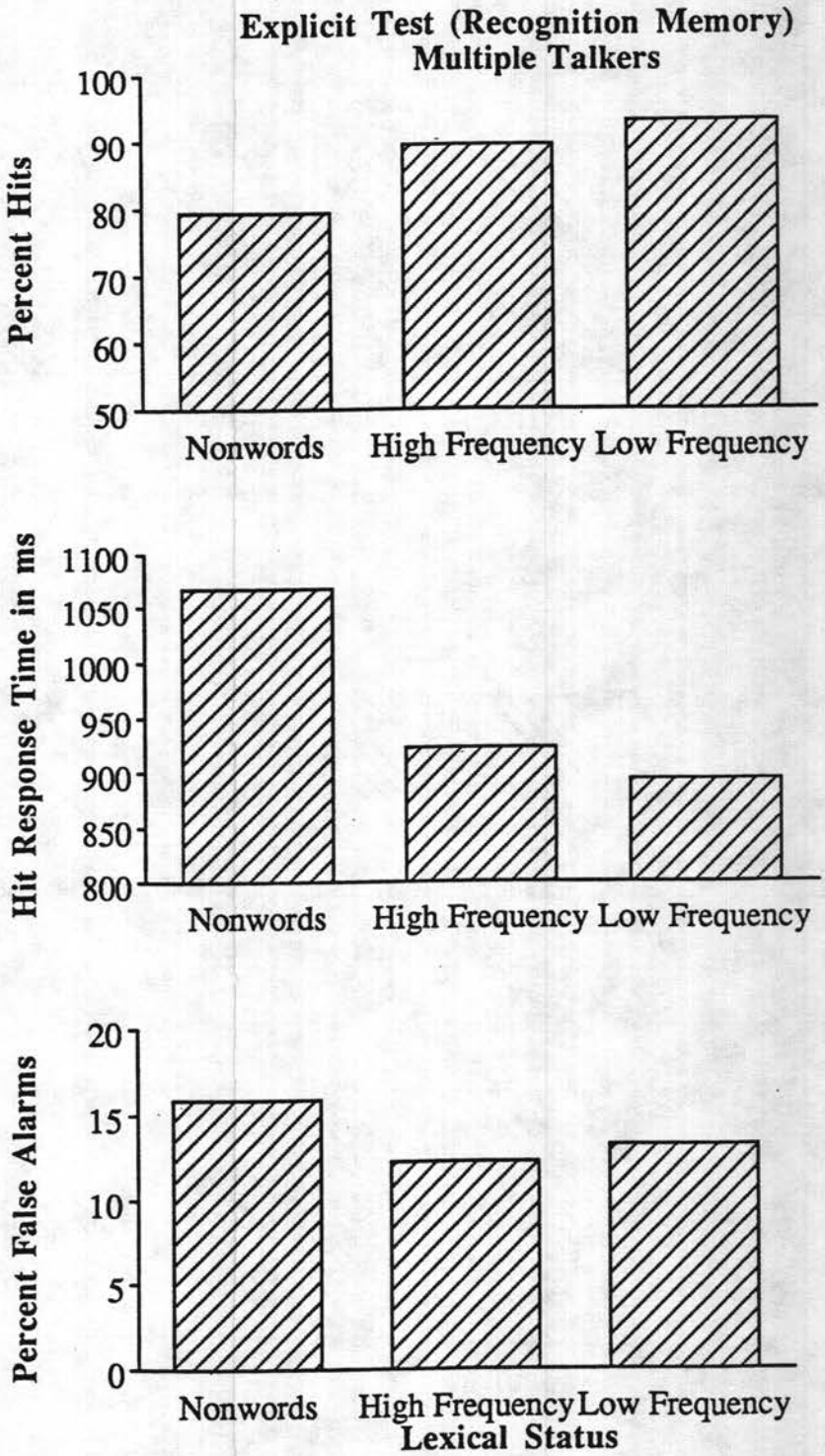

Figure 3.8. The upper panel shows mean hit rates as a function of lexical status in the recognition memory test of the multiple talker study condition. The middle panel shows mean response times for hits. The lower panel shows false alarm rates.

Some evidence was also obtained in the present experiment suggesting that recognition memory was affected by perceptual similarity among the talkers used in the test condition. In general, words and nonwords produced by the familiar voice from the study phase were recognized fastest and most accurately. Performance with the similar voice tended to be slightly, but not significantly, worse. Finally, speed and accuracy with a voice that was perceptually dissimilar to the old voice tended to be the slowest and least accurate. These results are similar to findings reported recently by Goldinger (1992), who found that recognition memory hit rates were moderately correlated with the perceptual distances among a set of male and female talkers.

The results from the recognition memory experiment contrast in several important ways with the findings obtained in Experiment 1A using the lexical decision task. Whereas changes in voice had a significant effect on recognition accuracy and latency, the same changes in voice did not influence the repetition effect in lexical decision response times or errors. Two observations support this conclusion. First, although there was a significant main effect for talker in the analyses of the lexical decision results, the ordering of voices was identical to the ordering of voices in the control condition. This suggests that any *a priori* differences among the voices were not overcome during the study phase. Second, in order to argue that changes in voices had an influence on the magnitude of the repetition effect, a significant interaction between talkers and repetition status would be expected. In this interaction, performance on old words and nonwords would have to be facilitated to a greater extent when they were produced by an old voice than when they were produced by the similar or dissimilar new voices.

The dissociation in the use of voice information across conditions can be explained by considering the types of information that subjects need to use in order to successfully complete the experimental tasks. The lexical decision task requires subjects to make a conceptual judgment about a spoken utterance. In this case, the listener is asked if the test item is in their mental lexicon or not. According to the framework laid out by Blaxton (1989), tests of conceptual knowledge tend to be very resistant to changes in surface forms. In contrast, the recognition memory test can be considered a data-driven task. Under the present experimental conditions, subjects use information about the voices as retrieval cues to facilitate recognition. Listeners in the present experiments were able to use this information regardless of whether they studied the items in the low variability, single-talker study condition or in the high variability, multiple-talker study condition.

Although the lexical decision task was insensitive to changes in voice between the study and test conditions, it should be noted that information about a speaker's voice was incidentally encoded into memory during the study phase, which also used a lexical decision task. Evidence for the encoding of talker-specific characteristics comes from considering the source of the effects of voice in recognition memory. If listeners had filtered out information about speakers' voices and discarded it as irrelevant noise, effects of voice would not have been found in the surprise recognition memory test. However, such effects were present in the analyses of the response times for hits in both study conditions and in the analysis of the hits for the single-talker study condition. Taken together, the findings from the recognition memory test and the lexical decision test indicate that voice information is incidentally encoded into memory, but that the information is only used when particular types of information are accessed from memory.

In addition to the effects of voice in the analyses of the results from the recognition memory task, recognition accuracy and response times also varied as a function of lexical status in the present experiment. Listeners recognized old high and low frequency words faster and more accurately than nonwords. Small, nonsignificant advantages were also obtained for low frequency words over high frequency words. Rao and Proctor (1984) obtained similar results using high frequency words, low frequency words and very low frequency words, which were functionally nonwords for the subjects. They suggested that the deficit for nonwords was due to ineffective processing of the nonwords during a lexical decision encoding task and they claimed that the processing deficit was caused by a lack of semantic information regarding the nonwords. As a consequence of this lack of information, subjects were forced to rely on an analysis of the surface characteristics

of the nonwords (Zechmeister et al., 1978). Subjects' reliance on surface features for nonwords, compared to the use of semantic information for words, may have led to the observed differences in recognition accuracy and response times found here.

EXPERIMENT 1C: Replication of Experiments 1A and 1B

Method

The purpose of Experiment 1C was to replicate the major findings of Experiments 1A and 1B with several modifications to the stimulus set and testing procedures. The first modification in the present experiment concerned the voices included in the study and test condition. In Experiments 1A and 1B talkers T1, T2, T5, T6, T7, T8, and T9 produced words and nonwords during the study phase of the experiment. Talker T2 served as the familiar voice during test. Talker T4 was included in the test as the similar voice and talker T3 was included as the dissimilar voice. Across both the implicit and explicit memory test conditions, responses were fastest to T2 and slowest to T3. Response times to T4 were intermediate between T2 and T3. One possible explanation for the consistency of this finding is that subjects were sensitive to the durations of the talkers' utterances and that these acoustic attributes were responsible for the differences observed among voices, rather than the encoding of talker-specific information. Measurement of the durations of the utterances produced by each of the talkers in the test condition indicates some support for this hypothesis. Mean durations for each talker in the test condition were ordered in the same pattern as the response times in each task: Talker T2 produced the shortest utterances (470 ms). The mean duration of T4's utterances (the similar talker) was 538 ms. Finally, the mean duration of T3's productions (the dissimilar talker) was 547 ms.

Because of the possibility that the results in Experiments 1A and 1B could be due to simple differences in stimulus durations, Experiment 1C was conducted using Talker T4 as the old talker and Talker T2 as the similar talker. As a result of this change in voices, the study session now consisted of T1, T4, T5, T6, T7, T8 and T9. The test session was conducted with talkers T4 (old talker), T2 (similar new talker), and T3 (dissimilar new talker). If the results of the present implicit and explicit memory tests follow the durations of the voices used in the test session, then the findings with regard to the use of voice information from Experiments 1A and 1B could be considered artifacts. However, if the pattern of response times replicates the pattern obtained in the previous experiments, then additional evidence will be provided for the hypothesis that voice information is used under some testing conditions, but not under others.

It should be noted that the hypothesis concerning the role of duration cannot be completely eliminated by the present modification because the dissimilar voice still displays the longest utterances. *A priori* differences among voices are particularly difficult to eliminate. As a consequence, researchers typically consider only "same" voice repetitions versus "different" voice repetitions. Counterbalancing the voices across conditions distributes the differences among voices evenly across conditions. In the present design, however, the role of similarity among the voices is of critical concern and such complete counterbalancing is not possible.

The second modification to Experiment 1C concerns the elimination of the single talker condition. The test conditions of the previous experiments and the present experiment were constructed by drawing on the perceptual similarity relations among talkers defined by the multidimensional scaling solution described in the previous chapter. This solution was defined over a range on nine voices. Subjects in the single talker study condition, however, heard only three of the nine voices over the course of the experiment. Thus, it is not clear what the nature of the similarity relationships are among the voices for the listeners in this condition. As a consequence, subjects in the experimental conditions of the present replication only participated in the multiple talker study condition. This condition more closely reflects the experimental conditions that were used to determine the perceptual similarities among the voices.

Finally, during the test session, old items produced by the old talker were exact repetitions of items produced during the study session. In the previous experiments, no differences were obtained between the exact repetitions and the new productions. The elimination of the new productions had two consequences for the composition of the test materials. First, each of the voices in the test conditions of the present experiment produced the same number of words and nonwords. In Experiments 1A and 1B the old talker produced the largest number of stimulus items during the test phase. Second, the number of test trials devoted to each experimental condition was increased from six to eight.

With the exception of the modifications described above, the study and test phases of the present experiment were identical to those used in Experiments 1A and 1B. Subjects participated in a common study phase in which they performed a lexical decision task. Following the completion of the study phase, subjects were either given a test of implicit memory (lexical decision) or a test of explicit memory (recognition memory). The main prediction for the experiment concerns the use of voice information across memory conditions. Listeners should not be sensitive to changes in voice during the lexical decision test because the task requires the use of conceptual information. However, listeners should exploit information about a speaker's voice during the recognition memory test in order to facilitate performance on the task.

Subjects

Subjects for the present experiment were 97 native speakers of English who were enrolled in an introductory psychology course at Indiana University. Listeners were given partial course credit for their participation. No subjects reported a history of speech or hearing problems at the time of testing. Thirty-five subjects participated in the control condition for the lexical decision experiment. These subjects only participated in the test phase of the experiment. Twenty-seven subjects participated in both study and test conditions of the implicit memory condition. Finally, 35 subjects participated in the study phase and the test of explicit memory.

Stimuli

The stimuli were identical to those used in Experiments 1A and 1B. As outlined above, talkers T1, T4, T5, T6, T7, T8, and T9 were assigned to the study phase of the experiment. Talker T4 served as the old talker during the test phase. Talker T2 was now the new, similar talker. Talker T3 remained the new, dissimilar talker.

Procedure

With the exception of the changes outlined above, the procedure for the present experiment was identical to the one used in Experiments 1A and 1B.

Results

Responses from each group of subjects were analyzed separately using analysis of variance. In analyses of the control condition and the implicit memory test, response times for correct responses and error rates were analyzed in separate ANOVAs. Talker (old vs. similar vs. dissimilar), repetition status (old vs. new), and lexical status (high frequency words vs. low frequency words vs. nonwords) were treated as within-subjects variables. In the analyses of the explicit memory test, hit rates, hit response times, and false alarms were analyzed in separate ANOVAs. Talker and lexical status were treated as within-subjects variables. All post-hoc tests were conducted using Tukey's HSD procedure.

Control Condition

Response times for correct responses and error rates varied significantly as a function of talker in the control condition [$F_{rt}(2,68)=18.51$, $p<.01$; $F_{err}(2,68)=5.93$, $p<.01$]. Subjects were slowest and least accurate when responding to the dissimilar talker (T3; rt: 1066 ms; 19% error). The differences between the old talker (T4) and the similar talker (T2) were small and did not approach significance (T4 rt: 1026 ms, 15% error; T2 rt: 1010 ms, 15% error).

Response times and error rates also varied as a function of lexical status [$F_{rt}(2,68)=128.76$, $p<.01$; $F_{err}(2,68)=67.00$, $p<.01$]. Responses to high and low frequency words were faster than to nonwords. Responses to high frequency words were faster than to low frequency words (high frequency words: 946 ms; low frequency words: 1019 ms; nonwords: 1137 ms). Error rates were significantly higher for nonwords and low frequency words than for high frequency words (high frequency words: 5% error; low frequency words: 22% error; nonwords: 22% error).

Implicit Memory Condition

Listeners were faster and more accurate when they responded to old words and nonwords than when they responded to new stimuli [old items: 936 ms, 9% error; new items: 1014 ms, 18% error; $F_{rt}(1,26)=73.12$, $p<.01$; $F_{err}(1,26)=102.38$, $p<.01$]. Response latencies and error rates also varied as function of lexical status [$F_{rt}(2,52)=107.49$, $p<.01$; $F_{err}(2,68)=44.64$, $p<.01$]. Responses to high and low frequency words were faster than to nonwords. In addition, latencies were marginally shorter to high frequency words than to low frequency words (high frequency words: 900 ms; low frequency words: 949 ms; nonwords: 1075 ms). Error rates for responses to high frequency words were significantly lower than to low frequency words or nonwords (high frequency words: 5% error; low frequency words: 20% error; nonwords: 16% error).

The top panel of Figure 3.9 shows mean response latencies as a function of talker. The light hatched bars show results from subjects who participated in the study and test conditions. The dark hatched bars show mean response times for subjects in the control condition. A main effect for voice was obtained in the analysis of the data collected for the subjects in the study and test condition [$F(2,52)=13.02$, $p<.01$]. Although post-hoc tests failed to localize the differences, responses to the old talker and the similar talker tended to be faster than responses to the dissimilar talker.

Insert Figure 3.9 about here.

The lower panel of Figure 3.9 shows error rates as a function of talker. The light hatched bars represent the subjects in the study and test condition. The dark hatched bars represent mean error rates for subjects in the control condition. Error rates in the study and test condition were lower when subjects responded to the old voice than when they responded to the dissimilar voice [$F(2,52)=9.98$, $p<.01$].

Significant interactions between lexical status and repetition status were obtained in the analyses of the response times and the error rates [$F_{rt}(2,52)=13.02$, $p<.01$; $F_{err}(2,52)=16.65$, $p<.01$]. The upper panel of Figure 3.10 shows the interaction in the mean response times. The lower panel shows a similar interaction in the error rates. Word frequency effects were severely attenuated for repeated items: Old high and low frequency words were responded to faster than old nonwords. However, response times to the two types of words did not differ significantly from each other. For new items, all pairwise comparisons among high and low frequency

words and nonwords were significant. Significant repetition effects were also obtained for high and low frequency words.

Insert Figure 3.10 about here.

Post-hoc tests on the mean error rates in each condition showed that old high frequency words were responded to more accurately than old low frequency words or nonwords. For new items, all pairwise comparisons among high frequency words, low frequency words and nonwords were significant. Error rates for old low frequency words were significantly lower than error rates for new low frequency words.

Explicit Memory Condition

The top panel of Figure 3.11 shows the mean percentage of hits as a function of talker during recognition memory. Listeners were more accurate when responding to words and nonwords produced by the old voice than when responding to stimuli produced by the dissimilar voice [$F(2,68)=7.98$, $p<.01$]. The middle panel shows response times for hits and the lower panel shows false alarms. Response times and false alarm rates did not vary as a function of talker.

Insert Figure 3.11 about here.

The three panels of Figure 3.12 show the mean percentage of hits, response times for hits and false alarms as a function of lexical status. High and low frequency words were recognized significantly faster than nonwords [$F_{hit}(2,68)=39.27$, $p<.01$; $F_{hit\text{-}rt}(2,68)=26.83$, $p<.01$]. Hit rates and response times did not differ significantly between high and low frequency words. False alarm rates were higher for nonwords than for high frequency words [$F(2,68)=5.47$, $p<.01$].

Insert Figure 3.12 about here.

Discussion

The purpose of the Experiment 1C was to insure that the findings from Experiments 1A and 1B concerning the role of changes in voice in lexical decision and recognition memory could not be accounted for by the specific choice of talkers. In the present experiments, the roles of talkers T2 and T4 were reversed in the test condition from their roles in Experiments 1A and 1B. T4, who was the similar talker in the previous experiments was now the old talker in the present experiment. T2, who was the old talker in Experiments 1A and 1B, served as the similar talker. The change in voices during the test condition was motivated by the observation that the mean duration of the talkers' utterances varied widely and that these variations in durations closely matched differences in response times to each of the voices. Switching the voices across experiments was one way to achieve partial counterbalancing across conditions.

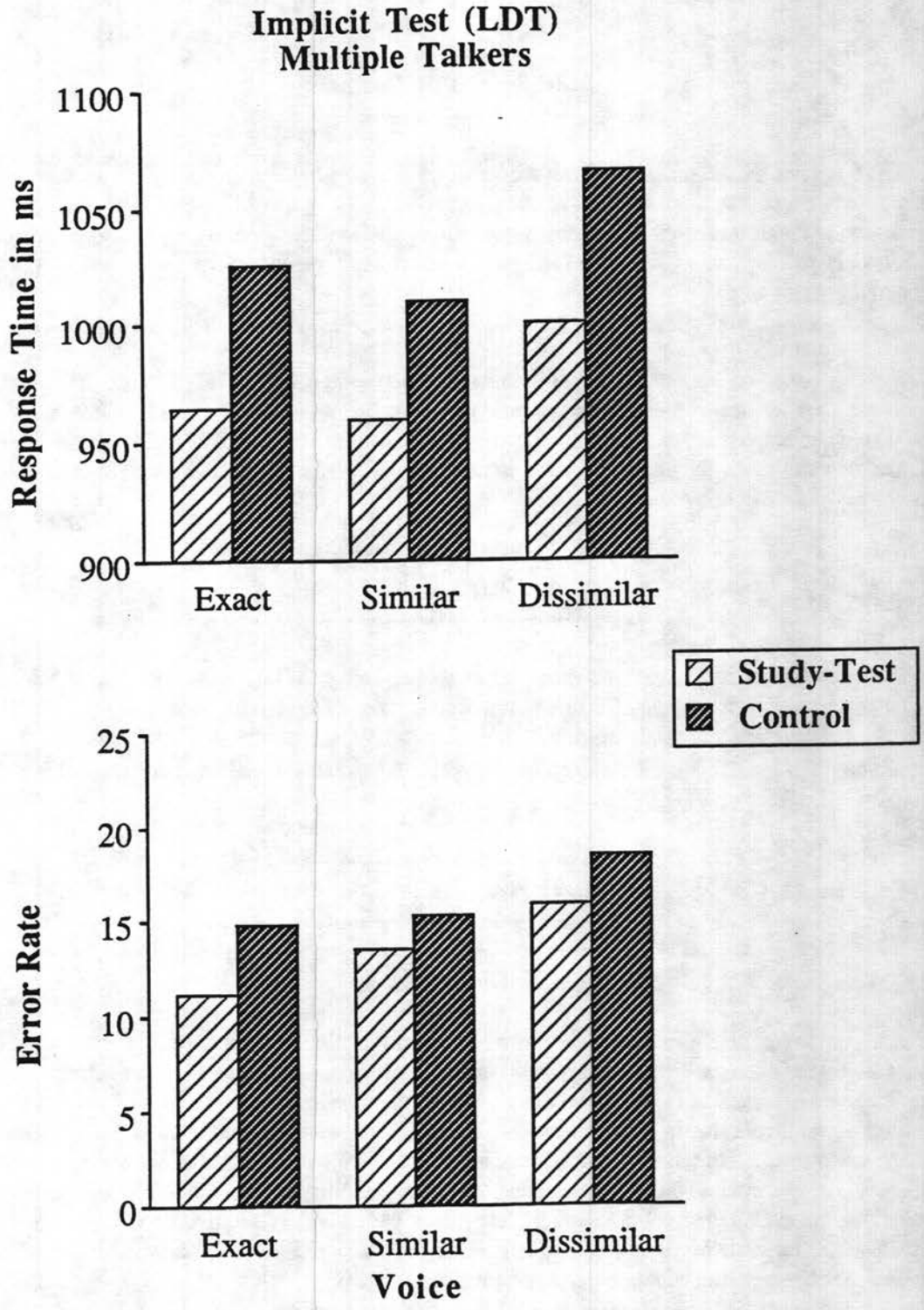

Figure 3.9. The upper panel shows mean response times to each talker in the lexical decision test. The lower panel shows error rates.

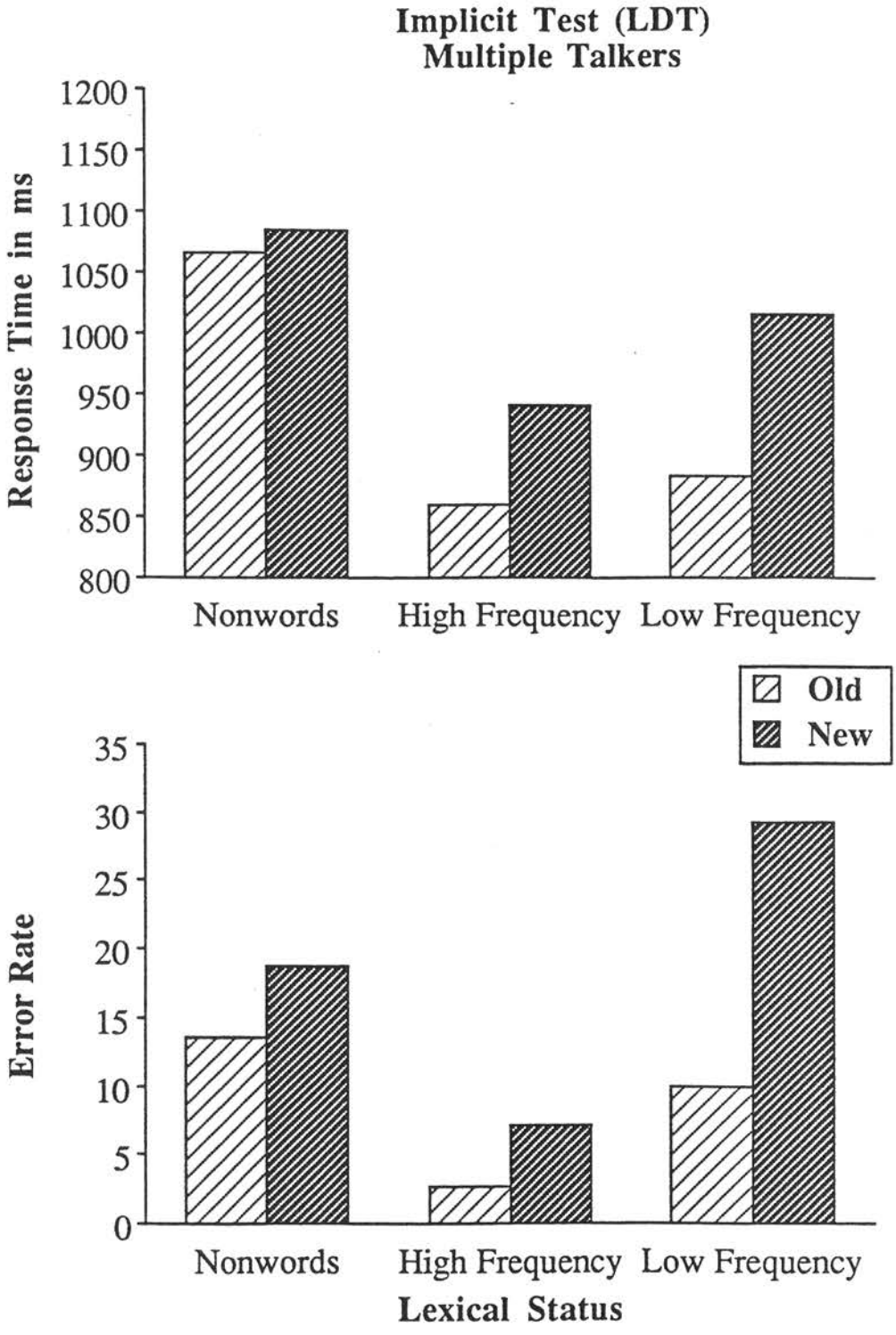

Figure 3.10. The upper panel shows mean response times to old and new high and low frequency words and nonwords in the lexical decision test. The lower panel shows error rates.

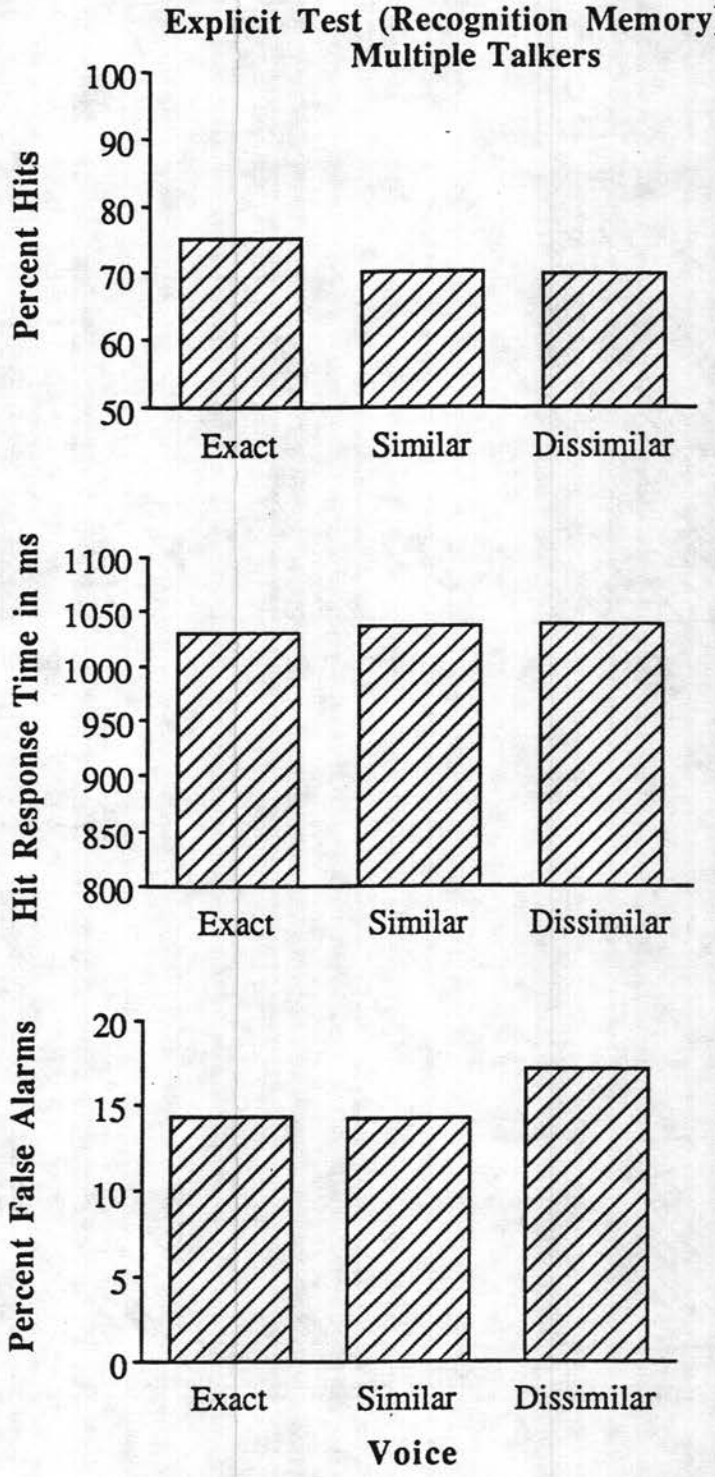

Figure 3.11. The upper panel shows mean hit rates as a function of voice in the recognition memory test. The middle panel shows mean response times for hits. The lower panel shows false alarm rates.

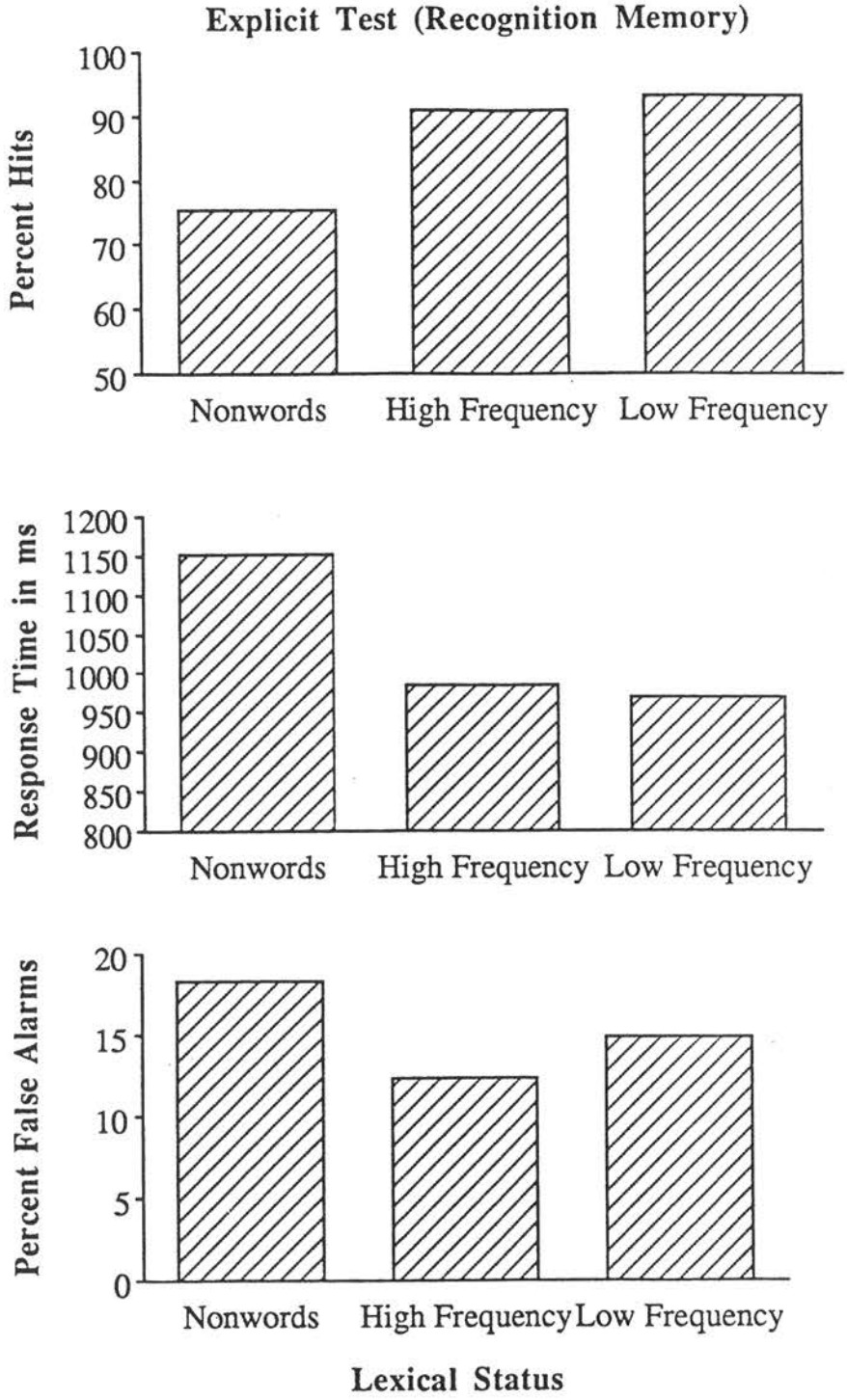

Figure 3.12. The upper panel shows mean hit rates as a function of lexical status in the recognition memory test. The middle panel shows mean response times for hits. The lower panel shows false alarm rates.

The results of the present experiment replicated the major findings from Experiments 1A and 1B. First, the importance of changes in voice during the lexical decision test of implicit memory was minimal. Although a significant main effect for voice was obtained, the ordering of voices was very similar to the pattern obtained in the control condition. This suggests that the familiarity gained with the old voice during the study phase of the experiment was insufficient to overcome any *a priori* differences among the voices. The absence of a significant interaction between the voice and repetition status variables also suggests that changes in voices did not have an impact on repetition effects during lexical decision. If changes in voices affected repetition effects in lexical decision, larger effects would have been anticipated when listeners responded to words and nonwords produced by the old talker. This did not happen. Furthermore, if repetition effects in lexical decision were sensitive to the similarity among talkers, a larger repetition effect would have been expected for the similar talker than for the dissimilar talker. This also did not occur.

Second, although changes in voice between the study and test conditions did not have an effect during a lexical decision test, changes in voice did influence explicit recognition memory. In the present experiment, listeners were more accurate at recognizing old and new nonwords that were repeated by one of the voices heard during the study session. This finding replicates the result obtained in the single-talker condition from Experiment 1B. However, it should be noted that an effect of voice was only obtained in the mean response times for hits in the multiple-talker study condition from Experiment 1B. The reason for the different effects across experiments is not clear at this time.

Taken together, the results of Experiments 1A, 1B, and 1C demonstrate that the lexical decision task is insensitive to changes in voice under conditions that produce repetition priming. In contrast, performance on recognition memory tests is improved by preserving the voice of the talker during the study and test conditions. This dissociation in the use of voice information suggests that the two tasks require subjects to make use of quite different sources of information contained in the acoustic signal. The lexical decision task requires subjects to draw on conceptual information, which is relatively insensitive to the original surface form of the word. In contrast, the recognition memory task relies on information about the talker's voice to guide processing and facilitate recognition.

In the next chapter, the retention of voice-specific information over time is examined. The results of Experiment 1B and 1C suggest that the use of talker-specific information facilitates recognition memory in a test given immediately after the study session. In the following chapter, the test is delayed for a period of twenty-four hours. If voice-specific information is preserved in long-term memory over time, then items produced by an old talker should be recognized more accurately than stimuli produced by a new talker. Furthermore, if voice information is retained in memory, then the dissociation observed in the present experiments between implicit and explicit tests of memory should also be obtained when the test is delayed.

CHAPTER IV: Retention of Voice Information Over Time

Introduction

In the preceding experiments, consistent dissociations in the use of voice information were obtained across implicit and explicit memory conditions. Subjects were not sensitive to changes in voice between the study and test sessions on the implicit test of memory (lexical decision). In the explicit test (recognition memory), however, subjects' accuracy and latency of responses to repeated words and nonwords were affected by changes in voice from the study session to the test session. Performance was better when stimuli were repeated in the same voice from the study session. In each of these experiments, the test phase was conducted immediately after the conclusion of the study phase. The main issue to be addressed in the present experiment concerns whether information related to a talker's voice is preserved over time and can be used to facilitate performance on delayed tests of implicit and explicit memory.

Several recent investigations of explicit and implicit memory have examined the retention of speaker-related characteristics over time. Using a continuous recognition memory task, Palmeri et al. (1992) found that listeners preserved information about the surface forms of spoken words for a period of at least five minutes. In their investigation, listeners recognized old words more accurately when they were repeated in the same voice than when the words were repeated in a different voice. The effect was the same whether the change in voices occurred within a gender or across genders. Goldinger (1992) found that the recognition memory accuracy varied as a function of similarity among voices over a span of 24 hours. After a week had elapsed between the initial study session and the test session, however, accuracy in recognition memory was no longer correlated with perceptual similarity.

While Goldinger (1992) found that the advantage for old voices decreased over time in a test of explicit memory, he also found that the correlation between perceptual similarity among voices and the magnitude of repetition effects did not decrease over a period of one week in perceptual identification, a test of implicit memory. Based on the difference in results obtained between the implicit and explicit memory conditions, Goldinger argued that conscious access to the surface features of spoken words stored in long-term memory is limited over time. However, despite this limited access, old instances could still influence the efficiency and accuracy of processing spoken words.

In the present investigation, the dissociation in the use of talker-related characteristics is examined again in both implicit and explicit memory conditions. Subjects participated in the same lexical decision encoding task as listeners from the previous experiments. Instead of taking the test phase immediately after the completion of the study session, however, listeners in the present experiment were tested after a delay of twenty-four hours. During the test phase, subjects were given either the lexical decision test of implicit memory or the recognition test of explicit memory. As in the previous experiments, changes in voice between the study and test session were not expected to influence performance on the lexical decision task. However, given Goldinger's (1992) findings, changes in the voice of the talker between the study and test session were predicted to affect recognition, even twenty four hours after the original incidental encoding phase. Taken together, these results would suggest that the surface forms of spoken words and nonwords are encoded into memory at the initial presentation and are subsequently retained over time. Furthermore, this information can be accessed and applied to the recognition of new test materials. However, this information is only engaged by tasks that rely on data-driven processing.

Method

Subjects

Thirty-two native speakers of English served as subjects in the present experiment. Fourteen listeners were randomly assigned to the explicit memory condition and eighteen listeners were assigned to the implicit memory condition. All subjects were enrolled in an introductory psychology course at Indiana University and were given partial course credit for their participation. Listeners reported no history of speech or hearing disorders at the time of testing.

Materials

The materials used in the present investigation were identical to those used in Experiment 1C. Talkers T1, T4, T5, T6, T7, T8, and T9 produced tokens during the study session. Talker T4 was the old talker during the test session. Talker T2 was the similar new talker, based on the results of the multidimensional scaling experiment. Talker T3 was the dissimilar new talker.

Procedure

The procedure for the study phase was identical to the one used in Experiment 1C. The test phase of the experiment was also identical to the test phase of Experiment 1C. However, in the present experiment, the explicit or implicit test of memory was delayed by twenty four hours.

Results

Responses from each group of subjects were analyzed separately using analysis of variance. In the analyses of the results from the implicit memory condition, mean response times for correct responses and error rates were analyzed in separate ANOVAs. Talker (old vs. similar vs. dissimilar), lexical status (high frequency words vs. low frequency words vs. nonwords) and repetition status were treated as within-subjects variables in each analysis. In the analyses of the explicit memory test, mean hit rates, mean response times for hits, and mean false alarm rates were analyzed in separate ANOVAs. Talker and lexical status were within-subjects variables in each analysis.

Implicit Memory Condition

The upper panel of Figure 4.1 shows mean response times as a function of lexical status and repetition status. The lower panel shows the corresponding error rates. Subjects were faster and more accurate when responding to old words and nonwords [$F_{rt}(1,17)=21.84$, $p<.01$, $F_{err}(1,17)=12.37$, $p<.01$]. Response latencies and error rates also varied as a function of lexical status [$F_{rt}(2,34)=77.17$, $p<.01$, $F_{err}(2,34)=36.05$, $p<.01$]. Responses were significantly faster to high and low frequency words than to nonwords. Mean response latencies were marginally shorter for high frequency words than for low frequency words. High frequency words were also responded to significantly more accurately than low frequency words or nonwords ($p<.05$).

Insert Figure 4.1 about here.

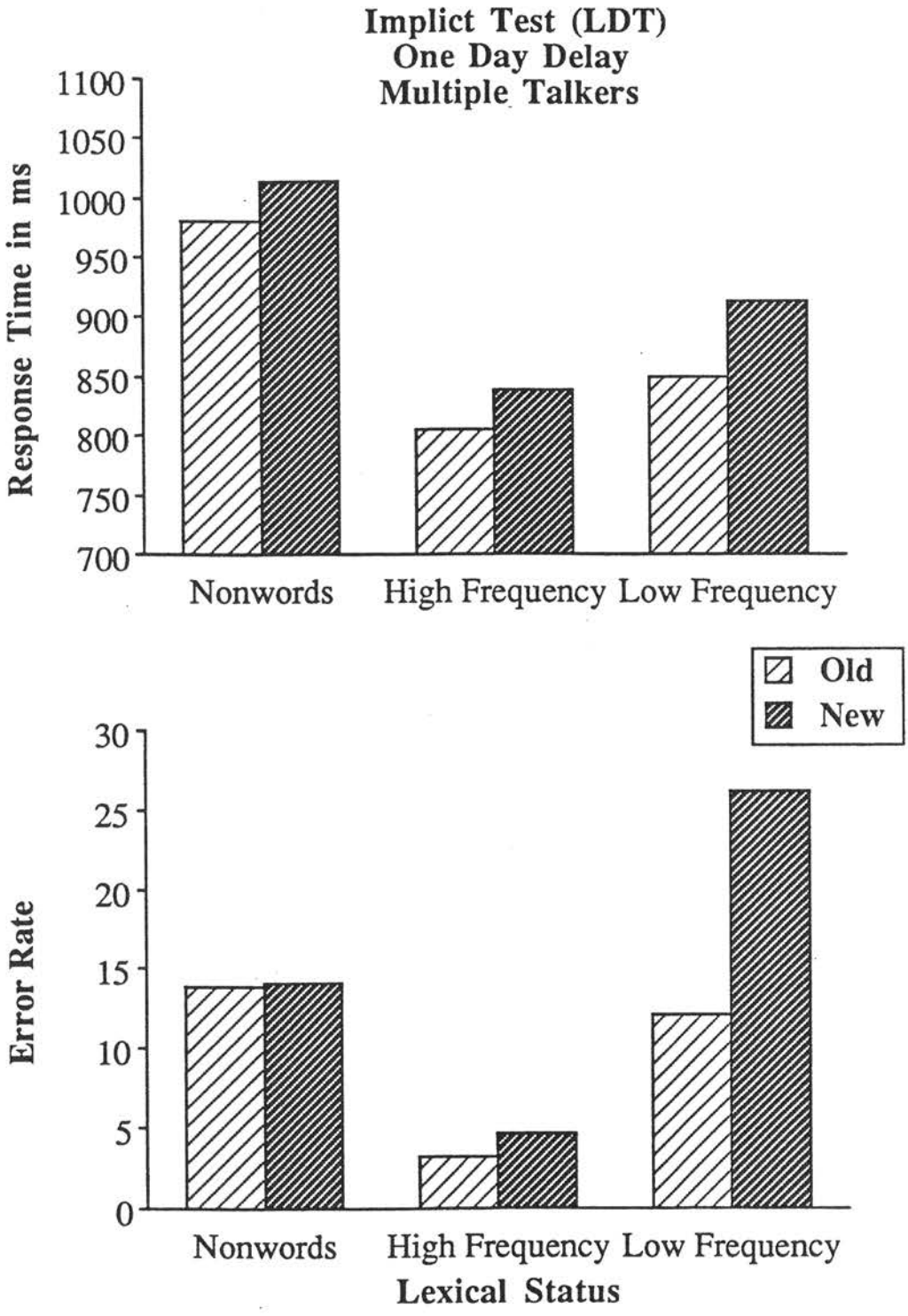

Figure 4.1. The upper panel shows mean response times to old and new high and low frequency words and nonwords in the delayed lexical decision test. The lower panel shows error rates.

In addition to the main effects for repetition status and lexical status, the interaction between the two variables was also significant in the analysis of the error rates [$F(2,34)=11.41$, $p<.01$]. For repeated items, error rates were significantly lower for high frequency words than for low frequency words or nonwords. For new items, all pairwise comparisons were significant. Thus, high frequency words were responded to more accurately than nonwords and low frequency words. Similarly, responses to nonwords were more accurate than responses to words.

The top panel of Figure 4.2 shows mean response latencies as a function of talker. The lower panel shows the corresponding error rates. Listeners were significantly faster when responding to the old talker and the similar talker than when responding to the dissimilar talker [$F(2,34)=12.21$, $p<.01$]. Listeners were also more accurate when responding to the old talker than when responding to the dissimilar talker [$F(2,34)=3.74$, $p<.01$].

Insert Figure 4.2 about here.

Explicit Memory Test

The top panel of Figure 4.3 shows mean percentage of hits as a function of voice. The middle panel shows the corresponding response times for hits and the lower panel shows false alarm rates. Hit rates varied significantly as a function of talker [$F(2,26)=10.16$, $p<.01$]. Although Tukey's tests failed to localize the differences among talkers, words and nonwords produced by the old talker were recognized the most accurately while tokens produced by the dissimilar talker were recognized the least accurately. Performance with the similar talker fell midway between the old talker and the dissimilar talker. Main effects for voice did not approach significance in the analyses of the mean response times for hits or the false alarms.

Insert Figure 4.3 about here.

Figure 4.4 shows mean hit percentages, response times for hits and false alarm rates as a function of lexical status. Subjects were the most accurate and fastest when recognizing low frequency words [$F_{hit}(2,26)=7.19$, $p<.01$; $F_{hit\text{-}rt}(2,26)=9.13$, $p<.01$]. Hit rates and response times were almost equivalent for high frequency words and nonwords. False alarm rates did not vary significantly as a function of lexical status.

Insert Figure 4.4 about here.

Discussion

The purpose of the present experiment was to investigate the retention of talker-specific information in long-term memory over time. Memory for characteristics of speakers' voices was assessed using both an implicit memory task (lexical decision) and an explicit memory task (recognition memory). As expected, the lexical decision task was insensitive to changes in voice between the study and test conditions. Although response times did vary as a function of talker during the test phase, the order of mean response times did not differ from the *a priori* ordering of voices determined by a control group of subjects in Experiment 1C.

In contrast to the results obtained with the lexical decision task, findings from the delayed recognition memory test indicated that subjects stored and had access to detailed information about a speaker's voice over a

period of at least twenty-four hours. This result extends findings reported by Goldinger (1992). Some evidence was also obtained in the present experiment that suggests recognition memory accuracy was affected by the similarity among the voices: Listeners tended to be most accurate when responding to the old voice and least accurate when responding to the new dissimilar voice. Performance with the dissimilar voice was intermediate between the old voice and the dissimilar voice.

In addition to the findings concerning the use of voice information during the lexical decision test and the recognition memory test, repetition effects during the lexical decision test should also be noted. Responses were faster and more accurate during the test session when subjects responded to repeated words and nonwords. Furthermore, an interaction between repetition status and lexical status was obtained in the analysis of the error rates: Subjects committed a remarkably high proportion of errors when responding to new low frequency words. Within the context of the model for lexical decision described by Balota and Chumbley (1984), these findings suggest that the criteria subjects set during the study phase for making fast decisions about words and nonwords were reinstated during the test phase. As a consequence, many of the new low frequency words fell below the lower criterion for committing to a careful analysis of the input pattern. Because these items fell below the lower criterion, incorrect "nonword" responses were assigned to the patterns.

Taken together, the results of the experiments reported in Chapters III and IV have been remarkably consistent. First, changes in voice had little effect on the repetition effect obtained in lexical decision. This finding contrasts with results obtained in other auditory word recognition tasks such as perceptual identification or auditory stem completion (see Goldinger, 1992; Schacter & Church, 1992). Second, changes in voice between the study and test sessions did affect recognition memory. Across three experiments (1B, 1C, and the present experiment), recognition of old words and nonwords was faster and more accurate when the tokens were produced by an old voice encountered during the study session than when tokens were produced by new voices.

The role of similarity among voices in recognition was ambiguous. Although stimuli produced by the voice that was similar to the old talker tended to be recognized faster and more accurately than tokens produced by the talker who was dissimilar to the old talker, statistical analysis did not support this trend in any of the experiments. Third, talker-specific information was incidentally encoded during lexical decision and subjects were able to access this information even after a delay of twenty-four hours. Fourth, one source for the dissociation in the use of voice information across implicit and explicit memory tasks is that the two tests require subjects to access different types of information. The lexical decision task draws heavily upon conceptual or semantic knowledge. In contrast, the recognition memory task allows subjects to exploit the surface features, or data-relevant aspects, of spoken words as cues to facilitate performance.

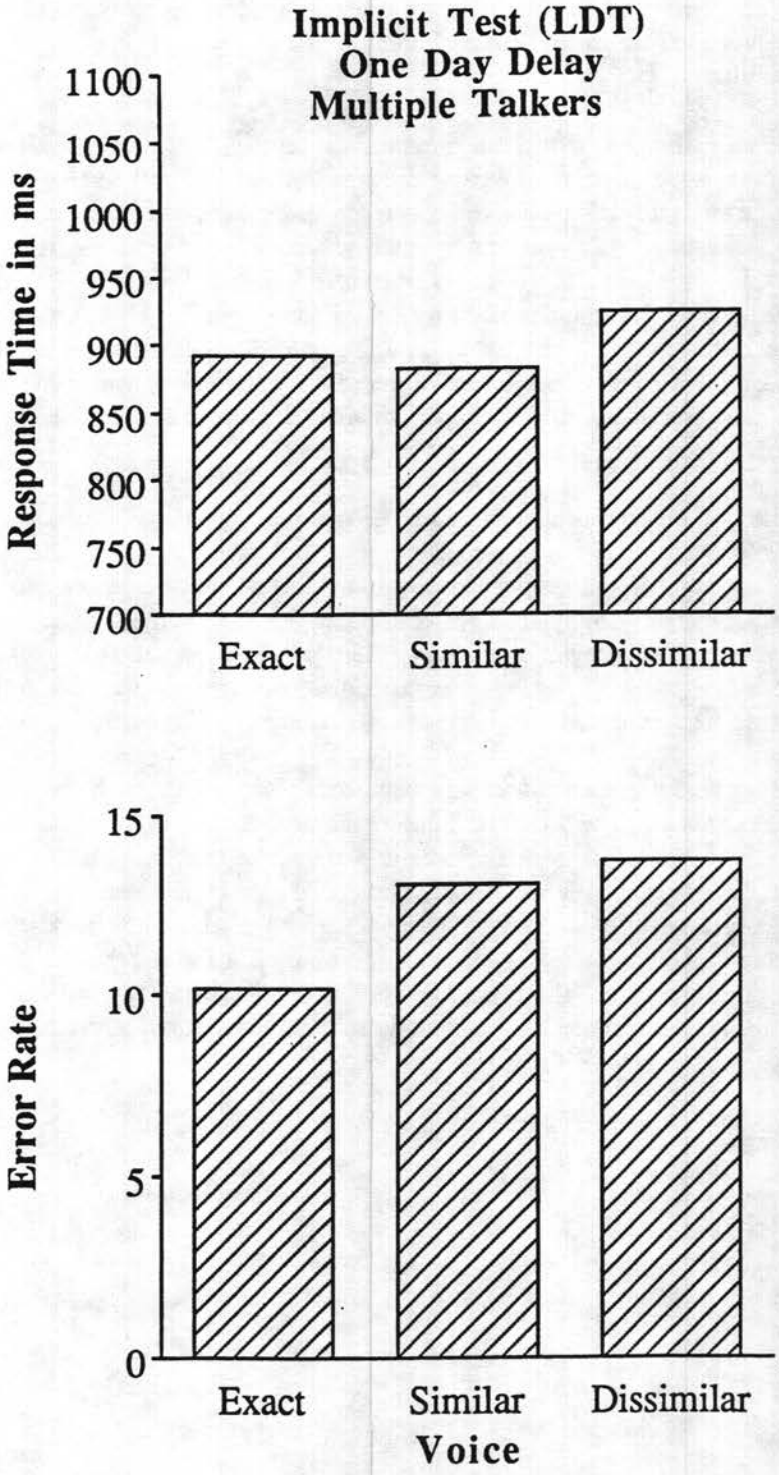

Figure 4.2. The upper panel shows mean response times to each talker in the delayed lexical decision test. The lower panel shows error rates.

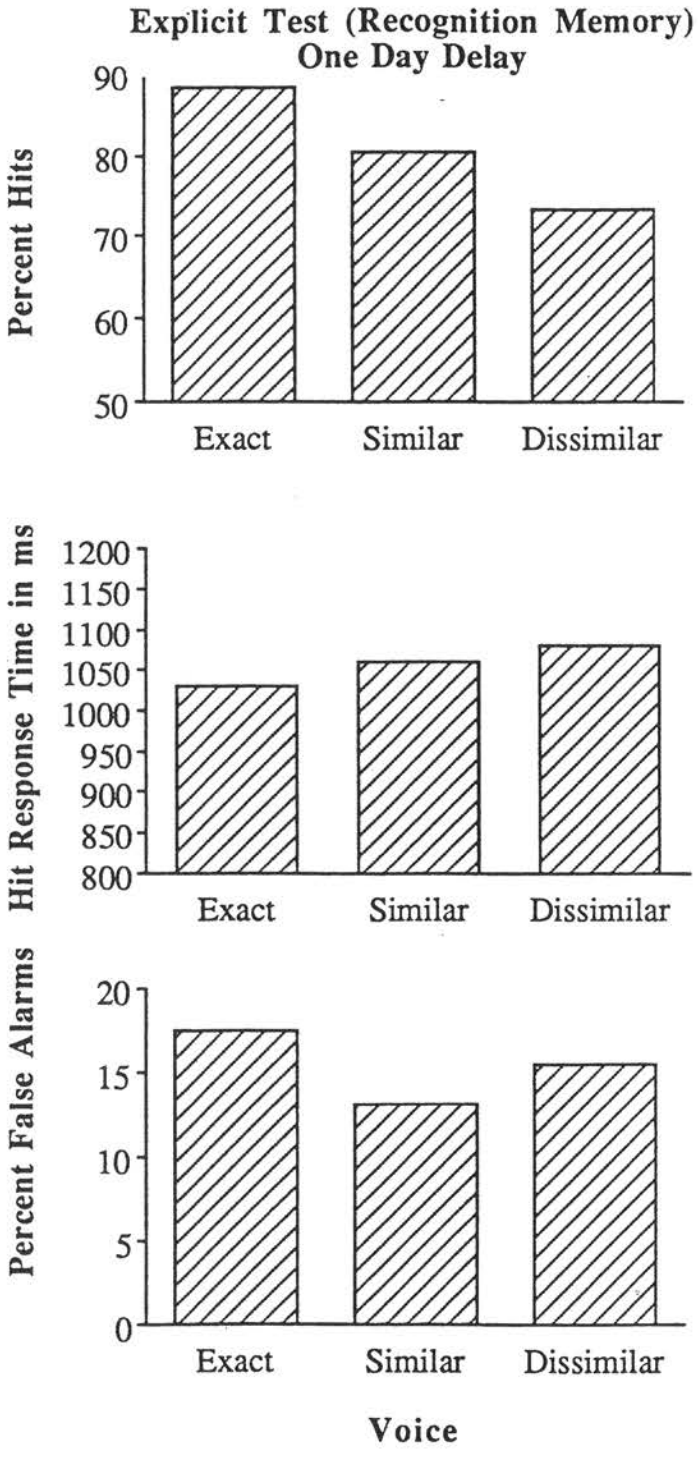

Figure 4.3. The upper panel shows mean hit rates as a function of voice in the delayed recognition memory test. The middle panel shows mean response times for hits. The lower panel shows false alarm rates.

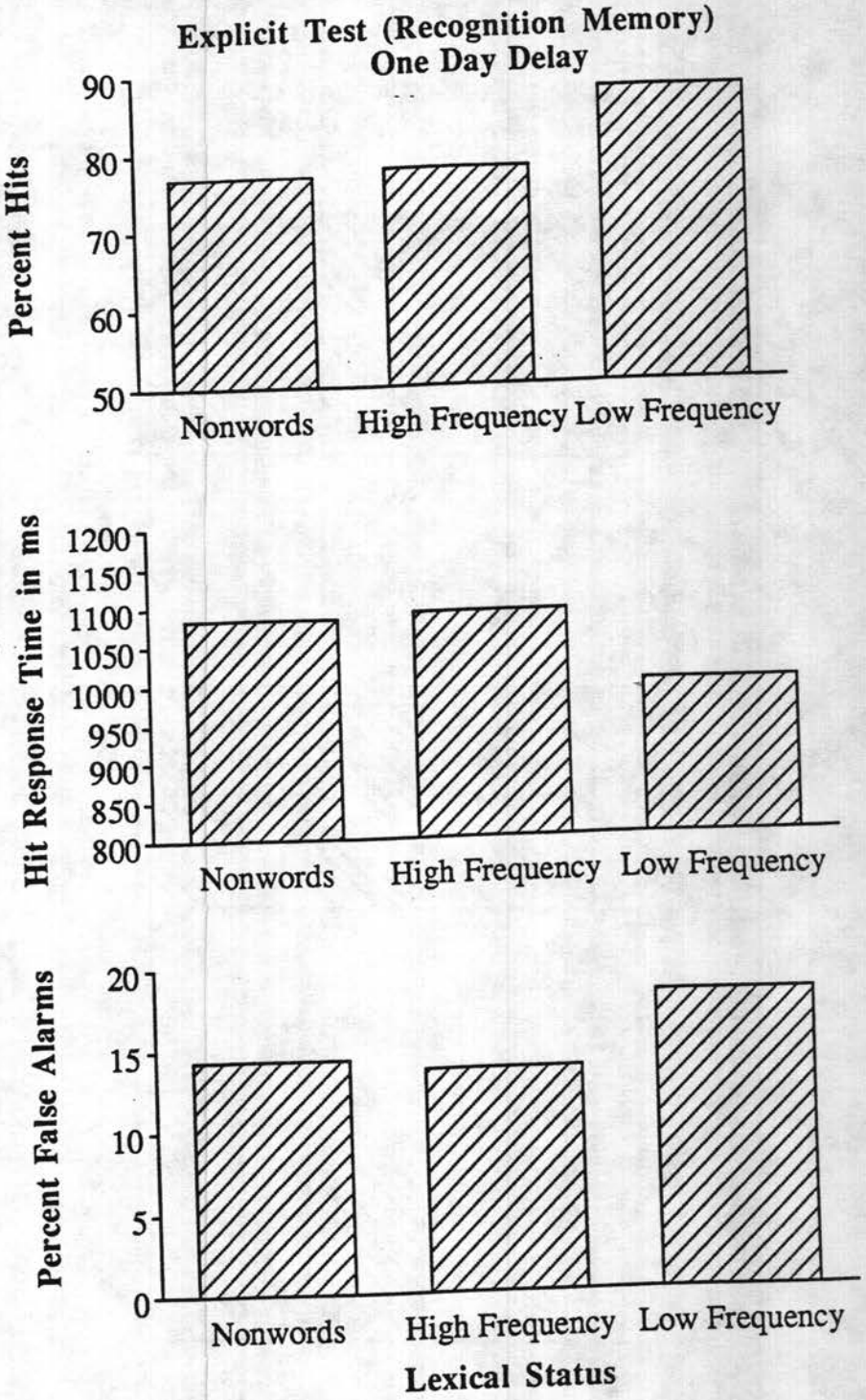

Figure 4.4. The upper panel shows mean hit rates as a function of lexical status in the delayed recognition memory test. The middle panel shows mean response times for hits. The lower panel shows false alarm rates.

CHAPTER V: Data-driven Tests of Implicit Memory

Introduction

In the two preceding chapters, dissociations in the use of talker-specific information were found across test of implicit and explicit memory. Changes in the voice of the talker between the study session and the test session did not influence repetition effects in lexical decision, a test of implicit memory. In contrast, changes in voice did have an influence on processing during the recognition memory task, an explicit test of memory: Words and nonwords produced by an old voice were recognized faster and more accurately than tokens produced by new voices. The difference in the effects obtained between the two conditions was explained in terms of the use of different types of information: The lexical decision task was assumed to tap conceptual information, whereas the recognition memory task was assumed to rely on the use of data-driven or surface information (Blaxton, 1989).

The dissociation is interesting because it indicates that while details related to a talker's voice are mandatorally processed and encoded into memory (Mullennix & Pisoni, 1990), these same characteristics are not necessarily used to facilitate access to the lexicon. However, it should be noted that the experimental design used in the previous chapters does not allow for a general statement to be made about the conditions under which characteristics related to a talker's voice will influence spoken language processing. This problem arises due to a confound between the type of memory task subjects were given and the type of information that was required to complete the task. As noted above, lexical decision, a test of implicit memory, relies mainly on conceptual information. In contrast, the recognition memory task, an explicit test of memory, has a strong data-driven component in which information about a speaker's voice is used to facilitate processing. Thus, the tasks differ along the dimensions of whether subjects must implicitly or explicitly access memory and whether they use conceptual or data-driven information.

In order to make a more definitive statement about the influence of talker-specific information on spoken language processing, these two factors need to be unconfounded. The results reported in this chapter were designed to partially unconfound these issues using measures of implicit memory. The purpose of the experiments reported here is to examine the effects of changes in voice between the study and test session using two implicit memory tasks that rely on careful attention to the data-driven components or surface forms of spoken words. In Experiments 3A and 3B, hypotheses regarding the encoding and use of voice-related information were tested using a gating task. In Experiment 3C, implicit memory for a talker's voice was assessed in an auditory stem completion task. If implicit memory conditions can be found that are sensitive to changes in voices between the study phase and the test phase, this would suggest that both implicit and explicit memory are sensitive to the surface forms of spoken words. However, reinstantiating the surface forms of repeated items may not be critical to obtaining repetition effects in all tests of implicit memory.

EXPERIMENT 3A: Gating I

Method

The gating task is a modified version of the standard perceptual identification task. During an experimental trial in the gating procedure, successively larger portions of spoken words are presented to listeners for identification until the entire stimulus is presented (Grosjean, 1980; Grosjean & Cotton, 1984). The advantage that the gating task has over the perceptual identification task is that gating does not require the addition of noise to the signal to reduce performance from ceiling levels. As noted by Schacter and Church (1992), the use of noise may mask critical voice-related features that influence the size of repetition effects. In contrast, gating preserves the surface forms of the stimuli, but makes them available to subjects gradually over the course of several presentations.

Grosjean (1980; Cotton & Grosjean, 1984) originally developed the task to study the distinction between the isolation and recognition points of spoken words. The isolation point for a word occurs when sufficient acoustic-phonetic information has been accumulated to eliminate all other possible words from consideration for recognition. In contrast, the recognition point refers to the amount of information a listener needs to identify a spoken word (Lively, Goldinger, & Pisoni, 1994; Marslen-Wilson, 1985). Isolation points and recognition points do not necessarily correspond to each other. For example, recognition points can be affected by factors such as surrounding context and whether the words are gated from their beginnings or from their ends (Salasoo & Pisoni, 1985).

In the present investigation, changes in recognition points were measured as a function of repetition in same or different voices. Experimental sessions for the present investigation were conducted in two phases. The study phase was identical to the study conditions of the preceding experiments: Subjects performed a lexical decision task on words and nonwords produced by seven different voices. During the test phase, subjects identified old and new words using the gating procedure. Nonwords were eliminated from the test procedure because no orthographic transcription exists for nonwords.

The critical predictions for the present experiment are similar to predictions that would be made for a perceptual identification experiment. However, instead of measuring the accuracy of identification, the critical data in the present experiment are the recognition points. We expected that the gating task, like perceptual identification, would also be sensitive to changes in voice between the study and test conditions: Recognition points for repeated words should occur earlier than for new words. Furthermore, old words produced in the same voice should be recognized with less signal duration than words produced by the similar new talker. Finally, words produced by the similar talker should be recognized at earlier presentations or gates than words produced by the dissimilar talker. An effect of voice in the gating task would strengthen the argument that information about a speaker's voice is mandatorally encoded and used in tasks that require careful analysis of the input signal.

Subjects

Thirty-five native speakers of English served as subjects in the gating experiment. Thirteen listeners were assigned to the control condition and were only given the gating portion of the experiment. Twenty-two listeners were assigned to the experimental condition and were given both the study phase and the test phase. Subjects reported no history of speech or hearing problems at the time of testing. All subjects were paid $5.00 for their participation.

Stimuli

The stimuli for the lexical decision task were identical to those used in the previous experiments. Talkers T1, T4, T5, T6, T7, T8, and T9 were assigned to the study phase. The stimuli for the gating test were the words from the previous experiments. Forty-eight high frequency words and 48 low frequency words were presented during each test session. Half of the words were old items from the study phase and half were new words. Talker T4 served as the old voice. Talker T2 was the similar new voice and talker T3 was the dissimilar new voice. Old and new words were counterbalanced across talkers so that each voice produced each item for approximately the same number of subjects.

Procedure

The procedure for the lexical decision task was identical to the method presented in the previous experiments. Subjects received 504 trials and were given a short break after 252 trials. Each study session lasted approximately 35 minutes.

All words presented during the gating test were presented using the forward gating mode: Thus, subjects heard words beginning from their initial portions and new information was digitally added to the end of the stimulus. Before the presentation of each new test word, a warning appeared on the CRT monitors in front of subjects alerting them that a new word was about to be presented. One second after the prompt appeared, subjects heard the initial 50 ms of the stimulus word. Subjects had ten seconds to type their guess as to the identity of the word into the monitor. Listeners were instructed to press "return" if they had no guess. After all of the subjects had responded, the next 75 ms of the word was added to the signal and subjects responded again. Successive presentations with the addition of 75 ms continued until the entire word was presented. Stimulus presentation, gating duration control and response collection were controlled by a PDP 11/34 computer. Each experimental session consisted of 96 words and lasted between 45 minutes and one hour.

Results

The main dependent variable for the gating portion of the experiment was the recognition point of a word, that is the stimulus duration at which subjects identified the stimulus word correctly for the first time. The value of this recognition point is reported in milliseconds. Separate analyses were conducted on the mean gate durations collected from subjects in the control condition and subjects in the experimental condition. Words frequency and talker were treated as within-subjects variables in the ANOVA conducted on data from subjects in the control condition. Repetition status, word frequency and talker were also treated as within-subjects variables in the ANOVA conducted on data from subjects in the study and test condition.

Control Condition

No main effect for voice or word frequency was obtained in the ANOVA [both $F<1$] and the interaction between the two variables did not approach significance either. Words produced by the old talker were correctly identified for the first time with an average signal duration of 283 ms. Stimuli from the similar new talker were identified with an average signal duration of 290 ms. Words produced by the dissimilar new talker were correctly identified with a stimulus duration of 298 ms.

Study-Test Condition

Only the main effect for repetition status was significant in the analysis conducted on the mean recognition points for subjects from the experimental condition [$F(1,21)=20.70$, $p<.01$]. Old words were identified with an average signal duration of 264 ms. New words were identified with an average stimulus duration of 295 ms. None of the remaining effects approached significance.

Discussion

The purpose of the present experiment was to examine the effects of changes in voice on repetition priming using the gating task. Listeners studied words and nonwords in a lexical decision task and then were tested with old and new words in the gating paradigm. A repetition effect was obtained: Subjects required shorter signal durations to identify old words than to identify new words. Old words were identified with a mean gate duration 264 ms while new words were identified with an average gate duration of 295 ms. Marslen-Wilson (1987) has claimed that the recognition point of most normally articulated words occurs within 200-250 ms of

the word's onset. The present results suggest that recognition points for high and low frequency words can be manipulated through repeated presentations. Salasoo and Pisoni (1985) found that recognition points are also sensitive to the meaningfulness of the surrounding context and whether stimuli are gated from their beginnings or their ends.

Although repetition effects were obtained in this experiment, the absence of voice effects in the gating task was somewhat surprising. Goldinger (1992) found that the auditory perceptual identification task was sensitive to changes in voice between study and test sessions. More recently, Schacter and Church (1992) reported larger repetition priming effects in auditory stem completion when words were repeated in the same voice. These results suggest under certain conditions, implicit memory tasks are sensitive to the surface features of spoken words.

The transfer appropriate processing framework provides one possible explanation for why voice effects may not have been obtained in the present experiment. Recall that this framework stressed the importance of a match between the encoding operations used during the initial presentation of a stimulus and its later repetition. In the case of the present experiment, the task used during the study phase, lexical decision, and the task used during the test session may have required subjects to perform somewhat different types of processing on the stimulus input. The lexical decision task requires subjects to make conceptual judgment along the dimension of familiarity or meaningfulness (Balota & Chumbley, 1984). In contrast, the gating procedure does not involve a conceptual judgment but does requires close attention to the acoustic-phonetic structure of the input. The consequence of this mismatch in processing operations may account for the absence of voice effects in the present experiment.

In order to examine the hypothesis that the mismatch in the study and test conditions was responsible for the lack of voice effects, the experiment was run again with a different encoding condition at the time of study. In Experiment 3B, subjects participated in a perceptual identification task during the study phase and a gating task during the test phase. The perceptual identification task does not require subjects to make a conceptual judgment. Rather, the task is successfully completed by attending to the acoustic-phonetic structure of the input. This change in encoding tasks should alleviate the mismatch between the task used during the study phase and the task used during the test phase. As a result, we expected that listeners would require less signal duration for correct identification when responding to old words produced by a familiar talker. Furthermore, listeners may be more efficient when responding to a new voice that is perceptually similar to the old voice than a new voice that is dissimilar to the old voice.

EXPERIMENT 3B: Gating II

Method

Subjects

Twenty-two paid listeners served as subjects in the present experiment. All subjects were native speakers of English and reported no history of speech or hearing problems at the time of testing. Listeners were paid $10.00 for their participation.

Stimuli

The words used during the study session were identical to those used in the previous experiment. All of the nonwords were eliminated because the task used during the study phase was perceptual identification and no standard orthographic transcription exists for nonwords. The stimulus materials used during the test session were identical to those used during the previous experiment.

Procedure

During the study session, subjects performed a perceptual identification task. Before each trial began, subjects saw a warning prompt for 500 ms on the CRT terminal in front of them. After the prompt appeared, an isolated word was presented over the headphones at a comfortable listening level. Subjects typed their responses on the computer terminal. Listeners were given a maximum of ten seconds to make a response on each trial before the next test item was presented. Over the course of the study phase, subjects identified 48 words (24 high frequency and 24 low frequency) seven times each for a total of 336 trials. Subjects were given a short break after 168 trials. Each word was produced by seven different talkers. The word and the talker selected on each trial were randomly chosen. The study session lasted approximately 35 minutes. Listeners were given a longer break between the end of the study session and the beginning of the test session.

The gating procedure used during the test session was identical to the procedure used in the previous experiment. After presentation of the initial 50 ms of a word, subjects heard stimuli gated in 75 ms increments until the entire item was presented. The test session lasted approximately 50 minutes.

Results

As in Experiment 3A, the primary dependent variable from the gating test was the mean signal duration at which subjects first correctly identified the test items. Repetition status, word frequency, and voice similarity were treated as within-subjects variables in an ANOVA conducted on the mean recognition point.

A main effect for repetition observed in Experiment 3A was obtained in the present experiment [$F(1,21)=88.00$, $p<.01$]. Old items were correctly identified with a mean signal duration of 227 ms. New words were recognized with a mean gate duration of 303 ms. In addition to the repetition effect, a main effect for word frequency was also obtained [$F(1,21)=10.62$, $p<.01$]. High frequency words required a longer signal duration for identification than low frequency words, 277 ms vs. 254 ms respectively. No main effects for voice were obtained [$F(2,42)<1$].

Discussion

The purpose of Experiment 3B was to examine the use of voice information in an implicit task that was assumed to require careful analysis of the acoustic-phonetic input. In contrast to Experiment 3A, the encoding and testing conditions of the present experiment were very similar. In the study phase, listeners performed perceptual identification. During the test phase, subjects identified words in a gating paradigm. Despite the similar processing requirements of the two tasks, no evidence was obtained suggesting that listeners processed old words produced by a familiar talker more efficiently than words produced by a perceptually similar or dissimilar unfamiliar talker.

In the discussion of Experiment 3A, we suggested that the difference between the processing demands of lexical decision and perceptual identification might be responsible for the lack of voice effects in the gating task. While this difference in task demands may be partially responsible for the observed pattern of results, another possibility needs to be considered. The gating task is very dissimilar to other tests of implicit memory, such as lexical decision, perceptual identification, or auditory stem completion. In each of the latter tasks, subjects are only presented with the test items once. In the gating task, however, subjects are exposed to the same item numerous times within a single trial. The role of sophisticated guessing strategies and contamination from explicit recognition of the test words cannot be discounted (Lively et al., 1994).

Because of the large difference between the gating task and other tests of implicit memory and the potential for contamination from explicit recognition strategies, Experiment 3C was conducted using a more traditional test of implicit memory, auditory stem completion. As noted above, Schacter and Church (1992) found that changes in voice between the study and test session affected the magnitude of the repetition effect obtained when subjects completed word stems that were unmasked by noise. In the present experiment, subjects participated in the same perceptual identification task as listeners in Experiment 3B. However, instead of receiving a gating test, subjects in Experiment 3C were given an auditory stem completion test. Given Schacter and Church's finding and the results obtained by Goldinger using a perceptual identification test, we predicted that subjects would complete a higher proportion of stems with old items when they were produced by an old talker. Furthermore, stem completion performance was also expected to vary as a function of similarity among the voices: Listeners were predicted to correctly complete more word stems with old items when the stems were produced by a voice that was perceptually similar to an old voice than when the voice was perceptually dissimilar to the old voice.

EXPERIMENT 3C: Auditory Stem Completion

Method

Subjects

Subjects in the present experiment were 29 paid undergraduates from Indiana University. Twenty-one of the subjects served in the study and test conditions of Experiment 3C. Eight subjects participated only in the control condition and were only given the stem completion test. All subjects were native speakers of English and reported no speech or hearing problems at the time of testing. Listeners were paid $5.00 for their participation.

Stimuli

The materials used during the study and test conditions of the present experiment were identical to those used in Experiment 3B.

Procedure

The same perceptual identification task used during Experiment 3B was used during the present experiment. Subjects participated in 336 trials during the study session. During the test phase of the experiment, subjects were given an auditory stem completion task. Prior to the presentation of a stem, a warning prompt on subjects' computer terminals alerted them that a new experimental trial was about to begin. After the prompt had been displayed, subjects heard the first 200 ms of one of the test items. Listeners were instructed to type a word into the terminal that was consistent with the auditory stem that they heard. Subjects were given ten seconds to make a response before the next stem was presented. Stimulus presentation, timing control, and response collection were under the control of a PDP 11/34 laboratory computer. A total of 96 trials were presented during the test session. Each test session lasted approximately 15 minutes.

Results

The primary dependent variable in the present experiment is the proportion of stems correctly completed with the intended stimulus item. Separate ANOVAs were conducted on the mean percentage of items correctly completed in the control condition and in the experimental condition. In the analysis of the control condition, talker (old vs. similar vs. dissimilar) and word frequency were treated as within-subjects variables. In the analysis of the experimental condition, talker, word frequency and repetition status were within-subjects variables.

Control Condition

A main effect for word frequency was obtained in the analysis of the results from the control condition [$F(1,7)=13.19$, $p<.01$]. Stems from high frequency words were completed more accurately than stems from low frequency words (high frequency words: 21% correct; low frequency words: 14% correct). A main effect for voice was also obtained in the analysis of the control condition [$F(2,14)=3.98$, $p<.05$]. Stems produced by the new talker who was perceptually similar to the old talker were identified most accurately (25% correct). Tokens produced by the old talker were identified correctly on 17% of the trials and stems produced by the dissimilar talker were only correctly identified on 11% of the trials.

Study and Test Condition

Figure 5.1 shows the mean percentage of stems correctly identified as a function of repetition status and word frequency. A main effect for repetition status was obtained [$F(1,20)=121.63$, $p<.01$]. Stems derived from old words were completed significantly more accurately than stems derived from new words. A significant interaction was also obtained between the repetition status and the word frequency variables [$F(1,20)=13.35$, $p<.01$]. Larger repetition effects were obtained for low frequency words than for high frequency words. In addition, new low frequency words were identified less accurately than new high frequency words.

Insert Figure 5.1 about here.

In addition to the interaction between word frequency and repetition status, word frequency also interacted with voice [$F(2,40)=4.99$, $p<.02$]. High frequency stems produced by the similar new voice were completed more accurately than low frequency stems produced by the old talker or the dissimilar talker (old talker: 36% correct; similar talker 46% correct; dissimilar talker: 36% correct). However, a different pattern emerged for the low frequency stems. Low frequency stems produced by the old talker were completed more accurately than low frequency stems produced by either of the new talkers (old talker: 40% correct; similar talker: 33% correct; dissimilar talker: 33% correct).

Discussion

In Experiment 3C, the effects of changes in voice between a study and test session were examined in an auditory stem completion task, a test of implicit memory. Little evidence was found to suggest that changes in voice affected the magnitude of repetition priming effects: Subjects' accuracy in completing word stems in this task did not vary significantly when responding to a familiar talker, a perceptually similar talker or a dissimilar talker. This finding replicates the results obtained from the gating paradigm in Experiments 3A and 3B. The results reported in this chapter contrast with the earlier findings reported by Schacter and Church (1992) using an auditory stem completion task and by Goldinger (1992) using a perceptual identification paradigm.

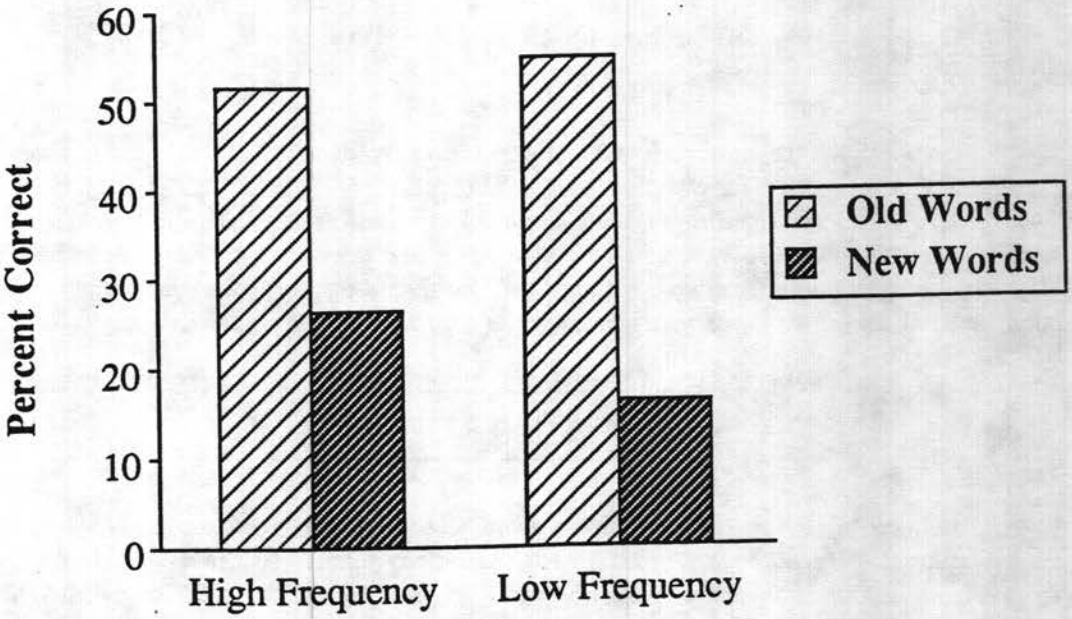

Figure 5.1. The figure shows the mean percentage of auditory stems completed correctly as a function of word frequency and repetition status.

Although the present results seem to contradict the hypothesis that talker-specific information is encoded and retained in long-term memory, several factors may have led to the failure to observe effects of changes in voice in the present experiments. First, consider the tasks used during the encoding phases of Experiments 3A, 3B, and 3C. As noted in Experiment 3A, subjects were given a lexical decision task during the study phase of the experiment and a gating task during the test phase. These two encoding tasks may require different types of processing and this mismatch may be responsible for the lack of voice effects during the test phase (Roediger, 1990). In order to alleviate this mismatch, a perceptual identification task was used as the encoding task during Experiment 3B and 3C, but no influence of changes in voice on repetition effects was observed in either of these experiments. However, it should be noted that the magnitude of the savings due to repetition was affected by changes in the encoding task. In Experiment 3A, a 31 ms advantage for old words over new words was obtained. In Experiment 3B, a 76 ms advantage was obtained for the same old words over the same new words. These findings indicate that the change in encoding tasks did have some effect on the test phase, but it did not influence the use of talker-specific information during the test conditions.

A second factor concerns the number of repetitions of each item during the study phase. In Experiments 3A, 3B, and 3C, subjects heard each item produced by seven different talkers during the study phase. The motivation for this manipulation was to expose listeners during the study phase to nearly the entire range of voices used to generate the multidimensional scaling solution. Recall that voices were assigned to the study and test phases of each experiment based on the scaling solution. In previous investigations, items presented during the study session have only been presented once (Goldinger, 1992; Schacter & Church, 1992). Multiple repetitions of items during the encoding phase may encourage subjects to engage explicit remembering strategies during test of implicit memory. This strategic use of explicit memory may mask the influence of changes in voices on repetition effects.

A third possibility relates to the choice of voices. In the present investigation, all of the voices were male native speakers of English with similar accents. As noted in Chapter 2, the range of durations across talkers was very limited because each speaker produced only monosyllabic words and nonwords. Furthermore, the range of fundamental frequencies across talkers was very narrow. The similarity among voices may have attenuated any small effects due to changes in voice between the study and test conditions.

Two points should be noted with regard to the range of voices selected in the present experiment and the similarity among them. First, the range of voices was not so narrow that voice effects were never obtained. In each of the recognition memory experiments reported in the present investigation, subjects were faster and/or more accurate in recognizing old words and nonwords repeated in the same voice. This suggests that experimental tasks may be differentially sensitive to the range of voices incorporated into the stimulus set. In cases where fine distinctions among speakers may serve as useful cues, such as recognition memory tasks, subjects' performance may be influenced even by a very narrow range of voices. However, in tasks where making small distinctions among similar talkers does not provide a useful set of cues, listeners' performance may be uninfluenced by a small range of voices in the stimulus set.

A second point to note is that other experiments have typically confounded changes in voice with changes in gender. Schacter and Church (1992) explicitly describe selecting their voices so that a change in voice between the study and test sessions also entailed a change in gender of the new talker. Goldinger (1992) manipulated voices so that changes in gender also occurred when different voices were presented during the study and test sessions. One problem with including male and female talkers in the test ensemble is that two categories of voices are present, *a priori*. This makes interpretation of results somewhat problematic because it is not clear whether a change in voice alone is responsible for an influence on the size of repetition effects or whether a change in gender is also important for observing such an influence (see, however, Palmeri et al., 1993). By including voices of only one gender and thus insuring that all changes in voice between study and test sessions occur within a gender, such problems in interpretation can be avoided.

A fourth factor that may have contributed to the present findings regarding changes in voice between the study and test sessions is the length of the words. In the auditory stem completion task, subjects were presented with the first 200 ms of monosyllabic words and were asked to complete the stems with the first word that came to mind that was consistent with the stem. The 200 ms stems, combined with the use of all male voices, may have been insufficient to differentiate among the voices. If subjects treated all of stems as being produced by the same voice, no effects of changes in talker between the study and test session would be expected. It should be noted that when Schacter and Church (1992) found an influence of changes in voice between the encoding and test phases of their experiments, they used multisyllabic words and also changed the gender of the talker.

Taken together, the results of Experiments 3A, 3B, and 3C failed to show any influence of changes in voice on the magnitude of the repetition priming effect. In the gating task, subjects required no more signal duration to correctly identify words when responding to stimuli produced by familiar or unfamiliar talkers. Similarly, listeners correctly completed the same proportion of word stems, regardless of the voice producing the stems. Several methodological factors, such as the choice of encoding tasks, the number of stimulus repetitions during the study phase, and the selection of specific voices, may have been responsible for the failure to observe voice-specific effects in the present experiments.

CHAPTER VI: General Discussion and Conclusions

Introduction

The purpose of the present investigation was to examine the influence of changes in voice on spoken word recognition. The traditional view of speech perception and spoken word recognition has assumed that information about a speaker's voice is discarded early on in spoken language processing and that this information is not useful for decoding the phonetic content contained in the acoustic waveform. However, a number of recent studies have demonstrated that trial-to-trial changes in the voice of the talker producing the test materials does have an effect on the speed and accuracy of word recognition. In the present investigation, the effects of changes in voice were examined in tasks that tapped either explicit or implicit memory. The explicit task was designed to tap the surface features of spoken words by allowing subjects to use information about a talker's voice as a cue to recognition. The implicit memory tasks encouraged subjects to either make use of conceptual information or to use information about the surface forms of spoken words.

The motivation for the methodology used in the present investigation was drawn from research carried out using the transfer appropriate processing framework (Roediger, 1990; Roediger & Blaxton, 1987) and findings reported by Blaxton (1989). Blaxton argued that experimental memory tasks could be described along at least two different dimensions. First, she suggested that memory tasks differ in the degree to which they require subjects to make explicit judgments about the contents of long-term memory. Tasks such as recognition and recall are considered to be explicit memory tasks because they require subjects to consciously probe the contents of long-term memory for particular items. In contrast, tasks such as lexical decision and auditory stem completion with repeated items are thought to be tests of implicit memory because they require little or any conscious access to the original presentation of the item.

The second dimension concerns the types of information that are used to successfully complete the task. Blaxton (1989) draws a distinction between tasks that rely heavily on data-driven information versus tasks that rely mainly on conceptual information. Procedures such as recognition memory in which old and new voices are used to create the test items may be considered data-driven tests because the voices or surface forms may serve as cues to facilitate recognition. In contrast, tests such as lexical decision or semantic categorization are thought to be more conceptually-based tasks because they require subjects to access facts about the items, rather than particular instances. Tests that emphasize careful analysis of the surface forms of stimuli may be considered episodic tests, whereas tests that stress the use of conceptual information may be considered more semantic in nature.

In the present investigation, we predicted that information about a speaker's voice would be used differentially according to the type of memory task subjects were asked to perform. Based on previous research, it was expected that recognition memory would be affected by changes in voice between the initial encoding phase and the subsequent test phase. Similarly, it was anticipated that tasks such as gating and auditory stem completion would be affected by changes in voice between the study and test phases of the experiment. The recognition memory task and the gating and auditory stem completion tasks are similar in that they are assumed to require careful attention to the surface forms of spoken words. The tests differ in several ways. The recognition memory task requires explicit access to stored instances, whereas the gating and auditory stem completion tasks only require implicit access to previously stored examples.

The predictions for the gating and auditory stem completion tasks can be contrasted with the predictions for the lexical decision test. All three tests are tests of implicit memory because they do not require subjects to consciously recollect and access previously stored instances. However, the three tasks are assumed to differ in terms of the types of information that they require for completion. Whereas gating and auditory stem completion may require careful attention to the acoustic-phonetic input, the lexical decision task relies more heavily upon

conceptual information. As a consequence, it was predicted that performance on the gating and stem completion tasks would be influenced by changes in the talkers' voices between the study and test sessions. However, it was also expected that changes in voice would not influence repetition effects in lexical decision.

A. Summary of Major Results

The "same-different" experiment in Chapter 2 was conducted to obtain a measure of similarity among the voices used in the stimulus ensemble. All of the talkers were male native speakers of English with similar accents. Multidimensional scaling analyses were applied to the results and talkers were assigned to the study conditions and test conditions of the later experiments. The talkers used in the study condition were all perceptually similar to each other. However, the talkers used in the test sessions were a familiar voice from the study sessions, a new voice that was perceptually similar to the old talker, and a new voice that was perceptually dissimilar to the old talker.

After the similarity space for the voices was derived, the effects of changes in voice between a study session and test session were examined in a series of explicit and implicit memory tasks. The findings from the explicit recognition memory tasks were very consistent across experiments: Subjects were always faster and/or more accurate at recognizing old words and nonwords repeated in the same voice as their original presentation. These results were obtained under conditions in which subjects heard a single talker or multiple talkers during the study session as well as when the test phase of the experiment was delayed by twenty-four hours.

Another consistent result from the recognition memory experiments was that only marginal evidence was obtained indicating that perceptual similarity among the voices had an impact on recognition. Although trends were observed in some experiments, many of the relevant comparisons among the old voice, the similar talker and the dissimilar talker failed to reach statistical significance. As noted in Chapter 5 with regard to the results from the gating and stem completion experiments, a very narrow range of voices was employed in the present investigation and this may have limited the ability to detect similarity effects in many of the experimental tasks.

The results from the lexical decision tests were also highly consistent across experiments. As anticipated by the framework described by Blaxton (1989), the magnitude of the repetition priming effect was not influenced by changes in voices between the study phase and the test phase of any experiment. Differences in response times among talkers observed during the test phases of each experiment closely match differences obtained with subjects from the control conditions. This similarity of results between the experimental and control conditions indicates that any *a priori* differences among talkers were not overcome by familiarizing subjects with one of the voices used during the test phase.

Although repetition effects in lexical decision were not influenced by changes in talkers, the magnitude of the repetition effect was affected by word frequency. Consistently larger benefits were observed when low frequency words were repeated than when high frequency words were repeated. In addition, error rates for new low frequency words were extremely high, relative to old low frequency words and new high frequency words. This bias against new low frequency words was observed both when subjects were tested immediately after the study session and when they were tested after a twenty-four hour delay.

In the final series of experiments, the influence of changes in voices between study and test conditions were examined in gating and auditory stem completion tasks. Significant repetition effects were observed in each experiment. Repeated words required shorter signal durations than new items for correct identification in the gating task. Similarly, subjects correctly completed a higher proportion of stems from old items than from new items. Although these implicit memory tasks were assumed to rely on data-driven processing, little if any

evidence was found to suggest that the magnitude of the repetition effect was influenced by changes in voices between study and test.

B. Conclusions from the Present Investigation

The findings from the present investigation suggest several conclusions regarding the nature of similarity among voices and the encoding and use of voice information during implicit and explicit tests of memory. First, the results of the same-different experiment suggest that acoustic factors play a large role in determining the similarity among voices. These factors appear to be independent of the lexical status of the items that the talkers produced: Highly similar multidimensional scaling solutions were obtained when responses to words and nonwords were scaled separately. However, it should also be noted that the acoustic correlates that correspond to the psychological dimensions that determine similarity may not be obvious. In the present investigation, factors such as pitch and duration did not account well for the underlying dimensions of the multidimensional scaling solution.

Second, with regard to the encoding of voice information, the results from the recognition memory experiments indicate that features of a speaker's voice appear to be automatically encoded into memory during lexical decision (Geiselman, 1979; Geiselman & Bellezza, 1977). These findings contrast with traditional abstractionist assumptions regarding speech perception and perceptual normalization, which suggest that the acoustic signal is rapidly converted into an idealized symbolic form and that source characteristics are lost during perceptual analysis.

Third, the present findings indicate that listeners have explicit access to encoded voice information even twenty-four hours after the stimuli were initially presented. This result was consistently obtained despite the fact that listeners heard several different voices during each study session. Based on results obtained by Goldinger (1992), however, it might be expected that explicit access voice information encoded in long-term memory may fade over time if the tests were carried out over longer periods of time.

Fourth, the results from the implicit tests, such as lexical decision, gating, and auditory stem completion, suggest that while information about a speaker's voice may be automatically encoded during lexical processing, this information is not always engaged to facilitate the later reprocessing of the same words. Several different factors may be responsible for the failure to find voice effects in the implicit memory tasks. With regard to the lexical decision task, subjects are required to make conceptual judgments about spoken utterances. Within the theoretical framework described by Blaxton (1989), we would not expect subjects to be sensitive to changes in surface forms in a such a task. In contrast, a number of stimulus and task-related factors may be responsible for the failure to find voice effects in the gating and stem completion tasks used here. Other recent studies have shown that stem completion and perceptual identification are sensitive to changes in surface forms between the study and test conditions (Church & Schacter, 1994; Goldinger, 1992; Schacter & Church, 1992). These findings argue against the possibility that all three implicit memory tasks used in the present investigation tap conceptual information and should therefore be insensitive to changes in voice between the encoding and test sessions.

Taken together, the results of the present investigation suggest that information about a speaker's voice is automatically encoded during spoken word recognition. Under some conditions this information can be accessed and used to facilitate recognition. These conditions typically require subjects to engage in a careful analysis of the inputs. However, speaker-related characteristics are not always used during the reprocessing of spoken words. Under some conditions, task demands may not require subjects to engage talker-specific details to facilitate processing. Under other conditions, a sufficiently wide range of voices may be needed to detect the influence of changes in voice on implicit memory.

C. Implications for Models of Speech Perception and Spoken Word Recognition

The results of the present investigation add to a growing body of research that suggests lexical processing takes place through the use of highly detailed information derived from perceptual analysis of the incoming signal. This perceptual analysis appears to operate by accessing previously stored, talker-specific information. In the present section, the results from the present investigation are combined with findings from previous reports to suggest that traditional models of speech perception and spoken word recognition are in need of changes that incorporate the use of talker-specific information as a part of the recognition process. In particular, models of spoken language processing need to be modified to incorporate aspects of exemplar-based models of categorization and long-term memory.

C.1. Implications for Models of Normalization

One problem that the incidental storage and use of talker-specific information raises for current theories of speech perception concerns information reduction. As noted in the introduction, information about a talker's voice has traditionally been treated as noise that is filtered out of the signal at a relatively early stage of processing because it has been assumed to be uninformative to the recovery of the phonetic content of the message (Liberman & Mattingly, 1986). By removing information about a speaker's voice from the input signal, the acoustic-phonetic stream can be treated as a discrete set of phonemes that can be manipulated and recombined to form syllables and words.

The problem with this approach to speech perception is that the filtering mechanism is often underspecified. For example, in their revision of the motor theory of speech perception, Liberman and Mattingly explicitly argue that characteristics of a speaker's voice are unknown to the "phonetic module" and that these features may be used as the input to another module that is responsible for identifying the voice of the speaker. Although this strategy of separating the linguistic message from the indexical properties of spoken language may simplify phonetic analysis, it is difficult to define the nature and scope of these processing modules and to develop testable models that incorporate constructs such as a phonetic module or a "voice" module (Klatt, 1989; Nearey, 1989; Jusczyk & Cohen, 1985).

Because most theories of speech perception do not give satisfactory accounts for how changes in voice affect listeners' efficiency in processing spoken language, theories of perceptual normalization have been developed to explain these effects. The guidelines for theories of normalization may be developed without reference to any particular model of speech perception (Nusbaum & Morin, 1992). Typically, these models of normalization have been limited in scope because they only address how listeners adapt to changes in voice (Gerstman, 1968; Johnson, 1989; Lieberman, 1973; Syrdal & Gopal, 1986). Normalization or compensation for other sources of variability in the speech waveform have been addressed by separate models that deal only with factors such as speaking rate (Miller & Volaitis, 1989; Volaitis & Miller, 1992).

As described in the introduction, two different types of models for speaker normalization have been proposed. These models are differentiated by the size of the speech sample that is required for the normalization process to function. Intrinsic models suggest that all of the information needed for normalization is contained within a single vowel or syllable. Acoustic cues such as the fundamental frequency and the third formant are relatively invariant across vowel contexts for a single speaker and these cues could provide a basis for interpreting the relationship between the first and second formant (Syrdal & Gopal, 1986). Intrinsic theories of normalization, on the other hand, can be subdivided into two classes of models. One class suggests that static cues, such as steady-state formant values, are used by the normalization mechanism. A second type of intrinsic normalization mechanism relies on dynamic information specified in vowels by the transitions of preceding and following consonants (Strange, 1989).

The other broad class of theories of normalization suggests that listeners rely on much longer portions of speech to calibrate a speaker's voice. In these extrinsic or contextual tuning theories, listeners are assumed to require more than a single vowel in order to map out the vowel space of a particular talker (Joos, 1948; Ladefoged & Broadbent, 1957). Information about point vowels is assumed to be particularly useful in this process because the point vowels define the extremes of a talker's vowel space (Gerstman, 1968; Sawusch, Nusbaum & Schwab, 1980). Once sufficient information about a voice has been gained, that information is used as a template for the normalization and recognition of subsequent speech from that talker.

Although some evidence has been gathered in favor of both models of normalization, no definitive empirical support has been provided for either type of model. For example, studies on listeners' accuracy of identification of "silent-center" vowels point to the importance of dynamic cues to vowel perception and support the intrinsic models (Jenkins, Strange, & Miranda, 1994; Verbrugge & Rakerd, 1986). Silent-center vowels occur when subjects are only presented with an initial and final consonant. The intermediate vowel is removed digitally from the waveform. The consonants contain critical information about the transitions into and out of the vowels and may be produced either by the same talker or by two different talkers. The typical finding is that the dynamic cues provide sufficient information for listeners to accurately identify the vowels (Shankweiler, Strange, & Verbrugge, 1977; Strange, 1987; Verbrugge & Rakerd, 1986) .

Other findings suggest the operation of an extrinsic normalization mechanism. For example, Ladefoged and Broadbent (1957) reported that a preceding sentence context improved vowel identification. However, Nusbaum and Morin (1992) found evidence for both intrinsic and extrinsic normalization mechanisms in a series of speeded classification tasks. They argued that an intrinsic normalization mechanism is engaged when the voice of the talker changes during an experimental trial. However, when a speaker's voice remains constant throughout a trial, listeners engage an extrinsic normalization mechanism that tunes to the voice over larger periods of time.

The present findings and other recent results indicating that talker-specific information is encoded and accessed to increase the efficiency of lexical processing are problematic for traditional models of normalization. Although most models do not explicitly claim that information about a speaker's voice is lost during the normalization process, the assumption is made tacitly when the purpose of the normalization mechanism is considered. Recall that the purpose of positing a normalization device was to "sanitize" the acoustic signal of sources of variability because these sources of variability were assumed to be uninformative to the recovery of the phonetic message. However, the present findings demonstrate that information about the acoustic signal is not filtered out early on during processing. Indeed, Mullennix and Pisoni's (1990) findings using a Garner speeded classification task showed that linguistic information and information about a speaker's voice are integrally perceived. Furthermore, the present results, combined with findings reported by Craik and Kirsner (1974), Palmeri et al. (1992), and Goldinger (1992), indicate that listeners do retain and have explicit access to information about a speaker's voice, at least over a period of twenty-four hours. Taken together, the findings suggest that a perceptual normalization mechanism may not be in operation during the processing of spoken language.

If a normalization mechanism for speakers' voices is assumed not to play a role in the early processing of speech signals, an account must be given for why listeners show deficits in performance when multiple talkers are represented in the stimulus ensemble and for why listeners also tend to show an advantage in some memory tasks when words are repeated in the same voices they were initially presented in. Models of normalization account for the finding that listeners are slower and less accurate to process words produced by multiple talkers by assuming that the normalization mechanism usurps attentional resources from the processes that are dedicated to recovering the linguistic content of the message (Mullennix et al., 1989; Nusbaum & Morrin, 1992). However, the attentional demands of the normalization mechanism do not account for the finding that recognition

is facilitated when words are repeated in the same voice because the tacit assumption of models of normalization is that the process entails a loss of information.

An alternative is to assume that variability in the waveform attracts attention. However, instead of discarding sources of variability, this information is encoded into long-term memory along with a linguistic interpretation of the input signal. One consequence of dividing processing attention between the linguistic and nonlinguistic aspects of the input is that the efficiency of spoken language processing suffers (Baddeley & Hitch, 1974; Luce, Feustel, & Pisoni, 1983; Mullennix et al., 1989; Navon & Gopher, 1979). This assumption may be used to account for the processing deficits that are typically associated with talker variability in identification paradigms (e.g. Mullennix et al., 1989).

It is interesting to note that all sources of variability in the speech waveform may not require the same amount of attentional resources. Sommers, Nygaard, and Pisoni (1992) investigated the independent and combined effects of several sources of variability on perceptual identification. They found that sources of variability such as changes in the voice of the talker and changes in speaking rate from trial to trial were detrimental to listeners' abilities to identify spoken words. The impact on performance was even larger when changes in voices were combined with variations in speaking rate. However, changes in the amplitude of the signal over a 30 dB range did not have an influence on perceptual identification accuracy. Taken together, the findings reported by Sommers et al. suggest that some sources of variability may usurp more attention from linguistic processing than others.

The assumption that sources of variability are encoded into memory, rather than lost during early processing, may be used to account for the findings that word recognition can be facilitated by repeating an item in the same voice. Many current models of long-term memory assume that information about an item and its surrounding context are encoded into memory as a result of perceptual processing (Gillund & Shiffrin, 1984; Hintzman, 1986). Information about a speaker's voice may also serve as a part of the context. When a new stimulus is presented, similar items in memory are activated. Stored items that match the context or voice of the new stimulus are more strongly activated than items that do not match the context. This difference in activation levels for stored items contributes to the observed advantage in recognition memory for items repeated in the same voice (Craik & Kirsner, 1974; Goldinger, 1992; Palmeri et al., 1992).

The account given above suggests that variability in the signal draws attention away from processing the linguistic message in the input signal. One consequence of the decrease in attention to the linguistic characteristics of the signal is that spoken language processing becomes less efficient and more error prone. Another consequence of this attentional account is that the sources of variability become automatically encoded into memory and may be available to facilitate the later reprocessing of spoken words. Such an account calls into question the necessity and viability of a normalization mechanism that reduces the speech waveform to a sequence of abstract, idealized symbols.

While the description of the role of attention in spoken language processing may give a reasonably good account for why variability has a negative impact on performance in some cases and a positive impact on performance in others, the account also points to an important direction for further empirical investigation. Goldinger (1992) correctly noted that all sources of variability do not have the same effects on perceptual identification accuracy (see Sommers et al., 1992). He argued that different sources of variability may not be equally salient to listeners. For example, a change in the gender of the talker may be highly salient to listeners, whereas a change in amplitude may not convey an important contrast. Future research needs to address the issue of how listeners scale different sources of variability and how these sources of variability may be equated in psychological terms. By equating these different sources of variability along the same scale, more precise investigations can be made into the nature of processes listeners employ to deal with sources of variability in the input signal.

C.2. Implications for Theories of Speech Perception

In addition to addressing issues related to models of talker normalization, results indicating the storage and use of talker-specific information during spoken language processing also have implications for a number of other issues in speech perception. One key issue concerns the modular nature of the mechanisms engaged during speech perception. In describing the "phonetic module," Liberman and Mattingly (1985) draw heavily on the ideas concerning modularity developed by Fodor (1983). The phonetic module is thought to be responsible for recovering the intended articulatory gestures of the speaker. Liberman and Mattingly argue that a specialized neural structure recovers a set of primitives that are used in both production and perception. These primitives do not represent the acoustics of the intended signal, rather they specify the movements of the articulators.

The phonetic module is a more specialized version of what Fodor (1983) described as the "language module." As such, the phonetic module has many of the same properties of the language module. First, all modules are assumed to be served by neurologically distinct mechanisms in the brain. Second, these distinct mechanisms preemptively operate only on domain-specific information. The phonetic module's domain is to recover the talker's intended articulatory gestures from the time-varying acoustic signal. As noted in the introduction, Liberman and Mattingly (1985) argue very explicitly that information related to the voice of the talker is shunted to another processing mechanism. Third, because the phonetic module is assumed to operate on only a limited range of information, the module operates very quickly and only gives a shallow output. In the case of the phonetic module, the output is related to the intended gestures of the talker. Fourth, the domain specificity and the speed of the module's operation lend it some degree of cognitive impenetrability: Information or representations recovered outside of the phonetic module cannot be used to influence or bias processing within the module. Furthermore, observation of intermediate representations computed within the module is assumed to be impossible. Once the phonetic module has been engaged by an auditory signal it mandatorally completes an analysis of the input before it makes a low-level representation available for inspection by other linguistic modules (Miller, 1987).

Several points can be made with regard to issues of modularity and the mandatory encoding of voice information during spoken language processing. First, it is extremely difficult to design empirical investigations that can test many of the assumptions of the modular framework. Klatt (1989) notes that the motor theory of speech perception has provided the motivation for intense research activity for many years. However, this theory serves more as a philosophy than a model because it is very difficult to make any testable predictions about the theory. Furthermore, the assumption that the objects of perception are the intended articulatory gestures of the speaker is underspecified. It is not at all clear how the phonetic module actually recovers intended articulatory gestures from the acoustic signal. Furthermore, given the assumption that modules are cognitively impenetrable, it may the case that no set of behavioral experiments can show how an intended gesture is recovered or what the nature of an intended gesture is.

A second point to note about the encoding of talker specific information and modularity concerns the domain of the phonetic module. According to the assumptions of the Motor Theory, the phonetic module only operates on information related to intended gestures. However, Mullennix and Pisoni (1990) have shown that voice and phonetic information are integrally perceived. Their findings indicate that the mechanism responsible for processing the phonetic form of the input may also have to consider the source or the voice producing the input. If information about a talker's voice does exert an influence on the recovery of a phonetic code, this suggests that the phonetic module may not be impenetrable to sources of information computed in other modules or that the scope to the input to the phonetic module needs to be broadened.

The assumptions of modular processing systems provide an important and controversial framework for conducting research. Issues related to the cognitive impenetrability and neural specialization of modules are

particularly challenging at this time. Research on the influence of the sources of variability on spoken language processing may provide some insight into both the number of proposed modules and their domains of operation. Furthermore, investigations conducted with brain damaged patients may give some insight into the autonomy of modules related to the processing of phonetic information and modules responsible for processing the identity of the talker (Van Lancker, 1991; Van Lancker et al., 1988; Van Lancker et al., 1989; Zattore, Evans, Meyer, & Gjedde, 1994).

Another important issue related to the encoding and use of talker-specific information concerns whether speech perception is direct or whether it is mediated by cognitive influences. The Motor Theory of speech perception provides one example of mediated speech perception: The computational power of the phonetic module is used to compare the input against a small set of candidates that describe the potential intended gestures in the signal. This is a form of analysis-by-synthesis (Stevens & Halle, 1968) in which actual information about the vocal tract is compared to idealized or intended parameters that describe the allowable and linguistically meaningful dynamics of different articulators (Liberman & Mattingly, 1985; Mattingly & Liberman, 1969). Perception is said to be mediated because it requires cognitive architecture to perform a comparison process in order to arrive at the identity of the phonetic percept.

The mediated perception of the motor theory can be contrasted with the direct-realist approach to speech perception offered by Fowler and her colleagues (Fowler, 1986, 1990; Fowler & Rosenblum, 1990, 1991). The two models are similar in many respects. Most importantly, both use phonetically relevant gestures as their primitives in perception. However, the direct realist perspective differs from the motor theory in its approach to modularity and the specialness of a phonetic module. Fowler and Rosenblum (1991) argue that speech perception is similar to visual perception in Gibsonian terms (Gibson 1966, 1979). In both speech perception and visual perception, perceivers recover distal events that structured the stimulation engaging the sensory organ. Thus, just as perceivers recover chairs from patterns of light, they recover linguistically relevant articulatory gestures from patterns of sound. These gestures are grouped together to form phonemes, syllables and words (Browman & Goldstein, 1985, 1986). Furthermore, they are recovered directly and do not require mediation by a specialized module. Whereas Liberman and Mattingly attempt to focus on the cognitive architecture that underlies speech perception, Fowler and her colleagues focus on the information that is contained in the signal itself.

Two aspects of the direct-realist approach to speech perception are appealing. First, careful consideration is given to the information that is in the signal, rather than trying to give an account of the underlying cognitive architecture. Second, the theory is appealing because the same principles are applied to perception, regardless of the input modality. This prevents one type of percept, speech for example, from taking on a privileged status over any other type of percept.

Despite these appealing qualities, it is unclear how a direct-realist theory of speech perception accounts for findings that talker-specific information is encoded and used under certain circumstances to facilitate the later reprocessing of a spoken word. Fowler and Rosenblum (1991) argue that learning does take place within the perceiver, while leaving the object of perception unchanged. Thus, listeners may become familiar with voices, while the voices themselves do not change. However, it is unclear how findings that specific perceptual episodes influence perception can be incorporated into a system that has no cognitive mediation (Goldinger, 1992).

A final critical issue that models of speech perception have to address concerns the nature of the primitive units used in perception. In the two theoretical frameworks outlined above, articulatory gestures played a key role in the recovery of phonetic segments. An alternative perspective is that listeners do not rely on the recovery of articulatory gestures at all. Instead, the acoustics of the speech waveform are used to generate linguistic representation (Klatt, 1979, 1980). Unfortunately, findings that indicate the incidental storage and use of talker-specific information are unlikely to provide evidence for either side of this debate. While inferences

can be made about the types of talker-specific information that get encoded into long-term memory, little more can be said about whether that information is encoded in terms of articulation or acoustics or both.

Although the Motor Theory (Liberman & Mattingly, 1985) and the direct realist framework offered by Fowler and her colleagues (Fowler, 1986, Fowler & Rosenblum, 1990, 1991) may not easily account for the encoding and use of episodic information related to a talker's voice, Klatt's Lexical Access from Spectra (LAFS) (Klatt, 1979, 1980, 1986) model does make an effort to deal with different sources of variability in the acoustic waveform. LAFS attempts to recognize spoken words directly from their spectral representations. Thus, it represents an example of an acoustic theory of speech perception rather than an articulatory theory. The LAFS model consists of a large precompiled network of interconnected diphones. Diphones are measured from the center of one phoneme to the center of the next phoneme and are assumed to be sufficiently invariant within a speaker to provide accurate recognition directly from the input spectrum. The spectral representations of the diphones are linked together using a set of phonological rules that handle phenomena such as reduction and deletion.

Syllables and words are recognized in LAFS by activating spectral templates that are consistent with the acoustic input. Multiple paths through the network are typically activated by an input signal. Commitment to any particular interpretation is delayed until a beam search is conducted back through the network to find the most highly activated path. Once this path has been determined, a phonetic representation can be given to the activated pattern of spectral diphone sequences. The beam search techniques and the idea of delayed commitment to a particular representation are very similar to concepts that were incorporated in the HARPY speech understanding system (Lowerre, 1976; Lowerre & Reddy, 1978).

As noted above the LAFS model was designed to be robust to stimulus variability in the ambient listening environment. Two mechanisms were incorporated to handle compensation for variations among talkers. First, Klatt (1979) assumed that the diphone templates in the model are talker-specific. Thus, knowledge of spectral sequences is not stored in the network in a talker-independent manner. Second, information about a speaker's dialect can be derived by examining how sequences of diphones are interconnected for a particular talker. The patterns of interconnection represent the phonological rules used by the speaker. By incorporating these two assumptions, LAFS is able to account for the storage and use of talker-specific information.

In general, findings indicating the incidental storage and retention of talker-specific information are difficult to account for within most frameworks proposed for speech perception. Many of these models were devised to account for phenomena that are very different from those discussed in the present report. Thus, the present findings may be considered beyond the scope of those models (Goldinger, 1992). While most models of spoken language processing attempt to explain how the speech signal is translated into a symbolic code, the present findings suggest that a broader set of issues needs to be addressed. Given that it is well established that information about a speaker's voice does have an impact on spoken language processing, it becomes increasingly important to develop a theoretical framework that can account for the traditional phenomena in speech perception as well as the growing body of new findings.

D. Implications for Models of Spoken Word Recognition

The present results and other related findings that indicate the incidental storage and use of talker-specific information during spoken language processing have several important implications for models of talker normalization. Whether the present findings also speak to issues of modularity, the specialness of speech, and the nature of the information used during perception is debatable. The present findings have little to contribute to these issues if speech perception is viewed within the narrow context of phoneme perception. However, if the traditional distinction between phoneme perception and word recognition is blurred (e.g. Klatt, 1979, 1980; McClelland & Elman, 1986), the present findings and related results suggest that some important changes need

to be incorporated into models of spoken language processing. In the present section several traditional design principles of models of spoken word recognition are outlined. Examples of models that embody these principles are then described. Finally, some new directions and design principles for models of spoken word recognition are introduced based on findings that suggest talker-specific information plays an important role in spoken word recognition.

Although models of word recognition differ to some degree in their specific assumptions, a number of common design principles have been incorporated into many of the current models of spoken word recognition. As a consequence, the models have become very similar in terms of their architectures and predictions (Forster, 1989; Marslen-Wilson, 1990). First, as noted in the introduction, most of the current models of spoken word recognition were developed as adaptations of models of visual word recognition. One result of this change in modalities has been that the nature of the input signal has also been changed. Instead of accepting an orthographic transcription as the input to the word recognition system, most models of spoken word recognition select candidates for recognition based on some type of phonetic or featural representation of the input (Luce, 1986; Marslen-Wilson, 1987; see, however, Klatt, 1979, 1980). It should be noted that just as models of visual word recognition are not typically concerned with how the letters themselves are perceived, models of spoken word recognition do not specify how the phonetic representations that drive the recognition system are derived.

A second common theme running through contemporary models of spoken word recognition is that words are recognized through an activation and selection process. In general, words in the mental lexicon are assumed to be activated to the degree that they are similar to the input representation. Typically, this involves the activation of a set of potential candidates. Words are ultimately recognized on the basis of a small set of selection rules, such as goodness-of-fit statistics or Luce's biased choice rule (Luce, 1959; Paap, Newsome, McDonald, & Schvaneveldt, 1982). It is interesting to note that models that were originally designed to accomplish lexical access through a strict serial search strategy have been modified in recent years to include an activation metaphor (Forster, 1976, 1987).

The final assumption adopted by all models of spoken word recognition is that a mental lexicon is accessed when words are recognized. The mental lexicon can be accessed very rapidly, is assumed to be quite large, containing as many as 75,000 words, and is assumed to be located in a specialized region of the brain's left hemisphere (Oldfield, 1966; Marslen-Wilson, 1987). Each lexical entry may contain information about a word's phonological or orthographic form, its syntactic class and its meaning. All of this information focuses on the linguistic aspects of a word. Typically, little consideration is given to "episodic" properties related to a word such as who produced it or in what context was the word presented. Because episodic information is not incorporated into lexical representations, the findings that talker-specific information are encoded and used to facilitate lexical processing are somewhat problematic for traditional models of spoken word recognition.

One example of a model that incorporates each of the three design principles described above is Morton's logogen model (1969, 1970, 1982). The logogen model was developed to account for findings from auditory and visual word recognition tasks. The model has separate input paths for auditory and visual information. In its present form, the model is underspecified as to how acoustic-phonetic information is derived from the speech waveform (Lively et al., 1994; Pisoni & Luce, 1987). The products of auditory and visual analysis of the input signal are used to directly access a set of passive detecting units, referred to as "logogens," stored in the mental lexicon. Each logogen is assumed to represent an individual word and contains abstract information about the orthography, phonology, syntactic class, and meaning of the word. Logogens are activated in memory according to their similarity to the acoustic-phonetic or orthographic input. The resting activation levels of individual logogens are set in a frequency sensitive manner, so that high frequency words cross a recognition threshold prior to low frequency words. Once a logogen has become sufficiently active to cross a recognition threshold, it becomes available to a more general cognitive system. With regard to the issue of the encoding and use of talker-specific information during lexical processing, Jackson and Morton (1984) have

maintained that logogens do not encode episodic details. They base this claim on a failure to observe a difference in the size of the repetition effect obtained for words repeated in the same voice versus words repeated in a different voice in a perceptual identification task.

Another example of a model that incorporates the design principles described above is Marslen-Wilson's cohort model (Marslen-Wilson, 1987, 1990; Marslen-Wilson & Welsh, 1978; Marslen-Wilson & Tyler, 1980). Marslen-Wilson developed the model to account for the speed of spoken word recognition. Cohort theory assumes that abstract entries in the mental lexicon are initially activated through bottom-up, left-to-right analysis of the incoming acoustic-phonetic signal (Cole & Jakimik, 1980). The set of candidates that is initially activated is referred to as the "word initial cohort." In the original version of the model (Marslen-Wilson & Welsh, 1978), all items in the word initial cohort were assumed to begin with the same phoneme. In later versions of the model, however, the concept of a word initial cohort was revised so that items in the cohort were defined in terms of features, rather than phonemes (Marslen-Wilson, 1987). Items are removed from the word initial cohort as they become inconsistent with the acoustic-phonetic input. Marslen-Wilson (1984, 1987) argues that the word recognition occurs when the target item diverges from all other possible candidate words. Thus, spoken word recognition involves making decisions about what is present in the signal (identification) and about what is not present (discrimination) (Luce, 1986).

The cohort model is vague with respect to its assumptions about the nature of lexical representation and how talker-specific information could be incorporated into lexical processing. Marslen-Wilson (1987) weakened the original form of the model by defining word initial cohorts in terms of features rather than phonemes. Similar to Klatt (1980), he argued that a highly categorized input signal may be prone to errors that are difficult to recover from. In order to correct this shortcoming, he suggested that an input representation, based on phonetic features, might preserve more of the original signal and could be useful in recovering from errors. An extension to the model that would help it to deal with the present findings would be to assume that lexical representations encode talker-specific information, in addition to linguistic information, and that this information about a talker's voice serves as additional bottom-up information that can be used to guide initial lexical processing.

Logogen theory was designed to explain findings from auditory and visual word recognition and cohort theory was developed to account for the efficiency of spoken language processing. Both models suffer from two general problems: First, the procedure through which the perceptual system derives an input representation for the word recognition system is basically left underspecified. Both models assume that a phonetic or featural code is derived from the speech waveform without an explanation of how that code is derived. The second problem for both models is that neither provides a mechanism for encoding or using talker-specific information. Supporters of logogen theory argue on the basis of a null result that the lexicon contains only abstract information (Jackson & Morton, 1984). Cohort theory is ambiguous on the issue too.

While the logogen model and cohort theory were designed mainly to account for phenomena in word recognition, TRACE was designed to capture the mutual influence of phonemic activation on lexical processing and lexical activation on phonetic processing (Elman & McClelland, 1986; Ganong, 1980; McClelland, 1991; McClelland & Elman, 1986; Samuel, 1981). TRACE is a connectionist model based on the interactive activation model of visual word recognition (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982). Information in TRACE is represented by three types of processing units or nodes: Feature nodes are designed after phonetic features from linguistic theory and accept input from the environment. Feature units are unidirectionally connected via excitatory links to phoneme units. Phoneme units are bidirectionally connected to word nodes. Both connections between phonemes and words are facilitatory in nature. Units within a particular level of representation are also interconnected by inhibitory connections. Activation and inhibition of units at all levels of representation are gated by the connection strengths between units and the excitation levels of the units sending the activation.

Processing in TRACE begins when a set of feature nodes are activated by an external input. This activation is fed forward to the phoneme units and from the phoneme units to the word nodes. Over time, the model accrues additional external activation and feedback is collected at lower levels of representation from higher levels of representation. Feedback makes TRACE a strongly interactive model. The pattern of activation that develops across the network is referred to as the "trace." As with most activation-based models of word recognition, several traces or candidates for recognition typically become active. Luce's biased choice rule is used to select from among the available traces.

TRACE is a powerful model of spoken language processing because it captures much of the interaction between speech perception and spoken word recognition (see however, Massaro, 1989). However, the model has no way to incorporate the present findings or other results indicating the storage and use of talker-specific information. Furthermore, it is not obvious how the concept of episodic traces could be incorporated into an interactive-activation processing framework.

Taken together, the traditional models of spoken word recognition described in the present section do not account well for the findings that listeners encode and use talker-specific information during spoken word recognition. To be fair to the models described above, it should be noted that they were not designed to handle these problems. Rather they were developed to address problems such as the word frequency effect, the efficiency of lexical processing and the interaction between phonetic and lexical activation. Thus, the problems explored in this report are somewhat outside of the domains of the models presented here. However, the findings reported here and elsewhere (Church & Schacter, 1994; Goldinger, 1992; Schacter & Church, 1992) do present several new and important theoretical challenges for future models of spoken word recognition to address.

E. Alternative Conceptions of the Mental Lexicon

Each of the models of spoken word recognition described in the previous section shared a number of theoretical assumptions with regard to the processing of the acoustic-phonetic input, the mechanisms used to access the mental lexicon, and the nature of lexical representation. In principle, the activation metaphor of lexical access is sufficient to capture effects due to the encoding of talker-specific information. The problem in accounting for the use of fine perceptual details related to a talker's voice comes when the architecture of the mental lexicon is considered. As noted above, each of the models of lexical access discussed in the preceding section assumed that contact was made with an abstract entry in the mental lexicon. Lexical entries were assumed to contain information only about the linguistic properties of the words. Episodic properties and contextual information were not assumed to be a part of lexical representations. It is this lack of episodic information that prevents traditional models of spoken word recognition from accounting for findings that demonstrate the encoding and use of talker-specific information.

Although researchers in spoken word recognition have traditionally ignored the contribution of episodic information to spoken language processing, several investigations have provided evidence suggesting that both instance-specific information and more general, abstract information are brought to bear on word recognition (Feustel, Shiffrin, & Salasoo, 1983; Salasoo, Shiffrin, & Feustel, 1985). Feustel et al. (1983) investigated repetition effects in word and nonword identification using a visual clarification procedure. They found repetition effects for both words and nonwords. Words also required less information for accurate identification than nonwords. In order to account for the repetition effects, Feustel et al. argued that two types of representations were activated. First, they suggested that subjects encoded instance-specific information during word identification and that this episodic information was reinstantiated during the items repetition. Bringing the episodic information to bear on the repetition of the item facilitated subjects' speed and accuracy in identifying old words and nonwords. Second, Feustel et al. (1983) argued that words have a "unitized" representation in

memory that is also contacted during perceptual identification. The presence of this additional source of information in memory leads to the observed advantage of words over nonwords.

Salasoo et al. (1985) used aspects of the representational scheme described by Feustel et al. (1983) to account for repetition effects in words and nonwords over time. Salasoo et al. found that visual nonwords can become "codified" in memory after several presentations. Once unified representations have been developed for nonwords, enduring repetition effects may be observed. In an identification test given one year after the initial experimental procedure, Salasoo et al. found that old nonwords were identified as accurately as old and new words. All three types of stimuli were identified more accurately than new nonwords. They argued that a unified representation in memory of the old nonwords was responsible of the observed facilitation in identification.

The accounts of repetition effects and lexical representation given by Feustel et al. (1983) and Salasoo et al. (1985) rely on two separate (possibly interacting) types of information. Hintzman (1986) has argued that separate, unified, semantic representations are unnecessary and that the same abstract information can be derived by examining groups of specific instances (see also Jacoby, 1983; Jacoby & Brooks, 1984). Hintzman's (1986) MIVERA 2 is an example of a pure exemplar model of long-term memory (Hintzman, 1988; Hintzman, Grandy, & Gold, 1981; Hintzman & Ludlam, 1980). The model assumes that each perceptual event that subjects attend to lays down a trace in memory. Each trace contains information about the perceptual features of the item and the context in which the item occurred. Separate traces are created for each event such that a previously stored object has several unique instances encoded in memory, rather than a single combined trace of all presentations. Categorization or identification of a new item occurs by comparing the perceptual representation of the new item to all of the stored traces in memory in parallel. The most similar traces in memory become activated by this comparison process and provide support for the identification of the new item. Information about the item and its surrounding context play an extremely important role in the activation of traces in memory and the categorization of new percepts.

The hybrid model described by Feustel et al. and the pure exemplar model proposed by Hintzman are both capable of handling the present results. In both cases, information about a talker's voice can be treated as contextual information that is preserved in an episodic trace. The appeal of the model offered by Feustel et al. is that it is an extension of the traditional notion of lexical representation: It incorporates the idea of an abstract canonical lexical representation with principles of exemplar-based representation. In contrast, MINERVA 2 is much more appealing because it provides a parsimonious explanation for how evidence of the use of abstract representations and the use of instance-specific examples can be incorporated into the same model.

Whether lexical representations consist of two separate stores of information, or just a single, composite store remains an issue for further debate. Another question that is raised by the use of episodic traces during spoken language processing concerns the nature of the information that is encoded in memory. As noted in the introduction, the transfer appropriate processing framework relied heavily on reinstantiating processing operations in accounting for priming effects (Roediger & Srinivas, 1993). Several types of processing operations may be considered as candidates for the information related to a speaker's voice that gets encoded into memory. For example, listeners may store the frequency-sensitive parameters extracted from the speech waveform that are used to drive a normalization mechanism. However, as noted above, the traditional conceptualization of perceptual normalization entails a loss of information and is fundamentally incompatible with the assumption that listeners are storing talker specific information. Storing normalization parameters is also incompatible with the assumption of the transfer appropriate processing framework which suggests that listeners encode perceptual operations.

Another alternative is that listeners store information about the procedures that were used to decode the speech waveform. These procedures may be related to processes that are used to extract acoustic cues from the

speech signal. Alternatively listeners may encode information about the procedures used to recover the articulatory gestures that structured the acoustic signal. As noted above, devising an experimental test to decide between the use of acoustic versus articulatory information is very difficult and is a topic of intense research interest (Kolers, 1976, 1979; Nygaard et al., 1994).

In summary, traditional models of spoken word recognition are unable to account for the use of talker-specific information in speech perception because they provide no means for incorporating episodic information into lexical representations. Examples of a hybrid model and an exemplar model of representation were introduced. These models are promising because they can be integrated with the traditional mechanisms used by models of word recognition. In addition, they are also sufficiently powerful to correct some of the shortcomings of traditional models of lexical access related to assumptions concerning a loss of information.

F. Future Directions

The goal of the present investigation was to examine some of the conditions under which talker-specific information is used to facilitate spoken-language processing. Given that this is a rather broad goal, a number of theoretical and empirical issues have either not been addressed or have been given only cursory examination. These issues include topics such as the auditory characteristics of voices that are encoded into long-term memory and the neural mechanisms that are responsible for processing talker-specific information. In this final section, a few important directions for future research are considered.

In the present investigation, we obtained little evidence to suggest that talker-specific information was used to facilitate spoken language processing in a series of data-driven implicit memory tasks. These results contrast markedly with a number of recent findings that show that tests such as perceptual identification and auditory stem completion tests are sensitive to the surface forms of spoken words (Church & Schacter, 1994; Goldinger, 1992; Schacter & Church, 1992). As discussed above, all of the changes in voice between the study and test conditions of the present experiments were within a gender, whereas other studies have allowed the gender of the voice to change between experimental conditions. The issue of the range of voices included in the stimulus set needs to be considered more thoroughly in order to address the issue of how specific memory representations are for voices.

Another important issue for further investigation concerns the underlying bases for the perceptual similarity among voices. Acoustic measurements of stimulus tokens may provide some insight into the underlying acoustic dimensions that listeners use when determining how similar two voices are. Such findings might be particularly useful in cases such as the present one, in which many obvious factors such as speaking rate, gender and dialect have been controlled for. Measurements might also provide some information about the features that are used to identify voices, such as characteristics related to fundamental frequency, duration and relative formant spacing.

The finding that listeners automatically encode characteristics of the speaker's voice during the lexical decision task suggests that some of the traditional assumptions about the role of a talker normalization mechanism need to be reassessed. In strong terms, the need for such a mechanism can be questioned. In weaker terms, the present data call into question the tacit assumption regarding the loss of information in all theories of perceptual normalization. As noted above, the evidence for a speaker normalization process is somewhat weak and requires further investigation (see also Goldinger, 1992).

Finally, with the advent of accessible neurophysiological techniques, more information is needed about the neural substrates responsible for processing linguistic and nonlinguistic auditory information. Van Lancker and her colleagues (1991; Van Lancker & Canter, 1982; Van Lancker et al., 1985; Van Lancker et al., 1985) have extensively investigated the identification and processing of famous voices in brain damaged subjects.

However, more research is needed on the changes that occur in the regions of the brain that are responsible for processing familiar voices as subjects are taught to identify new voices (Nygaard et al., 1994). Research on brain damaged patients also promises to provide exciting new information about the role of variability in spoken language processing.

G. Summary and Conclusions

The present investigation demonstrated that listeners automatically encode information about a speaker's voice during the course of spoken language processing. Talker-specific details are brought to bear on the recognition of spoken words in a task-sensitive manner. For example, listeners can gain explicit access to this information to facilitate the recognition of spoken words during a recognition memory task. However, under other conditions, such as a lexical decision task, subjects do not appear to engage talker-specific information as a part of the word recognition process. These findings add to a rapidly growing body of research which suggests that spoken language processing has access to and makes use of much more information than has previously been assumed. This body of findings promises to provide an important, and long overdue link, between research on spoken language processing and research on human memory. Both fields will benefit from such cross-talk.

References

Ainsworth, W. (1975). Intrinsic and extrinsic factors in vowel judgments. In G. Fant and M. Tatham (Eds.) *Auditory analysis and perception of speech* (pp. 103-113). London: Academic Press.

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.

Assmann, 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.

Atkinson, R. C. & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, pp. 89-105). New York: Academic Press.

Baddeley, A. D. & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and memory* (Vol. 8, pp. 47-89). New York: Academic Press.

Bricker, P., & Pruzansky, S. (1976). Speaker recognition. In N. J. Lass (Ed.) *Contemporary issues in experimental phonetics* (pp. 295-326). New York: Academic Press.

Brooks, L. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch and B. Lloyd (Eds.) *Cognition and categorization* (pp. 169-211). Hillsdale, N.J.: LEA.

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

Browman, C., & Goldstein, L. (1985). Dynamic modeling of phonetic structure. In V. Fromkin (Ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged* (pp. 35-53). Orlando, FL: Academic Press.

Browman, C., & Goldstein, L. (1986). Towards an articulatory phonology. In C. Ewan & J. Anderson (Eds.), *Phonology yearbook* (vol. 3, pp. 219-254). Cambridge, UK: Cambridge University Press.

Byrd, D. (1992). Sex, dialects, and reduction. *Proceedings of the 1992 International Conference on Spoken Language Processing (Banff, Alberta, Canada)*, **Vol. 1**, 827-830.

Byrd, D. (1993). 54,000 American Stops. *UCLA Working Papers in Phonetics*, **83**, 97-114.

Carterette, E. C., & Barneby, A. (1975). Recognition memory for voices. In E. Cohen and G. Nottebohn (Eds.) *Structure and process in speech perception* (pp. 256-265). New York: Springer.

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

Chomsky, N. & Miller, G. (1963). Introduction to formal analysis of natural languages. In R. D. Luce, R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology (Vol. 2)*. New York: Wiley.

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

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

Clifford, B. R. (1983). Memory for voices: The feasibility and quality of earwitness evidence. In S. M. A. Lloyd-Bostick and B. R. Clifford (Eds.) *Evaluating witness evidence* (pp. 189-218). New York: John Wiley & Sons, Ltd.

Clifford, B. R., Rathborn, H., & Bull, R. (1981). The effects of delay on voice recognition accuracy. *Law and Human Behavior*, **4**, 373-394.

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.

Cole, R. A., & Jakimik, J. (1980). A model of speech perception. In R. A. Cole (Ed.), *Perception and production of fluent speech*. Hillsdale, N.J.: LEA.

Cole, R. A., & Scott, B. (1974). The phantom in the phoneme: Invariant cues for stop consonants. *Perception & Psychophysics*, **15**, 101-107.

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., & Gerstman, L. J. (1952). Some experiments on the perception of synthetic speech sounds. *Journal of the Acoustical Society of America*, **24**, 597-606.

Cotton, S., & Grosjean, F. (1984). The gating paradigm: A comparison of successive and individual presentation formats. *Perception & Psychophysics*, **35**, 41-48.

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

Craik, F. I. M., & Lockhart, R. S. (1974). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, **11**, 671-684.

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

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

Crowder, R. G., & Morton, J. (1969). Precategorical acoustic storage (PAS). *Perception & Psychophysics*, **5**, 365-373.

Cutler, A. (1976). Phoneme monitoring reaction time as a function of preceding intonation contour. *Perception & Psychophysics*, **20**, 55-60.

Cutler, A. & Darwin, C. J. (1981). Phoneme-monitoring reaction time and preceding prosody: Effects of stop closure duration and of fundamental frequency. *Perception & Psychophysics*, **29**, 217-224.

Cutler, A. & Norris, D. G. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception & Performance*, **14**, 113-121.

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

Davison, M. L. (1992). *Multidimensional scaling*. Malabar, FL: Krieger Publishing Co.

DeCasper, A., & Fifer, W. (1980). On human bonding: Newborns prefer their mothers voices. *Science*, **208**, 1174-1176.

Dunn, J. C., & Kirsner, K. (1989). Implicit memory: Task or process. In S. Lewandowsky, J. C. Dunn, and K. Kirsner (Eds.), *Implicit memory: Theoretical issues* (pp. 17-31). Hillsdale, N. J.: Erlbaum.

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

Eimas, P. D. & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, **4**, 99-109.

Ellis, A. (1982). Modality-specific repetition priming of auditory word recognition. *Current Psychological Research*, **2**, 123-128.

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 processes* (pp. 360-385). Hillsdale, N.J.: LEA.

Evans, S. H., & Arnoult, M. D. (1967). Schematic concept formation: Demonstration in a free sorting task. *Psychonomic Science*, **9**, 221-222.

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

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

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

Forbach, G., Stanners, R., & Hochaus, L. (1974). Repetition and practice effects in a lexical decision task. *Memory & Cognition*, **2**, 337-339.

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

Forster, K. I. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper & E. C. T. Walker (Eds.) *Sentence processing: Psycholinguistic studies presented to Merill Garrett.*. Hillsdale, N.J.: LEA.

Forster, K. I. (1987). Form-priming with masked primes: The best-match hypothesis. In M. Coltheart (Ed.), *Attention and Performance XII*. Hillsdale, N.J.: LEA.

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

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

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

Fowler, C. A. (1990). Listener-talker attunements in speech. *Haskins Laboratories Status Report on Speech Research, SR-101/102*, 110-129.

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

Fowler, C. A. & Rosenblum, L. D. (1991). The perception of phonetic gestures. In I. G. Mattingly and M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception* (pp. 39-59). Hillsdale, N. J.: Erlbaum.

Franks, J. J., Plybon, C. J., & Auble, P. M. (1982). Units of episodic memory in perceptual recognition. *Memory & Cognition*, **10**, 62-68.

Gabrieli, J. D. E., Milberg, W., Keane, M. M., & Corkin, S. (1990). Intact priming of patterns despite impaired memory. *Neuropsychologia*, **28**, 417-428.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, **6**, 110-115.

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

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

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

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

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

Gerstman, L. J. (1968). Classification of self-normalized vowels. *IEEE Transactions on Audio Electronics*, **AU-16**, 78-80.

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

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

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

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, **13**, 8-20.

Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 5-16.

Glanzer, M. Adams, J. K., & Iverson, G. (1991). Forgetting and the mirror effect in recognition memory: Concentering of underlying distributions. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 81-93.

Glanzer, M., & Bowles, N. (1976). Analysis of the word-frequency effect in recognition memory. *Journal of Experimental Psychology: Human Learning and Memory*, **2**, 21-31.

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

Goldinger, S. D. (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.

Graf, P. & Schacter, D. L. (1985). Implicit and explicit memory for associations in normal and amnesiac subjects. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **11**, 501-518.

Gregg, V. H. (1976). Word frequency, recognition, and recall. In J. Brown (Ed.), *Recall and recognition.* London: Wiley.

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

Hecker, M. (1971). Speaker recognition: An interpretive survey of the literature. *ASHA monograph*, **16**.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, **93**, 411-423.

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

Hintzman, D. L., Caulton, D. A., & Curran, T. (1992). Retrieval constraints and the mirror effect. Unpublished manuscript.

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

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

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

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

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

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

Jacoby, L. L., Allan, L. G., Collins, J. C., & Larwill, L. K. (1988). Memory influences subjective experience: Noise judgments. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **14**, 240-247.

Jacoby, L. L., & Brooks, L. R. (1984). Nonanalytic cognition: Memory, perception and concept learning. In G. Bower (Ed.), *The psychology of learning and motivation, Vol. 18* (pp. 1-47). New York: Academic Press.

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

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

Jenkins, J. J., Strange, W., & Miranda, S. (1994). Vowel identification in mixed-speaker silent-center syllables. *Journal of the Acoustical Society of America*, **95**, 1030-1043.

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

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

Kelso, J. A. S., Saltzman, E. L., & Tuller, B. (1986). The dynamical theory on speech production: data and theory. *Journal of Phonetics*, **14**, 29-59.

Kewley-Port, D. (1982). Measurement of formant transitions in naturally produced consonant-vowel syllables. *Journal of the Acoustical Society of America*, **72**, 379-389.

Kewley-Port, D. (1983). Time-varying features as correlates of place of articulation in stop consonants. *Journal of the Acoustical Society of America*, **73**, 322-335.

Kewley-Port, D. & Watson, C. S. (1994). Formant-frequency discrimination for isolated English vowels. *Journal of the Acoustical Society of America*, **95**, 485-496.

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

Klatt, D. H. (1980). Speech perception: A model of acoustic-phonetic analysis and lexical access. In R. A. Cole (Ed.), *Perception and production of fluent speech*. Hillsdale, N. J.: LEA.

Klatt, D. H. (1986). The problem of variability in speech recognition and in models of speech perception. In J. S. Perkell and D. H. Klatt (Eds.) *Invariance and variability in speech perception*. Hillsdale, N. J.: LEA.

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

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

Kolers, P. A. (1979). A pattern-analyzing basis for recognition memory. In L. S. Cermack and F. I. M. Craik (Eds.) *Levels of processing and human memory*. Hillsdale, N.J.: LEA.

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

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling*. Beverly Hills, CA: Sage.

Kruskal, J. B., Young, J. B., F. W., & Seery, J. B. (1973). *How to use KYST, a very flexible program to do multidimensional scaling and unfolding*. Murray Hill, N.J.: Unpublished manuscript, Bell Laboratories.

Kucera, F., & Francis, W. (1967). *Computational analysis of present-day American English*. Providence, R. I.: Brown University Press.

Kuhl, P. K. (1992). Psychoacoustics and speech perception: Internal standards, perceptual anchors and prototypes. In L. A. Werner and E. W. Rubel (Eds.) *Developmental psychoacoustics* (pp. 293-332). Washington, DC: APA.

Ladefoged, P. (1967). *Three areas of experimental phonetics*. London: Oxford University Press.

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

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

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

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

Liberman, A. M. (1991). The relation of speech to reading and writing. Unpublished manuscript.

Liberman, A. M., Delattre, P. C., Cooper, F. S., & Gerstman, L. H. (1954). The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs*, **68**, 1-13.

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

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

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

Lieberman, P. (1973). On the evolution of language: A unified view. *Cognition*, **2**, 59-94.

Lightfoot, N. L. (1989). Effect of talker familiarity on serial recall of spoken word lists. *Research on speech perception progress report No. 15*. Bloomington, IN: Speech Research Laboratory, Department of Psychology, Indiana University.

Lisker, L. & Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. *Proceedings of the 6th international conference of phonetic sciences*. Prague: Academia.

Lively, S. E., Pisoni, D. B., & Goldinger, S. D. (1994). Spoken word recognition: Research and theory. In M. Gernsbacher (Ed.) *Handbook of psycholinguistics*. New York: Academic Press.

Lively, S. E., Pisoni, D. B., Summers, V. W., & Bernacki, R. H. (1993). Effects of cognitive workload on speech production: Acoustic analyses and perceptual consequences. *Journal of the Acoustical Society of America*, **93**, 2962-2973.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., & Yamada, T. (in press). Training Japanese listeners to identify English /r/ and /l/: III. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America*.

Lowerre, B. T. (1976). The HARPY speech recognition system. Unpublished doctoral dissertation. Carnegie Mellon University.

Lowerre, B. T., & Reddy, D. R. (1978). The HARPY speech understanding system. In W. A. Lea (Ed.), *Trends in speech recognition*. New York: Prentice Hall.

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

Luce, P. A., Feustel, T. C., & Pisoni, D. B. (1983). Capacity demands in short-term memory for synthetic and natural speech. *Human Factors*, **25**, 17-32.

Luce, R. D. (1959). *Individual choice behavior*. New York: Wiley.

Macmillan, N. A., Kaplan, H. L., & Creelman, C. D. (1977). The psychophysics of categorical perception. *Psychological Review*, **84**, 452-471.

Mandler, G., Goodman, G. O., & Wilkens-Gibbs, D. L. (1982). The word-frequency paradox in recognition memory. *Memory & Cognition*, **10**, 32-42.

Mann, V. A., Diamond, R., & Carey, S. (1979). Development of voice recognition: Parallels with face recognition. *Journal of Experimental Child Psychology*, **27**, 23-40.

Marslen-Wilson, W. D. (1984). Function and process in spoken word recognition. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance: Control of language processes* (pp. 125-150). Hillsdale, N.J.: LEA.

Marslen-Wilson, W. D. (1985). Speed shadowing and speech comprehension. *Speech Communication*, **4**, 55-73.

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

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

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 during word-recognition in continuous speech. *Cognitive Psychology*, **10**, 29-63.

Martin, C. S., Mullennix, J. W., Pisoni, D. B., & Summers, V. W. (1989). Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **15**, 676-684.

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

Massaro, D. W. (1989). Testing between the TRACE model and the fuzzy logical model of speech perception. *Cognitive Psychology*, **21**, 398-421.

Matsumoto, H., Hiki, S., Sone, T., & Nimura, T. (1973). Multidimensional representation of personal quality and its acoustical correlates. *IEEE Transactions of Audio Electroacoustics*, **AU-21**, 428-436.

Mattingly, I. G., & Liberman, A. M. (1969). The speech code and the physiology of language. In K. N. Leibovic (Ed.), *Information processing and in the nervous system*. New York: Springer-Verlag.

Mattingly, I. G., & Liberman, A. M. (1985). Verticality unparalleled. *The Behavioral and Brain Sciences*, **8**, 24-26.

McClelland, J. L. (1991). Stochastic interactive processes and the effect of context on perception. *Cognitive Psychology*, **23**, 1-44.

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

McClelland, J. L. & 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. (1981). An interactive-activation model of context effects in letter perception, Part I: An account of basic findings. *Psychological Review*, **88**, 375-407.

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

McGehee, F. (1944). An experimental investigation in voice recognition. *Journal of General Psychology*, **31**, 53-65.

McKoon, G., & Ratcliff, R. (1979). Priming in episodic and semantic memory. *Journal of Verbal Learning and Verbal Behavior*, **18**, 463-480

McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review*, **92**, 350-371.

McNeill, D. & Lindig, L. (1973). The perceptual reality of phonemes, syllables, words, and sentences. *Journal of Verbal Learning and Verbal Behavior*, **12**, 419-430.

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

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

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

Miller, J. D., Wier, C. C., Pastore, R. E., Kelly, W. J., & Dooling, R. J. (1976). Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception. *Journal of the Acoustical Society of America*, **60**, 410-417.

Miller, J. L. (1987). Mandatory processing in speech perception. In J. L. Garfield (Ed.), *Modularity in knowledge representation and natural-language understanding* (pp. 309-322). Cambridge, MA: MIT Press.

Miller, J. L. & Volaitis, L. E. (1989). Effects of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*, **46**, 505-512.

Miller, R. L. (1953). Auditory tests with synthetic vowels. *Journal of the Acoustical Society of America*, **25**, 114-121.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, **16**, 519-533.

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

Morton, J. (1970). Word recognition. In J. Morton & J. D. Marshall (Eds.), *Psycholinguistics 2: Structure and processes* (pp. 107-156). Cambridge, MA: MIT Press.

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

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

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

Murry, T., & Singh, S. (1980). Multidimensional analysis of male and female voices. *Journal of the Acoustical Society of America*, **68**, 1294-1300.

Muta, H., & Baer, T. (1988). A pitch-synchronous analysis of hoarseness in running speech. *Journal of the Acoustical Society of America*, **52**, 1238-1250.

Nakatani, L. H. & Schaffer, J. A. (1978). Hearing "words" without words: Prosodic cues for word perception. *Journal of the Acoustical Society of America*, **63**, 234-245.

Navon, D. & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review*, **86**, 214-255.

Nearey, T. M. (1978). *Phonetic feature systems for vowels*. Indiana University Linguistics Club, Bloomington, IN.

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

Neisser, U. (1967). Cognitive psychology. Englewood Cliffs, NJ: Prentice-Hall.

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

Nickerson, R. S. (1970). Binary classification reaction time: A review of some studies of human information processing capabilities. *Psychonomic Monograph Supplements*, **4**, 275-318.

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

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **14**, 700-708.

Nusbaum, H. C. & Morin, T. M. (1991). Paying attention to differences among talkers. In Y. Tohkura, E. Vatikiotis-Bateson, and Y. Sagisaka (Eds.) *Speech perception, production, and linguistic structure* (pp. 113-134). Tokyo: OHM Press.

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

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

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

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

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

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. (1961). Parameters of vowel quality. *Journal of Speech and Hearing Research*, **4**, 10-29.

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.

Pisoni, D. B. (1977). Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops. *Journal of the Acoustical Society of America*, **61**, 1352-1361.

Pisoni, D. B., & Lively, S. E. (1994). Variability and invariance in speech perception: A new look at some old problems in perceptual learning. In W. Strange (Ed.) *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research.* Timonium, MD: York Press.

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

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

Rakerd, B. & Verbrugge, R. R. (1987). Evidence that the dynamic information for vowels is talker independent in form. *Journal of Memory and Language*, **26**, 558-563.

Rao, K. V. (1983). The word frequency effect in situational frequency estimation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **9**, 73-81.

Rao, K. V., & Proctor, R. W. (1984). Study-phase processing and the word frequency effect in recognition memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **3**, 386-394.

Remez, R. E. (1987). Neural models of speech perception: A case history. In S. Harnad (Ed.) *Categorical perception: The groundwork of cognition* (pp. 199-225). Cambridge, UK: Cambridge University Press.

Repp, B. (1982). Phonetic trading relations and context effects: New evidence for a phonetic mode of perception. *Psychological Bulletin*, **92**, 81-110.

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

Roediger, H. L. (1990). Implicit memory: Retention without remembering. *American Psychologist*, **45**, 1043-1056.

Roediger, H. L., & Blaxton, T. A. (1987a). Retrieval modes produce dissociations in memory for surface information. In D. Gorfein and R. R. Hoffman (Eds.), *Memory and learning: The Ebbinghaus centennial conference*. Hillsdale, N. J.: Erlbaum.

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

Roediger, H. L., & McDermott, K. B. (1993). Implicit memory in normal human subjects. In F. Boller and J. Grafman (Eds.) *Handbook of neuropsychology, Vol. 8* (pp. 63-131). Amsterdam: Elsevier.

Roediger, H. L., & Srinivas, K. (1993). Specificity of operations in perceptual priming. In P. Graf and M. E. J. Masson (Eds.), *Implicit memory: New directions in cognition, development, and neuropsychology* (pp. 17-48). Hillsdale, N. J.: Erlbaum.

Roediger, H. L., Weldon, M. S., & Challis, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: A processing account. In H. L. Roediger and F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour or Endel Tulving*. Hillsdale, N. J.: Erlbaum.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive model of context effects in letter perception, Part 2: The perceptual enhancement effect and some tests and extensions of the model, *Psychological Review*, **89**, 60-94.

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.

Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, **110**, 474-494.

Sander, E. K., & Ripich, D. E. (1983). Vocal fatigue. *Annals of Otol Rhinol Lanyngology*, **92**, 141-145.

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

Savin, H. B., & Bever, T. G. (1970). The nonperceptual reality of the phoneme. *Journal of Verbal Learning and Verbal Behavior*, **9**, 295-302.

Sawusch, J. R., Nusbaum, H. C., & Schwab, E. C. (1980). Contextual effects in vowel perception II: Evidence for two processing mechanisms. *Perception & Psychophysics*, **27**, 421-434.

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

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

Schacter, D. L. & 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.

Scherer, R. C., Titze, I. R., Raphael, B. N., Wood, R. P., Ramig, L. A., & Blager, R. F. (1987). Vocal fatigue in a trained and untrained voice user. In T. Baer and L. Sasaki (Eds.) *Laryngeal function in phonation and respiration*. Boston: Little, Brown, & Co.

Shankweiler, D., Strange, W., & Verbrugge, R. (1977). Speech and the problem of perceptual constancy. In R. E. Shaw and J. Bransford (Eds.) *Perceiving, acting, and comprehending: Toward an ecological psychology* (pp. 315-345). Hillsdale, N.J.: Erlbaum.

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

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

Shimamura, A. P. (1986). Priming effects in amnesia: Evidence for a dissociable memory function. *Quarterly Journal of Experimental Psychology*, **38A**, 619-644.

Singh, S., & Murry, T. (1978). Multidimensional classification of normal voice qualities. *Journal of the Acoustical Society of America*, **64**, 81-87.

Skinner, B. F. (1957). *Verbal behavior*. New York, NY: Appleton-Century-Crofts.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, **4**, 592-604.

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

Stevens, K. N. & Blumstein, S. E. (1978) Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America*, 64, 1358-1368.

Stevens, K. N. & Blumstein, S. E. (1981). The search for invariant acoustic correlates of phonetic features. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech* (pp. 1-38). Hillsdale, N.J.: LEA.

Stevens, K. N. & Halle, M.(1967). Remarks on analysis by synthesis and distinctive features. In W. Wathen-Dunn (Ed.), *Models for the perception of speech and visual form*. Cambridge: MIT Press.

Strange, W. (1987). Information for vowels in formant transitions. *Journal of Memory and Language*, **26**, 550-557.

Strange, W., Verbrugge, R. R., Shankweiler, D. P., & Edman, T. R. (1976). Consonant environment specifies vowel identity. *Journal of the Acoustical Society of America*, **60**, 213-224.

Streeter, L. A., MacDonald, N. H., Apple, W., R. M, & Galotti, K. M. (1983). Acoustic and perceptual indicators of emotional stress. *Journal of the Acoustical Society of America*, **73**, 1354-1360.

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

Studdert-Kennedy, M. (1980). Speech perception. *Language and Speech*, **23**, 45-66.

Studdert-Kennedy, M. (1983). Perceiving phonetic events. *Haskins Laboratories Status Report on Speech Research, SR-74/75*. New Haven, CT: Haskins Laboratories, 53-69.

Summerfield, Q. (1975). Acoustic and phonetic components of the influence of voice changes on identification times for CVC syllables. *Speech Perception*, **2**, 73-98. Belfast: The Queen's University of Belfast.

Summerfield, Q. & Haggard, M. P. (1973). Vocal tract normalization as demonstrated by reaction times. *Speech Perception,* **3**, 75-86. Belfast: The Queen's University of Belfast.

Syrdal, A. K. (1984). Aspects of a model of the auditory representation of American English vowels. *Speech Communication*, **4**, 131-135.

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

Summerfield, A. Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 1074-1095.

Thompson, C. P. (1987). A language effect in voice identification. *Applied Cognitive Psychology*, **1**, 121-131.

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

Tulving, E. (1983). Elements of episodic memory. New York: Oxford University Press.

Tulving, E., & Thompson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, **80**, 352-373.

Van Lancker, D. (1991). Personal relevance and the human right hemisphere. *Brain and Cognition*, **17**, 64-92.

Van Lancker, D., & Canter, G. L. (1982). Impairment of voice and face recognition in patients with hemispheric damage. *Brain and Cognition*, **1**, 185-195.

Van Lancker, D. R., Cummings, J. L., Kreiman, J., & Dobkin, B. H. (1988). Phonagnosia: A dissociation between familiar and unfamiliar voices: *Cortex,* **24**, 195-209.

Van Lancker, D., & Kreiman, J. (1987). Voice discrimination and recognition are separate abilities. *Neuropsychologia*, **25**, 829-834.

Van Lancker, D., Kreiman, J., & Cummings, J. (1989). Voice perception deficits: Neuroanatomical correlates of phonagnosia. *Journal of Clinical and Experimental Neuropsychology*, **11**, 665-674.

Van Lancker, D., Kreiman, J., & Emmorey, K. (1985). Familiar voice recognition: Parameters and patterns. Part I. Backwards voices. *Journal of Phonetics*, **13**, 19-38.

Van Lancker, D., Kreiman, J., & Wickens, T. (1985). Familiar voice recognition: Parameters and patterns. Part II. Rate-altered voices. *Journal of Phonetics*, **13**, 39-52.

Verbrugge, R. R., & Rakerd, B. (1986). Evidence of talker-independent information for vowels. *Language and Speech*, **29**, 39-57.

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

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

Voiers, W. D. (1965). Performance evaluation of speech processing devices II.: The role of individual differences. *Report No. AFCRL-66-24*, Air Force Cambridge Research Laboratories, Office of Aerospace Research, Bedford, MA.

Volaitis, L. E. & Miller, J. L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America*, **92**, 723-735.

Walden, B. E., Montgomery, A. A., Gibeily, G. J., Prosek, R. A., & Schwartz, D. M. (1978). Correlates of psychological dimensions in talker similarity. *Journal of Speech and Hearing Research*, **21**, 265-275.

Waugh, N. C., & Norman, D. A. (1965). Primary memory. *Psychological Review*, **72**, 89-104.

Weiner, F., & Singh, S. (1974). Multidimensional analysis of voice reaction time judgments on pairs of English fricatives. *Journal of Experimental Psychology*, **102**, 613-620.

Wicklegren, W. A. (1969). Context-sensitive coding, associative memory, and serial order in (speech) behavior. *Psychological Review*, **76**, 1-15.

Yanigahara, N. (1967). Significance of harmonic changes and noise components in hoarseness. *Journal of Speech and Hearing Research*, **10**, 531-541.

Zatorre, R.J., Evans, A.C., Meyer, E. and Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science*, **256**, 846-849.

Zechmeister, E. B., Curt, C., & Sebastian, J. A. (1978). Errors in a recognition memory task are U-shaped functions of word frequency. *Bulletin of the Psychonomic Society*, **11**, 371-373.

Appendix A

Words and Nonwords used in Multidimensional Scaling of Voices

High Frequency Words	Low Frequency Words	Nonwords	
wide	surf	kime	pook
folk	cad	sadge	noop
cause	hoof	moun	herm
ship	vogue	vem	wule
cell	lure	nong	gat
coach	rouge	cham	wob
shook	peach	derd	jick
date	shine	yid	nedge
pace	coil	nug	looth
load	posh	gesh	chobe
pen	yawn	bape	souch
leg	czar	dom	titch
pull	kale	piff	vour
pair	soot	yock	shog
young	chalk	kouse	yach
deep	chop	sep	div
bit	goat	yar	kib
path	fetch	doudge	shoud
gear	moss	thup	meck
chief	gull	tade	lipe
cut	worm	tibe	nerb
phone	bike	mafe	ched
learn	mash	sib	jile
mud	limb	vap	zet
rush	mole	koof	losh
bag	gig	zin	kutch
ridge	toot	birge	toshe

Words and Nonwords used in Experiments 1A, 1B, 1C, and Chapter 4

Words

bag, bill, bus, cause, beach, bit, rang, cell, bike, birch, coot, chalk, boon, burp, cad, coil, get, chief, cut, date, folk, coach, choose, death, deep, doze, fang, gaff, gape, chop, fetch, geese, gig, join, lack, learn, mud, phone, gear, judge, leave, goat, gull, limb, mash, girth, hoof, kale, lure, pace, pair, path, ridge, leg, load, knife, pen, mole, pal, peach, rouge, moss, gnome, posh, pub, roof, share, ship, teeth, pope, pull, rush, shirt, sane, sham, soot, tomb, sire, shine, surf, tab, vain, wage, wide, yes, shook, thin, wish, young, vogue, worm, wove, zing, toot, wraith, yawn, czar

Nonwords

bape, ched, derd, doudge, cham, chobe, div, dom, herm, jile, kime, gesh, jick, kouse, kutch, lipe, mafe, meck, nedge, losh, moun, nong, noop, nerb, piff, sadge, sep, nug, pook, sib, souch, shog, tade, tibe, vap, thup, titch, toshe, vem, vour, wob, yach, yar, wule, yid, yock, zet

Words used in Experiments 3A, 3B, and 3C

bag, bill, bus, cause, beach, bit, rang, cell, bike, birch, coot, chalk, boon, burp, cad, coil, get, chief, cut, date, folk, coach, choose, death, deep, doze, fang, gaff, gape, chop, fetch, geese, gig, join, lack, learn, mud, phone, gear, judge, leave, goat, gull, limb, mash, girth, hoof, kale, lure, pace, pair, path, ridge, leg, load, knife, pen, mole, pal, peach, rouge, moss, gnome, posh, pub, roof, share, ship, teeth, pope, pull, rush, shirt, sane, sham, soot, tomb, sire, shine, surf, tab, vain, wage, wide, yes, shook, thin, wish, young, vogue, worm, wove, zing, toot, wraith, yawn, czar

Appendix B

Tables of Means

Experiment 1A: Lexical Decision Test
Single Talker Condition
High Frequency Words

	Old	Items	New	Items
Talker	RT	Error	RT	Error
Exact	832	1	896	7
Familiar	843	1	895	3
Similar	885	1	983	3
Dissimilar	905	4	976	15

Low Frequency Words

	Old	Items	New	Items
Talker	RT	Error	RT	Error
Exact	870	8	999	28
Familiar	868	15	1020	24
Similar	918	14	993	29
Dissimilar	922	12	1056	37

Nonwords

	Old	Items	New	Items
Talker	RT	Error	RT	Error
Exact	1047	19	1026	19
Familiar	1088	16	1063	18
Similar	1078	17	1005	16
Dissimilar	1074	20	1129	20

Experiment 1A: Lexical Decision Test
Multiple Talker Condition
High Frequency Words

	Old	Items	New	Items
Talker	RT	Error	RT	Error
Exact	884	2	989	9
Familiar	861	2	920	8
Similar	910	4	975	8
Dissimilar	929	9	1009	19

Low Frequency Words

	Old	Items	New	Items
Talker	RT	Error	RT	Error
Exact	925	13	1022	30
Familiar	900	10	1027	35
Similar	890	10	1080	29
Dissimilar	906	14	1083	33

Nonwords

	Old	Items	New	Items
Talker	RT	Error	RT	Error
Exact	1092	25	1083	24
Familiar	1105	22	1117	22
Similar	1097	16	1137	20
Dissimilar	1173	24	1127	25

Experiment 1B: Recognition Memory
Single Talker Condition
High Frequency Words

Talker	Percent Hits	Hit RT	Percent False Alarms
Exact	92	910	-
Familiar	94	934	17
Similar	90	939	16
Dissimilar	85	1033	15

Low Frequency Words

Talker	Percent Hits	Hit RT	Percent False Alarms
Exact	95	938	-
Familiar	95	910	16
Similar	92	925	13
Dissimilar	90	989	11

Nonwords

Talker	Percent Hits	Hit RT	Percent False Alarms
Exact	82	1046	-
Familiar	80	1068	16
Similar	76	1109	12
Dissimilar	71	1150	12

Experiment 1B: Recognition Memory
Multiple Talker Condition
High Frequency Words

Talker	Percent Hits	Hit RT	Percent False Alarms
Exact	91	904	-
Familiar	87	915	13
Similar	92	921	13
Dissimilar	89	947	11

Low Frequency Words

Talker	Percent Hits	Hit RT	Percent False Alarms
Exact	89	894	-
Familiar	96	882	15
Similar	95	872	12
Dissimilar	92	922	13

Nonwords

Talker	Percent Hits	Hit RT	Percent False Alarms
Exact	81	1032	-
Familiar	81	1041	17
Similar	77	1084	16
Dissimilar	80	1113	15

Experiment 1C: Lexical Decision Test
High Frequency Words

	Old	Items	New	Items
Talker	RT	Error	RT	Error
Old	866	1	942	5
Similar	846	4	935	5
Dissimilar	866	3	947	12

Low Frequency Words

	Old	Items	New	Items
Talker	RT	Error	RT	Error
Old	853	8	1012	28
Similar	860	11	1018	29
Dissimilar	937	11	1016	31

Nonwords

	Old	Items	New	Items
Talker	RT	Error	RT	Error
Old	1056	12	1058	14
Similar	1022	12	1078	20
Dissimilar	1117	17	1118	23

Recognition Memory Test
High Frequency Words

Talker	Percent Hits	Hit RT	Percent False Alarms
Old	77	825	10
Similar	72	838	9
Dissimilar	75	830	10

Low Frequency Words

Talker	Percent Hits	Hit RT	Percent False Alarms
Old	79	824	11
Similar	77	823	13
Dissimilar	77	803	12

Nonwords

Talker	Percent Hits	Hit RT	Percent False Alarms
Old	68	944	15
Similar	61	928	13
Dissimilar	58	996	18

Retention of Voices over Time

Lexical Decision Test
High Frequency Words

	Old	Items	New	Items
Talker	RT	Error	RT	Error
Old	818	2	833	2
Similar	795	1	815	4
Dissimilar	803	7	870	8

Low Frequency Words

	Old	Items	New	Items
Talker	RT	Error	RT	Error
Old	857	8	904	22
Similar	817	13	910	31
Dissimilar	876	15	927	25

Nonwords

	Old	Items	New	Items
Talker	RT	Error	RT	Error
Old	953	12	991	15
Similar	950	13	1012	16
Dissimilar	1036	16	1039	12

Recognition Memory Test
High Frequency Words

Talker	Percent Hits	Hit RT	Percent False Alarms
Old	89	1048	13
Similar	77	1127	12
Dissimilar	67	1097	15

Low Frequency Words

Talker	Percent Hits	Hit RT	Percent False Alarms
Old	94	970	23
Similar	85	990	12
Dissimilar	84	1031	18

Nonwords

Talker	Percent Hits	Hit RT	Percent False Alarms
Old	82	1070	16
Similar	79	1064	14
Dissimilar	69	1112	12

Gating I
Average Duration for Correct Completion

High Frequency Words

Talker	Old Items	New Items
Old	263	307
Similar	280	294
Dissimilar	268	294

Low Frequency Words

Talker	Old Items	New Items
Old	267	302
Similar	246	299
Dissimilar	260	277

Gating II
Average Duration for Correct Completion

High Frequency Words

Talker	Old Items	New Items
Old	242	315
Similar	246	298
Dissimilar	239	319

Low Frequency Words

Talker	Old Items	New Items
Old	195	305
Similar	218	281
Dissimilar	222	301

Auditory Stem Completion
Percent Correct

High Frequency Words

Talker	Old Items	New Items
Old	47	24
Similar	60	32
Dissimilar	48	23

Low Frequency Words

Talker	Old Items	New Items
Old	59	21
Similar	54	12
Dissimilar	51	15