

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

**Some Comments on Implicit Effects of Voices
on Spoken Word Recognition¹**

Carl W. Turner and David B. Pisoni

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

¹This research was supported by NIH Research Grant DC-00111 and NIDCD Training Grant DC-00012 to Indiana University in Bloomington, IN. We thank Sonya Sheffert for her helpful comments and suggestions on this manuscript.

Abstract

Church and Schacter (1994) reported the results of an experiment on speech perception which examined the effects of changing a speaker's voice between study and test on word identification and word recognition memory. They found that subjects scored a higher percentage of spoken word identifications on same-voice trials than different-voice trials; there was no significant difference due to study-to-test changes in the speaker's voice in a subsequent test of word recognition. These data were offered in support of a hypothetical auditory memory system (the PRS) that operates on implicit perceptual tests but has no effect on explicit recognition memory. In the present study, two experiments were conducted that revealed no difference in accuracy due to voice characteristics on a perceptual task but a reliable difference on a recognition memory task, demonstrating a dissociation in the opposite direction of that reported by Church and Schacter. We discuss the reasons for this set of apparently contradictory results, and argue that (1) voice information is incidentally encoded during on-going speech perception and spoken word recognition and that (2) voice information operates implicitly on both perceptual identification and recognition memory depending on the constraints of the situation. The findings reported in these experiments can not presently be accounted for by the PRS hypothesis.

Some Comments on Implicit Effects of Voices on Spoken Word Recognition

In a recent study of implicit memory and auditory priming, Church and Schacter (1994) showed that manipulating the surface characteristics of spoken words between study and test impaired performance in word identification and stem completion tasks. Changing variables such as speaker's voice, intonation, and fundamental frequency between the initial presentation of a list of spoken words and tests of perceptual identification impaired subjects' performance on these tests. Furthermore, subsequent recognition of the studied words was usually unaffected by these variables. Their study replicated the major findings of earlier studies by Schacter and colleagues which have demonstrated dissociations between perceptual identification tests and explicit memory tests (Schacter & Church, 1992; Schacter, Church, & Osowiecki, 1994). However, the data from the recognition memory tests reported in these experiments are inconsistent with several others reported in the literature with regard to the effects of speakers' voices on auditory recognition memory (Goldinger, 1992; Palmeri, Goldinger, & Pisoni, 1993; Sheffert & Fowler, in press).

The results reported by Church and Schacter (1994) are compatible with other findings in the memory literature that deal with dissociation between perceptual tasks and word recognition and the distinction between implicit and explicit memory (for comprehensive reviews, see Roediger and McDermott (1993) and Schacter (1987)). The general finding that has come from much of the recent work on implicit memory is that manipulations of the surface features of words have large effects on perceptual memory tests but have no little or no effect on explicit tests such as recall or recognition (see also Goldinger, 1992). Although most of the work on implicit memory for words has been in the visual domain, an increasing number of studies have attempted to demonstrate dissociations between implicit and explicit tests in the auditory domain.

The studies by Schacter and others that reveal priming on perceptual tests based on voice characteristics are important because they show that studies of visual word perception may generalize to the auditory domain. Although there is still relatively little work being done on implicit auditory memory, the findings are consistent across modalities and are observed both in subjects with unimpaired memory and among amnesic patients (Schacter, Church, & Bolton, in press; Schacter, Church, & Treadwell, 1994). The data have been interpreted by Schacter as support for a theoretical account of memory that entails separate physiological systems for implicit memory -- the perceptual representational systems (PRS) -- and explicit memory (Schacter, 1990, 1992; Squire, 1987; Tulving & Schacter, 1990). These systems theories of memory stand as alternatives to the so-called "processing" accounts of memory which ascribe successful identification and recognition to a match between processes carried out during encoding and retrieval (Roediger, 1990).

Consistent with the hypothesis that voice manipulations of spoken words do not affect explicit tests, Church and Schacter (1994) failed to find priming on recognition or on cued recall (see also Schacter & Church, 1992). However, these non-effects of voice on explicit tests stand in direct contrast with several recent experiments carried out in our laboratory and elsewhere in which auditory word recognition is facilitated when spoken words are presented in the same voice during both study and test (Goldinger, 1992; Lively, 1994; Palmeri et al., 1993; Sheffert & Fowler, in press). The apparent conflict between two sets of data has important implications for the PRS hypothesis and for current accounts of speech perception and spoken language processing.

In one recent study on the effects of voice on recognition memory, Palmeri et al. (1993) tested listeners' memory for spoken words in a "continuous recognition memory" task (Craik & Kirsner, 1974; Sheppard & Teghtsoonian, 1961). Words spoken by multiple talkers were presented twice for recognition within a single list; the number of intervening items between the first and second presentation of a word varied from one to 64. Subjects were more accurate at recognizing words presented in the same voice than words presented in a different voice; this overall advantage extended even to trials with 64 intervening items between initial word presentation and target. It is important to point out that no mention was made to subjects about the voice manipulation; they were instructed simply to make judgments about the words themselves. In a second experiment, subjects were instructed to attend to the speakers' voices. In this condition, subjects demonstrated above-chance performance in their ability to explicitly judge if the target words were spoken in the same voice or in a different voice. The results of both experiments suggest that the episodic memory of the speaker's voice is not lost but is preserved in long-term memory.

In another study, Goldinger (1992) compared perceptual and recognition tests using spoken words embedded in noise. In one condition, subjects studied lists of words and were then presented with the words for identification in the same voice or a different voice at delays of five minutes, one day, and one week. Across all delays, words presented in the same voice during test were more easily identified than words presented in a different voice. In a second condition, subjects listened to words presented in the clear, and were then tested for explicit recognition of the words at the same three delays. Subjects were better able to recognize old words presented in the same voice than in different voices after a 5-minute delay; there were no consistent differences due to voice at the 1-week delay. The data from the perceptual tests were consistent with those reported by Schacter and Church (1992) with respect to implicit memory for spoken words. The results also fit into the general framework of implicit memory whereby explicit memory is more affected by time than implicit memory, and are consistent with earlier studies on memory for visual words by Tulving, Schacter, and Stark (1982) and Masson (1984). In contrast to the findings reported by Church and Schacter (1994) and Schacter and Church (1992), however, Goldinger also found a significant effect of voice on recognition performance. Furthermore, the advantage for recognition memory during same-voice trials at a 5-minute delay held for speaker ensembles of two, six, and ten voices. Goldinger concluded from these results that detailed perceptual information about the speakers' voices was available during both word perception and word recognition and that this information is encoded in memory along with a more abstract representation of the lexical form of a word.

Sheffert and Fowler (in press) have since replicated the findings of Palmeri et al. (1993) using audio-visual stimuli instead of pure auditory stimuli. They found significant effects of the speaker's voice on word recognition; the effects of preserving the visible speaker across study and test were weak or entirely absent. They also found that explicit memory for faces was strong whereas explicit memory for voices was poor. The authors concluded from these data that detailed voice information is stored in memory along with the episodic memory for the spoken words, but that voice and visual face information is preserved independently. Together with the earlier findings of Craik and Kirsner (1984), these studies show that study-to-test voice matching can improve performance of auditory word recognition.

These experiments clearly demonstrate that voice information is not lost during initial encoding, but is preserved in memory and may, under some conditions, facilitate both perception and recognition (Pisoni, 1993). Church and Schacter's failure to find voice effects on explicit tests is therefore puzzling and warrants an explanation. How can we reconcile these apparent failures with the growing literature on the positive effects of voice on recognition memory? Within the theoretical orientation of PRS, recognition is considered to be an explicit test of memory; thus, voice characteristics should not affect performance on this test. But, as Goldinger (1992), Palmeri et al. (1993), Lively (1994) and Sheffert and Fowler (in press)

have all shown, voice manipulations do show a substantial and reliable effect on auditory recognition memory.

This article is not meant to dispute Church and Schacter's (1994) positive findings on implicit priming of voice information on word identification; the positive effects of voice matching on perceptual tasks are now well-documented. Rather, we take issue with their interpretation of some prominent null results, namely: their report of the lack of effect of voice on recognition and cued recall in their own experiments. Although Church and Schacter concede that "there are some circumstances in which variations in voice information reduce explicit recall and recognition (p. 532)," they do little to explicate those conditions or account for the earlier findings of Palmeri et al. (1993) and Goldinger (1992). In general, the PRS theory appears not to be able to account for these data. In discussing the outcome of their Experiment 1 and their failure to find an effect of voice on the recognition memory test, Church and Schacter (1994) commented, "This dissociation between implicit and explicit memory is consistent with the hypothesis that auditory priming is mediated by a PRS that stores information about the acoustical forms of the words, whereas explicit memory is mediated by an episodic system that generally relies on conceptually driven processes (p. 524)." In fact, the effect of voice information on word recognition is robust and important in understanding the nature of implicit and explicit memory for spoken words (Goldinger, 1992; Lively, 1994; Palmeri et al., 1993; Sheffert & Fowler, in press).

The present experiments were designed to follow up on and extend several earlier studies on voice and priming on perceptual and recognition tests of spoken words. The design included an implicit and explicit component (identification and recognition) that is characteristic of studies designed to uncover memory dissociations. Our goal was to replicate and extend several previous studies in which voice manipulations were found on tests of recognition memory (e.g., Palmeri et al., 1993). The results are discussed in terms of Church and Schacter's failure to find voice effects in explicit tests, and the conditions under which implicit priming of auditory recognition memory may or may not occur.

Experiment 1

If listeners incidentally encode the voice characteristics of the speaker during word perception, then such information may facilitate later perception of words presented by the same speaker. This experiment was intended, in part, to qualify the common finding that words repeatedly presented in the same voice are identified more accurately than words presented in different voices (Church & Schacter, 1994; Goldinger, 1992; Mullenix, Pisoni, & Martin, 1989; Schacter & Church, 1992).

Our design was different from previous studies that studied identification: Here we introduced a novel procedure for collecting identification data. Typically, words are presented twice; once during a study phase and then again during a test phase for open-set identification.² In Experiment 1, subjects were repeatedly presented with lists of words for identification. Subjects heard a word on each trial, then pressed a button corresponding to one of four that was visually presented on a computer screen in a four-alternative forced-choice design (4AFC). With this "continuous identification" design it was possible to track improvement both in accuracy and in reaction times across the several presentations of the same word lists. The identification task required only a shallow level of processing and no semantic encoding, conditions which have been shown in the past to be favorable for producing implicit effects (Roediger & McDermott,

²Church and Schacter's Experiment 1 used three presentations a stimulus word: two tokens presented in the clear and one of a degraded token. Subjects were instructed to ignore the second, degraded, stimulus presentation during the recognition test. Their inability to do so completely may have accounted for the relatively high false alarm rate (compared with our data in Experiment 2) in their experiment.

1993). The word lists used high frequency, single syllable words and had no obvious organizational component. Subjects were not instructed to remember the word list, nor were they instructed to attend to the speakers' voices. If there is memory for the words then that should be shown by increased accuracy and decreased response latency across six repetitions of the word lists.

An additional question has been raised about the role of noise in priming of spoken words. Church and Schacter (1994) refer to early studies by Schacter and Church (1992) and Jackson and Morton (1984) as two experiments in which no effects of same-voice priming were found using noisy (i.e., degraded) stimuli in tests of explicit memory. In order to investigate the effects of noise on identification and recognition, both of our experiments employed a novel signal processing technique to degrade stimuli; performance on a high-noise condition and a low-noise condition were compared to performance on a no-noise condition. Introducing noise to auditory perceptual tests has the predictable effect of reducing accuracy and slowing response times on correct trials (Mullennix et al., 1989).

It should be noted that the perceptual test used here may be thought of as a classification task, not a pure test of identification. Some of the effects often found in perceptual tests may be masked by a closed-set, forced-choice design (Sommers, Nygaard, & Pisoni, 1992). The classification test produces higher performance than a pure open-set identification test; there are fewer processing demands and therefore, less effort needed to complete the task. Thus, the effects of stimulus variability may be substantially attenuated. Thus, the identification task was constrained in two ways. If constraining the perceptual task to a 4AFC design provides cues to subjects otherwise unavailable, then the specific effects of voice may be masked by these two factors. Based on this we expected to find no difference in identification scores between same- and different-voice words.

Method

Subjects. Fifty-six subjects participated in this experiment: 18 were assigned to a no-noise condition, 17 to the low-noise condition, and 21 to the high-noise condition. All were recruited from the student population at Indiana University and received course credit for their participation. Subjects were tested in groups of six or fewer. All were native speakers of English who reported no speech or hearing disorders at the time of testing.

Stimuli. The words used in these experiments were single syllable, high frequency words drawn from the Modified Rhyme Test (MRT; House, Williams, Hecker, & Kryter, 1965). Each word was recorded on audiotape and digitized using a 12-bit analog-to-digital converter. The RMS amplitude levels for all words were digitally equated. All words were spoken by six different speakers (three male and three female).

The 96 words selected from the MRT list consisted of 48 word groups of two words each. Twenty-four word groups varied on the first phoneme, 24 on the last. On a given trial, subjects viewed four words in a word group printed in capital letters on the bottom line of a computer screen, and simultaneously heard one of the words presented through headphones. Thus, the words "BAD," "BACK," "BASS," and "BAN" might appear on the CRT screen as subjects heard the word "bad" (refer to the Appendix for the complete list of stimulus materials).

Design. Subjects were randomly assigned to one of three groups: No-noise, low-noise, or high-noise. The stimuli used in the two noise conditions were created by re-recording the digitized word files and flipping the sign bits in the files by a predetermined probability. Low-noise stimuli were bit-flipped

with a .05 probability; high-noise stimuli were bit-flipped with a .15 probability. This technique of stimulus degradation differs from white noise because it preserves the spectral envelope of the speakers' voices.

Using one word from each of the 48 selected MRT word groups, three 16-word lists were constructed. In a trial block, a 16-word list was presented six times for a total of 96 trials per block. A two-minute break preceded the next block of trials. On every trial, three alternatives were selected randomly from the remaining five words for presentation on the screen. Within a trial block, half the words were spoken by the same speaker on six successive list presentations, and half were spoken one time by each of the six speakers. Speaker order within a trial block was randomized. Finally, the words were arranged for display on the computer screen so that the target word appeared in a random location.

Procedure. Subjects were tested in sound-attenuated booths equipped with a computer screen, headphones, and a four-button response box. Buttons on the response box were labeled "1" to "4" from left to right. Subjects were instructed to respond as quickly as possible to the spoken word by pressing one of the four buttons corresponding to the words on the screen. After each trial, a light above the correct button was lit to indicate the correct response. The deadline for responding was two seconds, after which the screen went blank and the next trial began. The time between trials was 1500 msec.

Results

There was no difference in percentage of correct classifications based on same- or different-voice presentations. An ANOVA performed on accuracy scores revealed no effect of Voice, $F(1, 53) = 0.13$. The only effects in accuracy were due to Noise ($F(2, 53) = 23.9, p < .0001$) and Repetition ($F(5, 265) = 67.1, p < .0001$). Accuracy increased across the six list repetitions and decreased as the level of noise increased. As shown in Figure 1, subjects in the same-voice and different-voice no-noise conditions performed at ceiling throughout the six repetitions of the word lists. By contrast, subjects in the high-noise conditions scored a mean of 62.2% correct in the first repetition and reached asymptote only by the fourth repetition. There was also also a Noise x Repetition interaction, $F(10, 265) = 17.5, p < .0001$.

Insert Figure 1 about here

Response latency, however, showed a strong effect of the voice manipulations. Subjects were faster to correctly classify same-voice words than different-voice words, $F(1, 53) = 10.1, p < .01$. The shorter latencies on the correct same-voice trials may have reflected the listeners' greater confidence in their judgments. There was a main effect of Noise on latency, $F(2, 53) = 5.0, p < .01$: Subjects in the no-noise conditions responded fastest and subjects in the high-noise conditions responded the slowest. There was also an effect of Repetition, $F(5, 265) = 68.1, p < .0001$, as the latencies decreased for all subjects across the six word list repetitions. And, there was a Repetition x Noise interaction, $F(10, 265) = 2.1, p < .05$; see Figure 2.

Insert Figure 2 about here

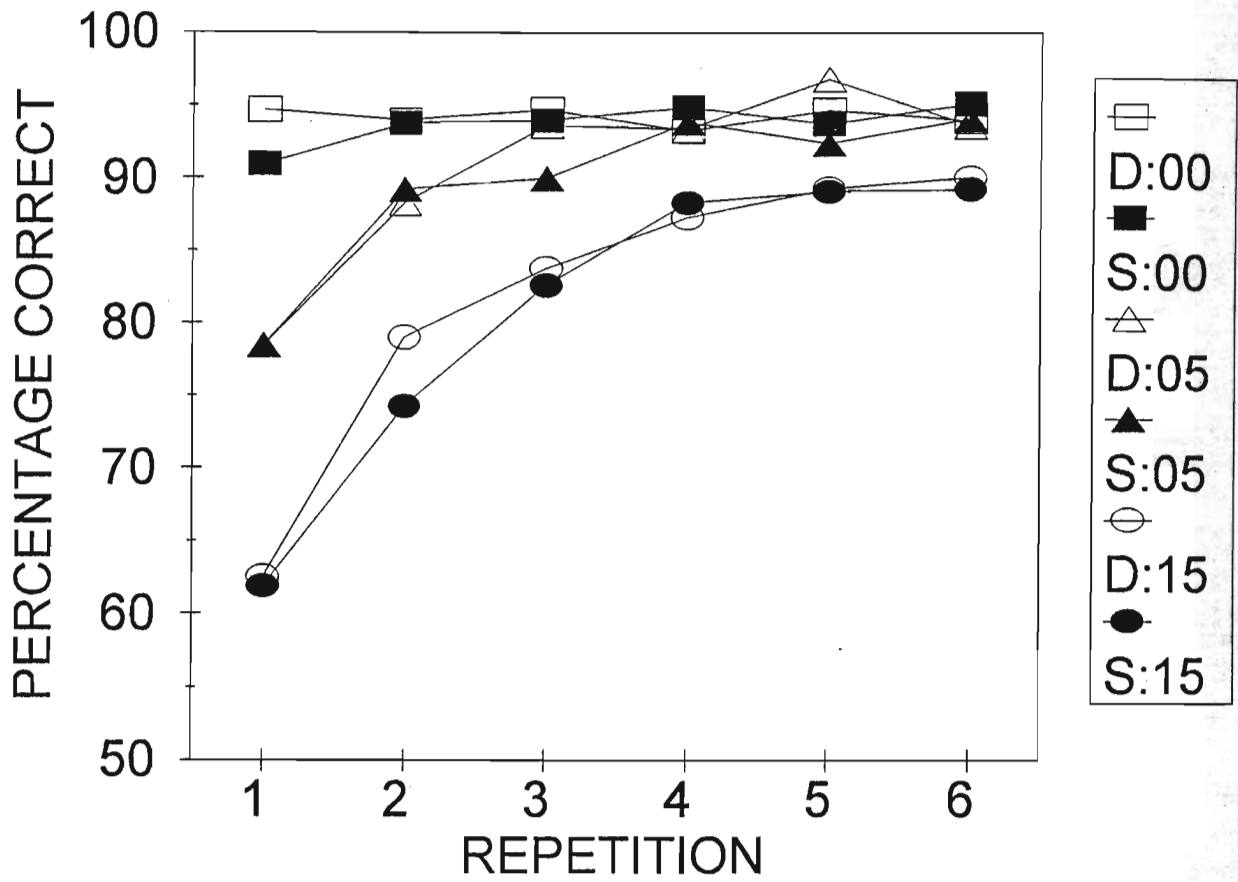


Figure 1. Percentage of correct word identifications in Experiment 1 by level of noise (00 = no noise, 05 = low noise, 15 = high noise), repetition, and voice (S = same, D = different).

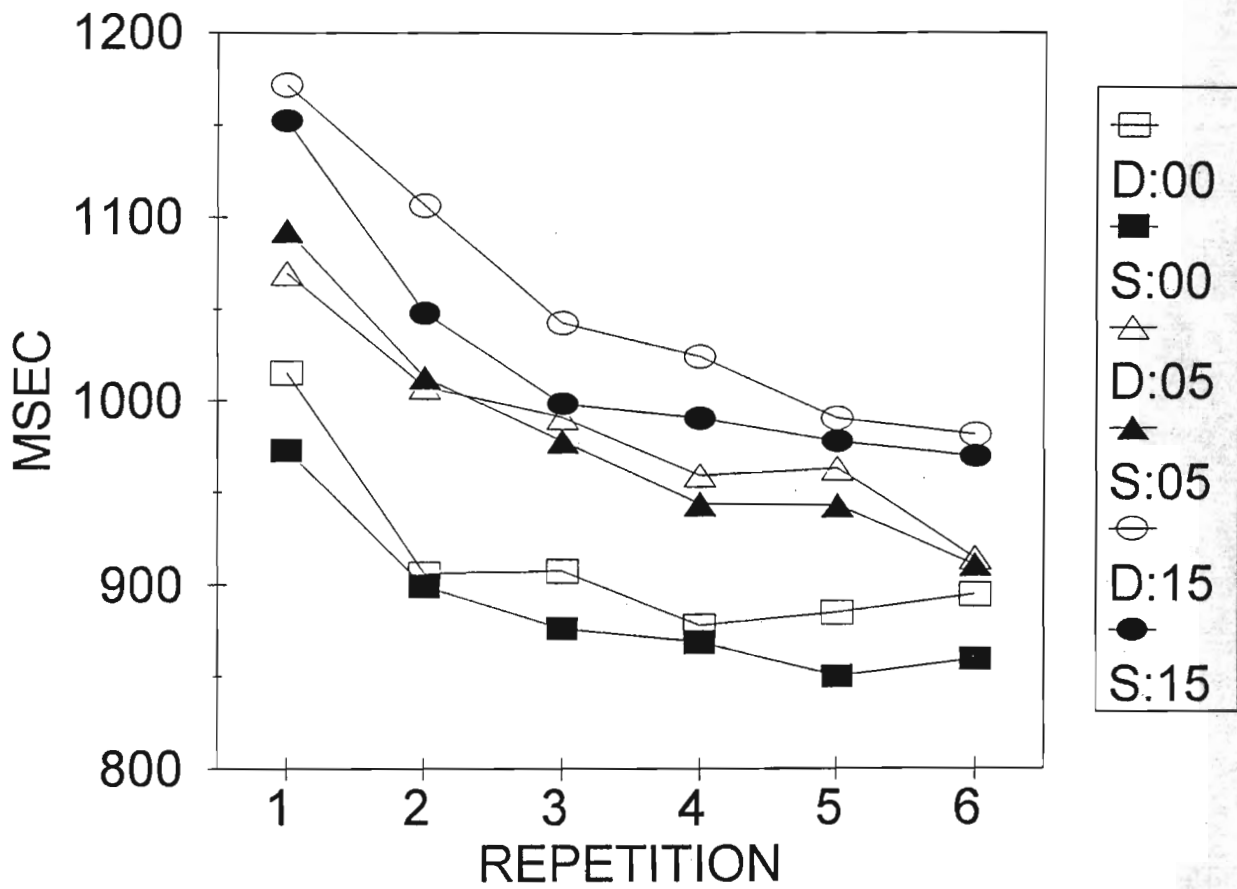


Figure 2. Latency of correct word identifications in Experiment 1 by level of noise (00 = no noise, 05 = low noise, 15 = high noise), repetition, and voice (S = same, D = different).

Discussion

We found no difference in accuracy between same- and different-voice trials. Subjects were, however, faster to classify same-voice words than different-words. As discussed previously, studies that show differences in percentage of spoken words correctly identified based on voice manipulations have all employed an open-set identification design (Church & Schacter, 1994; Goldinger, 1992). Previous research has shown that closed-set identifications may lead to a different set of strategies on the part of the listeners; variables found to affect spoken word identification, such as word frequency or lexical density, are known to be reduced or eliminated in some closed-set tests (Sommers et al., 1992). In this case, the ordering of words within a list may have provided the words a contextual component that increased explicit recognition for the words and attenuated the perceptual effects of the voice manipulation.

The present findings differ from those reported by Church and Schacter (1994). They found a reduced level of priming on word identification in their Experiment 1 due to study-to-test voice changes. In an earlier study (Schacter & Church, 1992), they also reported better performance for same-voice trials on auditory stem-completion tasks (Experiments 3 and 4). The lack of an effect of voice on accuracy in the present study in no way suggests that the effect does not exist; Church and Schacter's studies demonstrated that priming based on voice manipulations may be found under the right set of conditions. Rather, the present results demonstrate clearly that it is possible to construct perceptual tests that will not reveal priming due to voice variables, and that the details of the test must be considered carefully in drawing conclusions about differences (or lack thereof) between groups due to the manipulation of speaker characteristics. By the same token, it is also possible to construct auditory recognition tests that will or will not show effects of voice variables on performance. Experiment 2 was designed with the aim of maximizing the effects of implicit processing on explicit word recognition memory due to speaker characteristics.

Experiment 2

Presenting words in the same voice in both study and test sessions has been shown to have a facilitating effect on recognition in some studies (Craik & Kirsner, 1974; Goldinger, 1992; Palmeri et al., 1993; Sheffert & Fowler, 1994) but has failed to affect performance in others (Church & Schacter, 1994; Jackson & Morton, 1984). Experiment 2 extended the previous experiment by including a surprise recognition test in order to investigate the effects of study-to-test changes in voice on explicit memory performance.

Church and Schacter (1994) have argued that the failure of study-to-test voice matching to improve recognition was due to the conceptual nature of recognition memory. "The dissociation between implicit and explicit memory is consistent with the hypothesis that auditory priming is mediated by a PRS that stores information about the acoustical form of the words, whereas explicit memory is mediated by an episodic system that generally relies on conceptually driven processes (pp. 523-524)." If this claim is correct, then eliminating the conditions which promote conceptual processing during study and test should permit effects of perceptual encoding to occur, if they exist. The design used in Experiment 1 served this purpose well: The words used in that experiment were high frequency and bore no obvious relationship to each other, the classification task itself did not require elaborative processing, and subjects were not instructed to attend to the voices speaking the words. Thus, Experiment 1 was appropriate for use as the training phase in the present experiment.

A recognition test was added to this design after an initial training phase; the procedure was based on the methods used in Experiment 1 in which a word was presented for identification in the same voice

across six repetitions of the word list. During the recognition test, words were presented in the same or different voices. We hypothesized that target words in the same voice would be better recognized than words in a different voice.

As noted earlier, there has been some concern expressed about the effect of noise on implicit priming of recognition. Schacter and Church (1992) and Jackson and Morton (1984) found no effect of voice on words presented in noise. Church and Schacter (1994) attributed this result to the specific effects of white noise on word recognition. As in Experiment 1, we also included three levels of stimulus degradation in order to assess the effects of noise and voice manipulations on word recognition.

Method

Subjects. Sixteen subjects served in the no-noise condition, 21 were assigned to the low-noise condition, and 21 were assigned to the high-noise condition. All were recruited from the student population at Indiana University and received course credit for their participation. Subjects were tested in groups of six or fewer. All were native speakers of English who reported no speech or hearing disorders at the time of testing.

Procedure. The training phase followed the same procedure used in the previous experiment, with a single modification to the stimuli: All 48 words were spoken in the same voices across all six presentations of the words. The six speakers spoke 8 words each. Training was followed by a surprise recognition test in which subjects were presented with all 48 target words plus 48 foils. The foils were selected from the 48 MRT word groups. Each target had an associated foil word (e.g., "bad" and "bat"); none of the foil words appeared as alternatives during the training phase. Half of the targets were presented in the same voice as in the training phase, and half in a different voice. Likewise, half of the foils were presented in the same voice as their associated target words, half in a different voice.

Results

Figure 3 shows the percentage of correct responses during the training phase for all three levels of stimulus degradation across six repetitions of the word lists. An analysis of the perceptual identification data was conducted and it largely replicated the results obtained in Experiment 1 for accuracy and response latency. An ANOVA conducted for accuracy revealed main effects due to Noise, $F(2, 55) = 66.1, p < .0001$, and Repetition, $F(5, 275) = 76.3, p < .0001$. The analysis also revealed a Noise x Repetition interaction, $F(10, 275) = 15.1, p < .0001$.

Insert Figure 3 about here

Likewise, the data for latency of correct responses were also highly similar to that collected in Experiment 1. There were main effects of Noise and Repetition, $F(2, 55) = 15.6, p < .0001$, and $F(5, 275) = 74.8, p < .0001$, respectively. The Noise x Repetition interaction was not significant, $F(10, 275) = 1.3, p > 0.2$. As shown in Figure 4, subjects in the no-noise condition performed at asymptote throughout the six repetitions, while subjects in the noise conditions reached asymptotic performance on the third or fourth repetition.

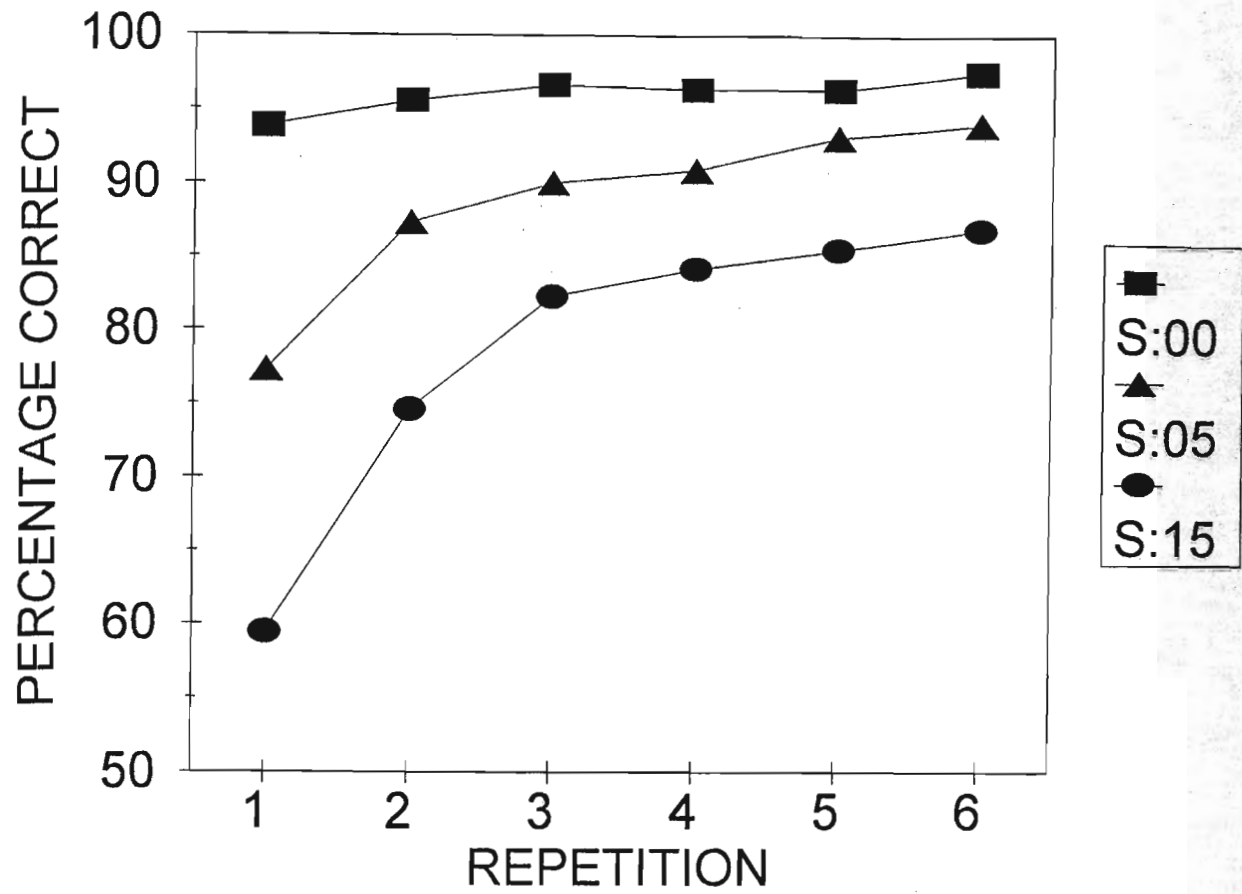


Figure 3. Percentage of correct word identifications in Experiment 2 by level of noise (00 = no noise, 05 = low noise, 15 = high noise) and repetition.

Insert Figure 4 about here

For the recognition memory data, accuracy scores were computed for same- and different-voice targets and foils for all three noise conditions. As hypothesized, subjects were more accurate in recognizing the same-voice targets than the different-voice targets, $F(1, 55) = 11.2, p < .001$. There was no effect of Voice on the accuracy of detecting foils, $F < .2$. As expected, we found a main effect of Noise on accuracy, $F(2, 55) = 7.5, p < .001$, but no Voice \times Noise interaction, $F(2, 55) = .425$. Table 1 shows percentage of correct responses for each of the three noise conditions.

Insert Table 1 about here

A signal-detection analysis confirmed that subjects were better able to discriminate between targets and foils on same-voice trials than on different-voice trials. The d 's were calculated for each subject then averaged across subjects within each noise condition. The advantage for recognition of same-voice words over different-voice words was consistent across all three levels of stimulus degradation. See Table 2.

Insert Table 2 about here

Discussion

As anticipated, subjects were more accurate in recognizing same-voice target words than different-voice target words in the recognition memory task. This finding replicated the studies by Palmeri et al. (1993) and Sheffert and Fowler (in press) with respect to the facilitating effects of voice on recognition memory. There was also a main effect of noise on recognition. In contrast to Schacter and Church (1992) and Jackson and Morton (1984), we found an effect of voice for words presented in the clear as well as for the degraded stimuli; in fact, recognition performance was *higher* for target words in the noise conditions than for target words presented in the clear. Together with Experiment 1, the results of Experiment 2 demonstrate a dissociation between implicit and explicit memory tests. In this case, however, the dissociation observed was in the opposite direction of that reported by Church and Schacter. In our experiments, we found no effect of voice on accuracy in the perceptual task but a significant effect of voice on recognition memory performance.

It should be noted again that the stimuli used in the present experiments were not degraded by white noise but by a signal-processing algorithm that preserves the spectral envelope of the speech waveform. Perhaps this difference accounts for the success of the voice manipulation on the recognition tests. On the other hand, previous studies have found similar identification performance scores independent of the stimulus degradation technique (Sommers, Nygaard, & Pisoni, 1994). More research on the effects

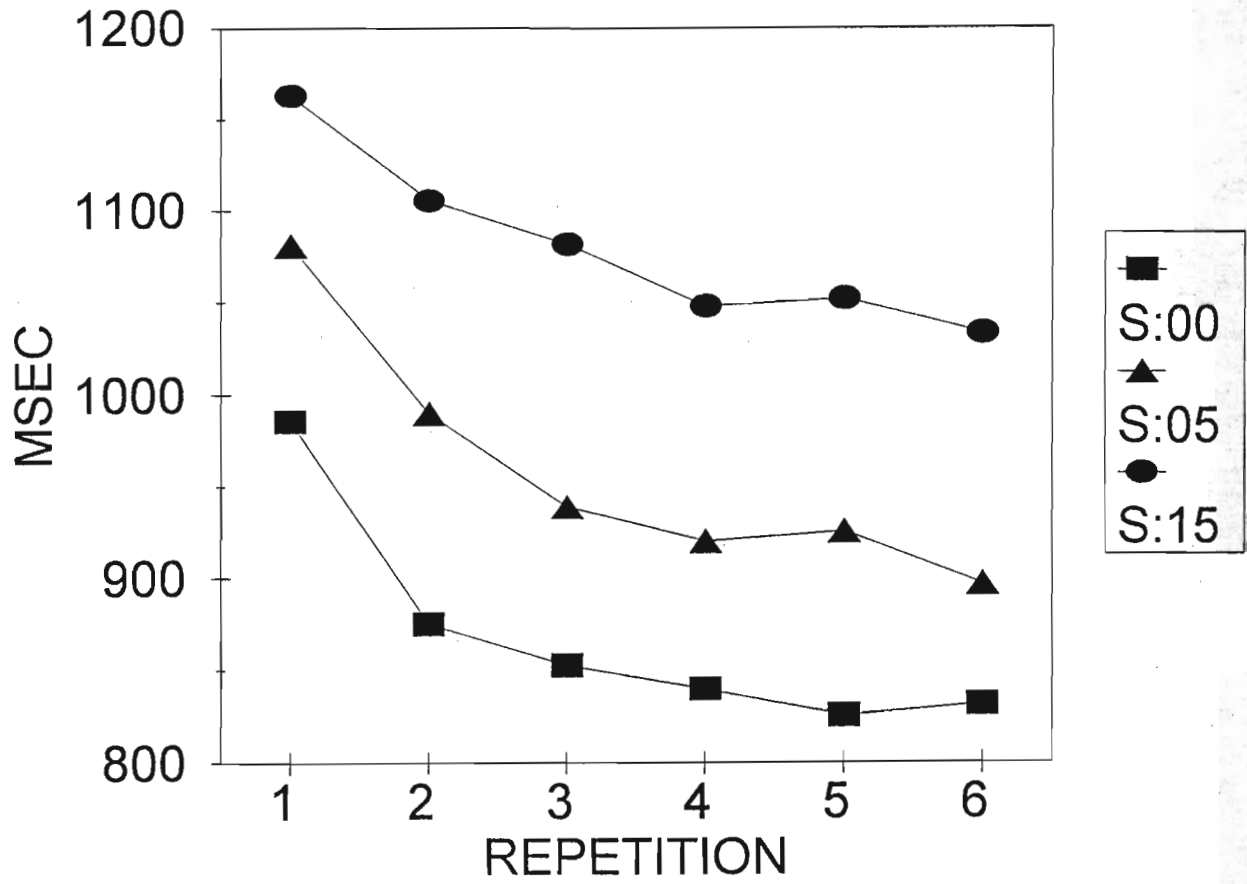


Figure 4. Latency of correct word identifications in Experiment 2 by level of noise (00 = no noise, 05 = low noise, 15 = high noise) and repetition.

Table 1

Mean percentage of correct recognition scores for same- and different-voice words in Experiment 2.

	Target		Foil	
	Same	Different	Same	Different
No noise	80.2	76.8	88.5	88.8
RBF05	70.8	64.7	83.3	82.1
RBF15	66.8	59.7	55.7	58.3
	72.6	67.1	75.8	76.4

Table 2

Mean d's for correct recognition scores for same- and different-voice words in Experiment 2.

	Voice	
	Same	Different
No noise	2.40	2.31
RBF05	1.51	1.47
RBF15	0.61	0.49

of various stimulus degradation techniques and their effects on different perceptual and memory tests is needed. Nevertheless, our results show that the effects of voice on auditory recognition memory hold for degraded words as well as for words presented in the clear.

"Implicit memory is revealed when previous experiences facilitate performance on a task that does not require conscious or intentional recollection of those experiences" (Schacter, 1987, p. 501). Word identification, naming, or lexical decision have been considered to be implicit memory tasks to the extent that subjects are not required to consciously recollect previous events in order to perform the task. At the same time, the so-called explicit tests, such as recognition memory, may show implicit effects if subjects are not directed to attend to the variable, such as voice, that account for differences in performance.

One of the assumptions of our research on speech perception has been that listeners, in perceiving speech, encode some of the acoustic attributes of the speaker's voice, especially those characteristics that permit the listener to identify voices in the speaker ensemble (Nygaard, Sommers, & Pisoni, 1994). This information can facilitate performance in tests of spoken word recognition, recall, and perceptual identification. We have further argued that voice information is encoded incidentally and utilized implicitly (Pisoni, 1993). This hypothesis continues to gain support as further studies are conducted that employ experimental procedures for dissociating implicit memory from explicit tests of memory such as cued recall and recognition.

General Discussion

The distinction between implicit and explicit memory has created an accepted framework within which to address issues of memory and the apparent dissociation between different classes of memory tests. Since the earliest days of research on implicit memory, a consistent finding has been that the manipulation of the surface features of words produce effects on perceptual tests but none on explicit tests (Roediger & McDermott, 1993; Schacter, 1987). It is useful to remember that this framework was developed largely from experiments with visually presented words; it is acknowledged that few studies have dealt with memory dissociations using spoken words (Goldinger, 1992; Schacter & Church, 1992). Furthermore, it is within this general framework that Schacter and colleagues have derived their ideas about the auditory perceptual representational system. Research on spoken language, though, carries with it a set of additional concerns that are not addressed by written language. Indeed, the literature dealing with the manipulation of typography and other surface characteristics of written language on memory is often messy and contradictory (Roediger & McDermott, 1993). The general framework on memory developed out of work with visual language is important, but it is also important to consider the particular sensory modalities involved and the specific details of the tests that are carried out (Mandler, 1980).

The results reported here show that the episodic memory for a speaker's voice are preserved and can prime recognition memory. This priming is by definition implicit (see Schacter, 1987), as subjects are not directed to attend to the speaker's voice during either study or test. These findings are in sharp contrast to the recent findings of Church and Schacter (1994) and conflict with recent theoretical accounts concerning the PRS, which describes explicit recognition memory as a largely conceptual memory system, unaffected by the perceptual characteristics of stimuli.

Throughout this article we have referred to "explicit" and "implicit" tests of memory. In fact, there may be no "pure" tests of implicit or explicit memory; the terminology confuses tests and processes (Jacoby, 1991). The identification test used in Experiment 1 had a substantial explicit component which may have masked any differences due to implicit factors. By the same token, the recognition memory test

may have both implicit and explicit components. With factors such as semantic encoding and context held constant as we attempted to do here, implicit effects due to a match between voice during study and test could account for the observed differences in recognition scores. Although it may seem to some readers that the size of the effects on target recognition in Experiment 2 (ranging from 3.4% to 7.1%, depending on the condition) are less than overwhelming, they are as large as the identification data reported by Church and Schacter (1994). Both Palmeri et al. (1993) and Sheffert and Fowler (in press) also reported effects on recognition memory tests in the range of 5% to 10%.

At the present time, there is much debate on the nature of memory dissociation. Both the processing theories and the systems theories attempt to account for observed dissociation between perceptual and recollective tasks. The present experiments on spoken word identification and recognition do not select between these competing theories of memory dissociation. However, processing accounts would seem to be more amenable for describing the types of implicit effects on recognition tests described in Experiment 2 (see Jacoby, 1991). The PRS theory does not, in its present form, seem to be able to account for the possibility of implicit influences of voice on auditory word recognition. A systems theory should not necessarily preclude the possibility of multiple systems acting together on recognition. It may turn out that separate memory systems operate independently, but in parallel, each contributing to the identification and recognition of a given item. In light of the recent findings reported by Palmeri et al. (1993), Sheffert and Fowler (in press), and the present experiment showing robust voice effects on recognition tests, it seems necessary not only to acknowledge the influence of perceptual effects on explicit tests of memory but to attempt to account for the apparent failure of Church and Schacter to find these results in their experiments.

A number of theorists have argued that the multiple systems theories of memory seem to be superior to processing theories in accounting for memory deficits (Roediger, 1990). Interestingly, dissociations with respect to voice identification and recognition have been observed among clinical patients with specific perceptual deficits (VanLancker, Cummings, Kreiman, & Dobkin, 1988; VanLancker, Kreiman, & Cummings, 1989). The anatomic correlates of memory for voices are only now being explored (VanLancker et al., 1989). While preliminary, one study concluded that temporal lobe damage impairs general voice discrimination; inferior and parietal regions of the right hemisphere impair familiar voice recognition (VanLancker et al., 1988). Behavioral studies with trained listeners on word identification tests suggest that words that are spoken in familiar voices may be processed in a manner somewhat differently than words spoken by unfamiliar voices (Nygaard et al., 1994). These studies tentatively suggest the existence of a unique system for the recognition of familiar voices.

It would seem plausible that the same system is engaged during both the perceptual identification and the recognition of spoken words (VanLancker et al., 1988). Obviously, the recognition of a word must begin with the perception of the word. It is at this level that the implicit effects of speaker characteristics may have their initial effect. With what we have discovered in the past few years about the role of speaker information on spoken language processing, it is safe to say that any comprehensive theory of memory for spoken language must now take into account the robust effects of voice on perception and memory. It is hoped that further research in this area will accommodate theory from both basic memory research and from speech perception and spoken language processing.

References

- 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.
- Craik, F. I. M., & Kirsner, K. (1974). The effect of speaker's voice on word recognition. *Quarterly Journal of Experimental Psychology*, *26*, 274-284.
- Goldinger, S. D. (1992). Words and voices: Implicit and explicit memory for spoken words. *Research on Speech Perception Technical Report No. 7* (Indiana University, Bloomington, IN).
- House, A. S., Williams, C. E., Hecker, M. H. L., & Kryter, K. D. (1965). Articulation-testing methods: Consonantal differentiation with a closed-response set. *Journal of the Acoustical Society of America*, *37*, 158-166.
- Jackson, A., & Morton, J. (1984). Facilitation of auditory word recognition. *Memory & Cognition*, *12*, 568-574.
- Jacoby, L. L. (1991). A process dissociation framework: separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513-541.
- Lively, S. E. (1994). Preserving the perceptual record: Retention of talker-specific information in long-term memory. *Research on Speech Perception Technical Report No. 9*. Indiana University, Bloomington, IN.
- Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*, 252-271.
- Masson, M. E. J. (1984). Memory for the surface structure of sentences: Remembering with and without awareness. *Journal of Verbal Learning and Verbal Behavior*, *23*, 579-592.
- 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.
- Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, *5*, 42-46.
- Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 1-20.
- Pisoni, D. B. (1993). Long-term memory in speech perception: Some new findings on talker variability, speaking rate and perceptual learning. *Speech Communication*, *13*, 109-125.
- Roediger, H. L. (1990). Implicit memory: Retention without remembering. *American Psychologist*, *45*, 1043-1056.

- Roediger, H. L., & McDermott, K. B. (1993). Implicit memory in normal human subjects. In F. Boller & J. Grafman (Eds.), *Handbook of Neuropsychology Vol. 8* (pp. 63-131). Elsevier.
- Schacter, D. L. (1987). Implicit memory: history and current status. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *13*, 501-518.
- Schacter, D. L. (1990). Perceptual representation systems and implicit memory: Toward a resolution of the multiple memory systems debate. In A. Diamond (Ed.), *Annals of the New York Academy of Sciences: Volume 608. Development and neural bases of higher cortical functions* (pp. 543-571). New York: New York Academy of Sciences.
- Schacter, D. L. (1992). Understanding implicit memory: A cognitive neuroscience approach. *American Psychologist*, *47*, 559-569.
- Schacter, D. L., & Church, B. (1992). Auditory priming: Implicit and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 915-930.
- Schacter, D. L., Church, B., & Bolton, E. (in press). Implicit memory in amnesic patients: Impairment of voice-specific priming. *Psychological Science*.
- Schacter, D. L., Church, B., & Osowiecki, D. M. (1994). Auditory priming in elderly adults: Impairment of voice-specific implicit memory. *Memory*, *2*, 295-323.
- Schacter, D. L., Church, B., & Treadwell, J. (1994). Implicit memory in amnesic patients: Evidence for spared auditory priming. *Psychological Science*, *5*, 20-25.
- Sheffert, S. M., & Fowler, C. A. (in press). The effects of voice and visible speaker change on memory for spoken words. *Journal of Memory and Language*.
- Shepard, R. N., & Teghtsoonian, M. (1961). Retention of information under conditions approaching a steady state. *Journal of Experimental Psychology*, *62*, 302-309.
- Sommers, M. S., Nygaard, L. C., & Pisoni, D. B. (1992, October). Stimulus variability and the perception of spoken words: Effects of variations in speaking rate and overall amplitude. *Proceedings of the 1992 International Conference on Spoken Language Processing*, Banff, Canada.
- Sommers, M. S., Nygaard, L. C., & Pisoni, D. B. (1994). Stimulus variability and spoken word recognition. I. Effects of variability in speaking rate and overall amplitude. *Journal of the Acoustical Society of America*, *96*, 1324-1324.
- Squire, L. R. (1987). *Memory and brain*. New York: Oxford University Press.
- Tulving, E., Schacter, D. L., & Stark, H. A. (1982). Priming effects in word fragment completion are independent of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*, 336-342.
- VanLancker, D. R., Cummings, J. L., Kreiman, J., & Dobkin, B. H. (1988). Phonagnosia: a dissociation between familiar and unfamiliar voices. *Cortex*, *24*, 195-209.

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

Appendix

Words used as auditory stimuli in Experiments 1 and 2.

1. BAT	BAD	25. LED	SHED
2. BEAN	BEACH	26. SOLD	TOLD
3. BUS	BUT	27. BIG	RIG
4. CAPE	CANE	28. LICK	TICK
5. CUT	CUFF	29. BOOK	TOOK
6. DIG	DIP	30. DARK	BARK
7. DUCK	DUD	31. MALE	TALE
8. FIG	FIZZ	32. REEL	FEEL
9. HEAR	HEAL	33. HILL	BILL
10. KING	KID	34. FOIL	COIL
11. LATE	LAKE	35. FAME	NAME
12. MAT	MAN	36. TEN	PEN
13. PAGE	PAIN	37. GUN	FUN
14. PASS	PAT	38. RANG	FANG
15. PEAS	PEAK	39. TENT	BENT
16. PILL	PIT	40. RIP	TIP
17. PUFF	PUP	41. TOP	HOP
18. RAKE	RACE	42. MEAT	FEAT
19. SAKE	SAVE	43. BIT	HIT
20. SAD	SACK	44. HOT	GOT
21. SEEM	SEEK	45. NEST	TEST
22. DUB	SUB	46. BUST	JUST
23. TACK	TAP	47. RAW	PAW
24. TEACH	TEASE	48. WAY	SAY