



Published in final edited form as:

*Percept Psychophys.* 1999 February ; 61(2): 206–219.

## Effects of talker, rate, and amplitude variation on recognition memory for spoken words

**ANN R. BRADLOW,**

Northwestern University, Evanston, Illinois

**LYNNE C. NYGAARD,** and

Emory University, Atlanta, Georgia

**DAVID B. PISONI**

Indiana University, Bloomington, Indiana

### Abstract

This study investigated the encoding of the surface form of spoken words using a continuous recognition memory task. The purpose was to compare and contrast three sources of stimulus variability—talker, speaking rate, and overall amplitude—to determine the extent to which each source of variability is retained in episodic memory. In Experiment 1, listeners judged whether each word in a list of spoken words was “old” (had occurred previously in the list) or “new.” Listeners were more accurate at recognizing a word as old if it was repeated by the same talker and at the same speaking rate; however, there was no recognition advantage for words repeated at the same overall amplitude. In Experiment 2, listeners were first asked to judge whether each word was old or new, as before, and then they had to explicitly judge whether it was repeated by the same talker, at the same rate, or at the same amplitude. On the first task, listeners again showed an advantage in recognition memory for words repeated by the same talker and at same speaking rate, but no advantage occurred for the amplitude condition. However, in all three conditions, listeners were able to explicitly detect whether an old word was repeated by the same talker, at the same rate, or at the same amplitude. These data suggest that although information about all three properties of spoken words is encoded and retained in memory, each source of stimulus variation differs in the extent to which it affects episodic memory for spoken words.

---

A long-standing problem for theories of speech perception and spoken word recognition has been perceptual constancy in the face of a highly variable speech signal. Listeners extract stable linguistic percepts from an acoustic speech signal that varies substantially due to idiosyncratic differences in the size and shape of individual talkers’ vocal tracts as well as to differences within and among talkers in factors such as speaking rate, dialect, speaking style, and vocal effort. Traditionally, researchers have adopted an abstractionist approach to the problem of perceptual constancy, assuming that variability in the speech signal is perceptual “noise” that must be “stripped away” during perception to arrive at a series of abstract, canonical linguistic units (see Pisoni, 1997). Research has typically focused either on searching for sets of acoustic, articulatory, or relational invariants hypothesized to allow access to phoneme- and ultimately word-sized units (e.g., Kewley-Port, 1983; Stevens & Blumstein, 1978) or on normalization algorithms and processes that would successfully filter out stimulus variation to arrive at the abstract units thought to underlie further

---

Copyright 1999 Psychonomic Society, Inc.

Correspondence should be addressed to A. R. Bradlow, Department of Linguistics, Northwestern University, 2016 Sheridan Road, Evanston, IL 60208-4090 (abradlow@nwu.edu).

An earlier version of this study was presented at the 131st meeting of the Acoustical Society of America in Indianapolis, May 1996.

linguistic processing (Gerstman, 1968; Halle, 1985; Johnson, 1990; Joos, 1948; Ladefoged & Broadbent, 1957; Nearey, 1989).

Recently, however, a growing body of research has begun to call into question the fundamental underlying assumptions of the traditional abstractionist approach, and researchers have begun to explicitly investigate the effects of stimulus variability on a variety of speech perception and spoken word recognition tasks (e.g., Mullennix, Pisoni, & Martin, 1989; Sommers, Nygaard, & Pisoni, 1994; Stevens, 1996). The general orientation of this research regards the inherent variability in the speech signal due to different talker- and other instance-specific characteristics as a useful source of information to the listener about the communicative situation (Laver, 1989; Laver & Trudgill, 1979). Indeed, even aspects of a talker's voice can provide considerable information about the individual identity, health, age, and even emotional state of the talker and can be used by listeners both to interpret linguistic content and to guide their own speech production (see Nygaard, Sommers, & Pisoni, 1994).

Due to the importance of surface characteristics during speech communication, an alternative approach has emerged that assumes that variability is incorporated into lexical representations along with linguistic content (Goldinger, 1996; Pisoni, 1993, 1997). Linguistic representations in long-term memory are hypothesized to be extremely detailed, preserving in memory the constantly changing surface form of each spoken word (Goldinger, 1996). In particular, drawing from exemplar-based theories of memory (Eich, 1982; Hintzman, 1986) and categorization (Nosofsky, 1991), episodic theories of the lexicon have been proposed that assume that collections of detailed traces represent individual words (Goldinger, 1996; Pisoni, 1997; Tenpenny, 1995). In a strong form, episodic theories of speech perception predict that all aspects of surface form are included in lexical representations and that they affect both the perception and the memory of spoken words.

With respect to the perception of spoken words, there is now considerable evidence that at least some aspects of the surface form of spoken words influence spoken word recognition (Cole, Coltheart, & Allard, 1974; Creelman, 1957). For example, Mullennix et al. (1989) investigated the effects of talker variability<sup>1</sup> on perception and showed that word recognition accuracy decreased and response times increased when listeners were presented with lists of spoken words produced by multiple talkers relative to a condition in which listeners were presented with the identical words produced by only a single talker. In addition, Mullennix and Pisoni (1990) found that listeners had a difficult time ignoring irrelevant talker variation when they were asked to classify words by initial phoneme in a Garner (1974) speeded classification task. Their results suggest that these two perceptual dimensions, talker and phoneme, are processed in an integral fashion.

Variability in speaking rate has also been shown to affect speech perception and spoken word recognition (Miller & Liberman, 1979; Miller & Volaitis, 1989; Summerfield, 1981; Volaitis & Miller, 1992). For example, Sommers et al. (1994) showed a decrease in word identification scores for mixed-speaking-rate lists relative to single-speaking-rate lists. Interestingly, however, no decrease in word identification scores was found for lists of words that had mixed overall amplitudes relative to lists of words presented at a single

---

<sup>1</sup>Throughout this manuscript, we maintain a distinction between the terms *voice* and *talker*. We reserve the term *voice* for reference to qualitative aspects of an utterance (involving both glottal source and vocal tract filter characteristics), and we use *talker* to refer to the individual who produced the utterance. Whereas a change in talker necessarily entails a change in voice, a change in voice does not necessarily entail a change in talker. In other words, it is possible for a given talker to produce two utterances with different voice characteristics—for example, due to a change in physical or mental state, or due to an intentional vocal disguise. In our study, as in most others that we cite, we investigated the effects of talker variability. However, since a change in talker entails a change in voice, it remains for future research to disentangle the effects of talker and voice variability.

overall amplitude. These results suggest that variability due to talker characteristics and speaking rate is both time and resource demanding. As the talker or speaking rate changes from trial to trial in these tasks, fewer processing resources are available for extracting the phonetic content of each word, resulting in higher error rates and longer response times in high- rather than low-variability contexts. Variability in overall amplitude, however, does not appear to be resource demanding. Error rates and response times did not change with high versus low variability in overall amplitude.

Taken together, these previous findings indicate that although some forms of stimulus variability in the speech signal do seem to affect the perception of spoken words, all sources of variability may not. One explanation, consistent with the traditional abstractionist approach, is that a time- and resource-consuming normalization process is responsible for the effects of variability on perception. Differences among talker, rate, and overall amplitude variability occur because each source of variability engages these resource-demanding normalization processes to a greater or lesser extent. In particular, variability in talker characteristics and speaking rate may require more extensive normalization procedures than variation due to changes in overall amplitude. An alternative explanation is that decreased performance in high- as opposed to low-variability contexts reflects the additional time and resources needed to encode information provided by the changing surface form (Goldinger, Pisoni, & Logan, 1991). According to this account, differences among sources of variability may reflect differences in perceptual saliency among the different perceptual dimensions under consideration. For example, Sommers et al. (1994) suggested that the effects of talker and rate variability on word recognition may be due to the relevance of these dimensions for the perception of phonetic contrasts (Ladefoged & Broadbent, 1957; Miller, 1987). In contrast, variability in overall amplitude has not been found to affect phonetic categorization, and therefore may not demand limited processing resources during word recognition.

## Memory for Talkers

To determine whether perceptual effects of stimulus variation are due to time-consuming normalization processes or due to the increased encoding time for perceptual detail, other studies have investigated the effects of stimulus variability on memory for spoken words (e.g., Goldinger et al., 1991; Martin, Mullennix, Pisoni, & Summers, 1989; Nygaard, Sommers, & Pisoni, 1995; Palmeri, Goldinger, & Pisoni, 1993; for reviews, see Pisoni, 1993, 1997). The idea is that if these perceptual effects are due to a normalization process, then memory for spoken words should be based on abstract symbolic linguistic representations. If, however, perceptual effects are due to increased resources needed to encode perceptual detail, then memory for spoken words should be affected by the surface form of the speech signal. Turning first to talker variability, Martin et al. found that listeners performed better in a serial recall task when the to-be-remembered words were produced by a single talker than when the same words were produced by multiple talkers. This difference in serial recall of spoken words was selective in nature and was located only in the primacy portion of the serial position curve—that is, for the first three words in 10-word lists. Martin et al. proposed that this finding arose from the increased processing demands incurred by increased stimulus variability. Additional processing requirements interfered with listeners' ability to maintain and rehearse information in working memory and to transfer this information to long-term memory.

In a follow-up study, Goldinger et al. (1991) investigated further the nature of talker variability effects on serial recall of spoken word lists by varying the rate of presentation of the items in the list to be recalled. Goldinger et al. hypothesized that rate of presentation would selectively affect the listener's ability to encode the distinctive talker information for

multiple-talker lists. If given enough rehearsal time, listeners might be able to use the distinctive talker information as an additional retrieval cue, and thus the multiple-talker lists would be more accurately recalled than the single-talker lists. Indeed, Goldinger et al. found that at fast presentation rates (one word every 250 msec), words in the primacy portion of the single-talker lists were more accurately recalled than those from multiple-talker lists, whereas at slow presentation rates (one word every 4,000 msec), this difference in recall accuracy was reversed. These results showed that information about a talker is encoded into long-term memory and can be used as an effective retrieval cue under optimal conditions.

In a study of recognition memory for spoken words, Palmeri et al. (1993) found that detailed information about a talker is retained in memory and facilitates recognition of a previously encountered word. Specifically, Palmeri et al. found that listeners were better at recognizing a word as a repeated item (i.e., “old”) in a continuous list of spoken words when the word was repeated by the same talker than when the talker differed from first to second repetition. Furthermore, Palmeri et al. showed that when listeners recognized that the word was a repeated word in the list, they were also able to explicitly recognize whether the talker was the same or different as in the first occurrence of the word.

Additional evidence for the retention of talker characteristics in long-term memory comes from a series of studies conducted by Schacter and Church (1992) and Church and Schacter (1994). Using an auditory priming task, these studies showed that talker information had a significant effect on measures of implicit memory such as auditory stem completion, but not on measures of explicit memory such as cued recall or recognition. For example, words were more likely to be produced as a stem completion if the stem was repeated by the same talker at study and test. Church and Schacter also found effects of intonation and fundamental frequency on measures of implicit memory. Similarly, Goldinger (1996) found that detailed talker information appeared to be retained in memory and used in perceptual identification (implicit task) and recognition memory (explicit task) tasks. Words were better identified and recognized when they were repeated by the same talker than when they were repeated by a different talker (see also Sheffert, 1998a, 1998b). Further, voice similarity affected the amount of repetition benefit. Words repeated in a similar voice were more likely to be identified or recognized than words repeated in a less similar voice.

Taken together, these recent studies suggest that talker information is encoded and retained in long-term memory representations of spoken words (see also Craik & Kirsner, 1974; Geiselman, 1979; Geiselman & Bellezza, 1976, 1977; Geiselman & Crawley, 1983; Sheffert & Fowler, 1995). Thus, the effects of talker variability on word recognition appear to be due to the additional time and resources needed to encode and store distinctive talker information, rather than to a time-consuming normalization process.

## Memory for Speaking Rate and Overall Amplitude

Although the evidence for the retention of talker information is relatively clear, considerably less attention has been paid to the encoding and retention of other sources of stimulus variability in surface form. Is it the case that all sources of stimulus variation are encoded and represented in long-term memory to the same extent? Do episodic traces of spoken words preserve all perceptual details of surface form, or does some abstraction and loss occur during spoken word recognition? If abstraction does occur, what factors might determine which aspects of surface form are preserved and which ones are lost? The aim of the present experiments was to investigate these questions by comparing and contrasting three different kinds of surface characteristics—talker, speaking rate, and overall amplitude—using a continuous recognition memory task. Our goal was to determine whether these

sources of variation differ in the extent to which each is encoded and used during recognition memory.

Some clues to the retention of surface form other than talker's voice come from an earlier study conducted by Nygaard et al. (1995). Their study compared the effects of speaking rate and overall amplitude on serial recall with the effects of talker variability. Nygaard et al. (1995) found that at fast presentation rates, items presented early in lists spoken either by a single talker or at a single speaking rate were better recalled than the same items spoken by multiple talkers or at multiple speaking rates. At a slow presentation rate, however, early items in the multiple-talker lists were better recalled than those in the single-talker lists; however, this reversal of recall accuracy was not obtained for the items in the multiple-rate lists relative to those in the single-rate lists. Rather, at the slow presentation rate, there was no difference between recall of items in the multiple- and single-rate lists. Furthermore, Nygaard et al. (1995) found no differences between serial recall of single- and multiple-amplitude lists at fast or at slow presentation rates. Taken together, these results suggest again that distinctive talker information is encoded in the long-term memory representation of spoken words, and if given sufficient rehearsal time, this additional distinctive information can be used as a retrieval cue by the listener. In contrast, the data from these serial recall experiments did not provide any evidence that either speaking rate or overall amplitude is encoded in long-term memory along with the linguistic content of a spoken word.

The results of the Nygaard et al. (1995) study were somewhat surprising. Although both talker and speaking rate variability had been shown in earlier experiments to have substantial effects on the perception of spoken words (Mullennix et al., 1989; Sommers et al., 1994), only talker information appeared to be retained and used in the serial recall task. Consequently, the present experiments were designed to further examine talker, speaking rate, and overall amplitude variability to determine the extent to which each source of variability is encoded and represented in episodic memory. In particular, we were interested in addressing four issues. The first was to determine whether or not talker, rate, and amplitude would differ in their effects on recognition memory. Thus, one purpose of the study was to evaluate the specificity of the lexical representations that are formed in this task and to evaluate the strong version of episodic-based theoretical accounts of lexical representation. The rationale was to determine whether all sources of variability affect recognition memory to the same extent. If so, lexical representations are likely to consist of collections of veridical perceptual traces. However, if all sources of variation do not affect recognition memory to the same extent, then some mechanism for selectively preserving details of surface form must be included in accounts of lexical representation.

The second issue we sought to address was a broad comparison of the effects of variability on serial recall versus recognition memory. Although not a direct comparison, the present experiments sought to evaluate recognition memory for the same types of surface variability previously evaluated in a serial recall task. Both Goldinger et al. (1991) and Nygaard et al. (1995) found reliable but small effects of talker variability on serial recall, whereas Nygaard et al. (1995) failed to find comparable effects of speaking rate and overall amplitude.

In the present experiments, stimulus variability was investigated in a paradigm that had shown robust effects of talker variability and that might provide a more sensitive test than serial recall of the effects of speaking rate and overall amplitude on memory retention. Serial recall tasks assess the extent to which listeners can explicitly remember the serial order of a list of items. For talker variation, each word in the list was produced by a different talker, therefore providing distinctive cues to serial order and item identity for each word in the serial recall list. In contrast, for speaking rate variation, three different speaking rates were

used, and approximately one third of the words in each list was produced at each rate. Arguably, in the case of speaking rate, the information for serial order was not as distinctive (see Nygaard et al., 1995, for a complete discussion). In the continuous recognition memory task, listeners are not required to explicitly encode word identity and serial order; rather, word recognition is simply facilitated by previous presentations of the word and may be more sensitive to effects of surface form. If continuous recognition proves to be a more sensitive test of long-term memory for surface characteristics, the outcome of these experiments could be used to evaluate whether the salience or relevance of each surface form (as judged by the effect of each surface form on perceptual tests of spoken word recognition) affects the degree to which different surface characteristics are encoded and retained in memory.

The third issue we sought to address was the time course of the effects of stimulus variability. In the continuous recognition memory paradigm, words are presented and later repeated after a varying number of intervening items (lag). Half the words are repeated with the same surface characteristics and half with different surface characteristics at each lag. Responses to words repeated at shorter lags are assumed to reflect short-term memory processes whereas responses to words repeated at longer lags are assumed to reflect long-term memory. If repetition with same versus different surface characteristics affected items at short lags, this outcome would suggest that stimulus variability affects the rehearsal and encoding of spoken words. If repetition of same versus different surface form affected items at long lags, the results would suggest that stimulus variability is retained in long-term memory and used during spoken word recognition. Further, the design of the present experiment allowed us to compare and contrast the effects of each source of stimulus variability in order to determine whether the retention of different aspects of surface form varies as a function of time.

Finally, the fourth issue we sought to address in this paper was a comparison of implicit and explicit judgments of surface form. By comparing the effects of stimulus variability on recognition memory for word identity alone with its effects on explicit recognition memory for surface form in a second experiment, we hoped to tease apart the retention of surface characteristics in memory from the use of those surface characteristics during word recognition. That is, we wanted to determine whether information about talker, rate, and amplitude could possibly be retained without affecting recognition memory for word identity. For instance, although overall amplitude was predicted to have little effect on recognition memory for word identity, it is unclear whether variability in overall amplitude might be recognized explicitly. By asking listeners if words were repeated at the same or different amplitudes, we were able to determine whether explicit recognition memory for surface form differs from recognition memory for word identity.

In sum, the purpose of the present study was to further investigate the role of different sources of stimulus variability in the encoding of spoken words in memory. By comparing the effects of talker, rate, and amplitude variability in a continuous recognition memory task, we hoped to develop a more comprehensive understanding of the effects of different item-specific features on speech perception and spoken word recognition.

## EXPERIMENT 1

Experiment 1 investigated whether listeners were more accurate at recognizing a word as “old” (i.e., whether the word had occurred previously in a list of spoken words) if it was repeated by the same talker (Condition 1), at the same speaking rate (Condition 2), or at the same overall amplitude (Condition 3). The talker condition was a replication of Palmeri et

al. (1993). The rate and amplitude conditions were designed to extend the findings on talker to conditions in which the stimuli incorporated other sources of variability.

On the basis of previous research, talker variability was expected to affect recognition memory for spoken words. Items repeated by the same talker were expected to be better recognized than items repeated by a different talker. Likewise, due to its effects on the perception of spoken words, speaking rate was expected to affect recognition memory for spoken words. Although Nygaard et al. (1995) found little evidence of effects of speaking rate on serial recall, recognition memory was expected to be more sensitive to changes in speaking rate. Words repeated at the same speaking rate were expected to be better recognized than words repeated at a different speaking rate. Finally, overall amplitude was expected to have little effect on recognition memory for word identity. Overall amplitude was manipulated by scaling the signal presentation levels up or down over a 25-dB range. This manipulation was assumed to be irrelevant with respect to perceiving the linguistic content of the signal. In previous experiments, overall amplitude has not been shown to affect perceptual processing of spoken words. Thus, words repeated at different overall amplitudes were expected to be recognized as well as words repeated at the same (constant) overall amplitude.

## Method

**Listeners**—One hundred and twenty students enrolled in undergraduate introductory psychology courses at Indiana University served as listeners. All listeners received partial course credit for their participation. All were native speakers of American English who reported no history of speech or hearing disorder at the time of testing.

**Stimuli**—The stimuli used in Experiment 1 came from a digital database of 200 monosyllabic words spoken by two talkers (one male and one female) at three different rates of speech (fast, medium, and slow). The words were selected from four 50-item phonetically balanced (PB) word lists (American National Standards Institute, 1971) and were originally recorded embedded in the carrier sentence, “Please say the word \_\_\_\_\_.” For each rate of speech, the full set of 200 sentences was presented to the talkers in random order on a CRT screen located in a sound-attenuated booth (IAC 401A). Speakers were simply asked to read each sentence aloud at a fast, medium, or slow rate of speech. Productions were monitored by an experimenter via a loudspeaker located outside the recording booth so that the mispronounced sentences could be noted and re-recorded.

The stimuli were transduced with a Shure (SM98) microphone and digitized on-line in real time via a 12-bit analogue-to-digital converter (DT2801) at a sampling rate of 10 kHz. The stimuli were then low-pass filtered at 4.8 kHz and the target words were digitally edited from the carrier sentences. The average root mean square amplitude of each of the stimuli was equated using a signal processing software package (Luce & Carrell, 1981). In order to create different presentation levels for the amplitude condition (Condition 3), high- and low-amplitude versions of the medium rate tokens from each of the two talkers were created. These tokens were generated by setting the maximum waveform amplitude level to a specified value. The remaining amplitude values in the digital files were then rescaled relative to this specified maximum so that relative amplitude differences in the signal were preserved. For the high- and low-amplitude sets, the maximum amplitude values were set at 60 dB SPL and 35 dB SPL, respectively. All other stimuli were leveled at 50 dB SPL.

For each of the three stimulus conditions (talker, rate, and amplitude), word lists were constructed in which each test word was presented and then repeated once after a lag of 2, 8, 16, or 32 intervening items. The test word itself counted as an intervening item. Each list began with 15 practice trials, which were used to familiarize the listeners with the test

procedure. Practice trials consisted of several same and different surface form repetitions, depending, of course, on condition. None of these 15 words was repeated in the experiment. The next 30 trials were used to establish a memory load and were not used in the final data analyses. A memory load was used in an attempt to equate performance on stimulus pairs occurring early in the list with pairs occurring later in the list. Thus, by discarding the first 30 trials, we were able to evaluate memory performance in listeners whose memory “buffers” were already full. The actual test list consisted of 144 test word pairs. Twenty-one filler items were interspersed in the test list and were not included in any analyses. The test pairs were distributed evenly across the four lags, with half of the repetitions at each lag having the same talker, rate, or amplitude and half having a different talker, rate, or amplitude as the original presentation of the test word. The total number of items in each list was 354. Eight separate randomizations were created, resulting in eight separate lists for each condition. Listeners were assigned randomly to the eight lists. For the final analyses, data were collapsed across randomizations.

For all three conditions, the lag between the first and second repetition of a word was manipulated as a within-subjects variable (2, 8, 16, or 32 words). However, experimental condition itself (whether talker, rate, or amplitude characteristics were repeated) was manipulated as a between-subjects variable. For the talker condition (Condition 1), only medium rate tokens were used to construct the lists. Two talkers were used, a male and a female, and words were repeated by either the same or a different talker. Thus, the talker for the second repetition of the target words was a within-subjects variable. Forty-two listeners participated in Condition 1.

For the rate condition (Condition 2), only the fast and slow rate tokens from both talkers were used. For this condition, two sets of lists were constructed—one using tokens produced by the male talker and one using tokens produced by the female talker. We used two different talkers as a control to ensure that any observed rate effects would not be specific to a particular talker's productions. Within a list, however, the talker did not vary. Thus, talker was a between-subjects variable, with half the listeners responding to tokens produced by the male talker ( $n = 20$ ) and half responding to tokens produced by the female talker ( $n = 20$ ). The speaking rate of the second repetition of the target words was a within-subjects variable (same vs. different rate).

Finally, for the amplitude condition (Condition 3), only the medium rate tokens from both talkers were used. As for the rate condition, two sets of lists were constructed—one using the rescaled items from the male talker and one using the rescaled items from the female talker. Again, we wanted to control for potential differences between talkers' amplitude tokens. However, within a list, the talker did not vary. Thus, talker was a between-subjects variable, with half the listeners responding to tokens produced by the male talker ( $n = 19$ ) and half responding to tokens produced by the female talker ( $n = 19$ ). The overall amplitude of the second repetition of the target words was a within-subjects variable (same vs. different amplitude).

**Procedure**—Listeners were tested in groups of 5 or fewer in a quiet room used for speech perception experiments. The presentation of stimuli and collection of responses were controlled on-line by a PDP-11/34 computer. Each digital stimulus was output using a 12-bit digital-to-analogue converter and was low-pass filtered at 4.8 kHz. The stimuli were presented binaurally over matched and calibrated headphones (TDH-39) at a comfortable listening level. On each trial, listeners heard a spoken word and had up to 5 sec to enter a response of “old” (i.e., the word had appeared previously in the list of spoken words) or “new” (i.e., the word was new to the list). Listeners entered their responses on appropriately labeled two-button response boxes that were interfaced to the computer. If no response was



entered after 5 sec, that trial was not recorded and the program proceeded to the next trial. No feedback was provided. The entire session of 354 trials lasted approximated 25–35 min.

## Results and Discussion

Figure 1 shows the item recognition accuracies (percent correct “old” responses) for the same-talker and different-talker repetitions (Figure 1a), same-rate and different-rate repetitions (Figure 1b), and same-amplitude and different-amplitude repetitions (Figure 1c) as a function of lag. For the talker condition, a two-factor analysis of variance (ANOVA) with lag (2, 8, 16, 32) and repetition (same talker vs. different talker) as factors showed significant main effects for both factors. Accuracy decreased with increasing lag [ $F(3,328) = 24.518, p < .0001$ ], and same-talker repetitions were recognized better overall than were different-talker repetitions [ $F(1,328) = 5.516, p < .0194$ ]. The two-way interaction was not significant. This result replicates the previous findings of Palmeri et al. (1993), who showed a same-talker advantage for recognizing a word as a repeated item without any explicit instructions to the listeners to attend to the identity of the talker.

For the rate condition, a three-factor repeated measures ANOVA with lag (2, 8, 16, 32), repetition (same-rate vs. different-rate), and talker (male vs. female) as factors showed significant main effects for lag and repetition but not for talker (indicating no difference in recognition memory for words spoken by a male or a female talker). Accuracy decreased with increasing lag [ $F(3,152) = 17.057, p < .0001$ ], and same-rate repetitions were better recognized than different-rate repetitions [ $F(1,152) = 39.895, p < .0001$ ]. There was no main effect of talker [ $F(1,152) = .323, p = .5708$ ], and none of the interactions were significant, indicating that regardless of the talker, there were consistent and reliable effects of lag and repetition. This finding extends the same-talker advantage in recognition memory found by Palmeri et al. (1993) to a different source of variability in speech, and thus demonstrates that both talker and rate information are encoded in memory along with the symbolic/linguistic information about a spoken word.

For the amplitude condition, a three-factor repeated measures ANOVA with lag (2, 8, 16, 32), repetition (same-amplitude vs. different-amplitude), and talker (male vs. female) as factors showed significant main effects for lag and talker. Accuracy decreased with increasing lag [ $F(3,144) = 38.474, p < .0001$ ], and accuracy was generally higher for the male talker than for the female talker [ $F(1,144) = 4.319, p < .0395$ ]. However, there was no main effect of amplitude repetition, and none of the interactions were significant. Thus, whereas recognition accuracy decreased with increasing lag, there was no difference in recognition accuracy between the same-amplitude and different-amplitude trials. Furthermore, this pattern of results was obtained for both talkers even though the overall accuracy scores for the male talker were slightly higher than for the female talker (91.2% and 88.9% correct item recognition, respectively). The fact that there was no same-amplitude advantage in recognition memory for both talkers suggests that overall amplitude information may not be a source of variability in speech that is encoded into long-term memory in the same way as talker and rate information, and that different item-specific stimulus characteristics can have distinct effects on speech perception and spoken word recognition.

In order to compare the overall level of discrimination between old and new items across the three conditions, we computed  $d'$  scores for each listener in each condition. The mean  $d'$  score in all three conditions was significantly greater than zero ( $p < .0001$  in all three conditions by a one-sample  $t$  test), indicating good discrimination in all conditions (see Table 1). Furthermore, a one-factor ANOVA with condition as the factor showed a significant main effect of condition [ $F(2,117) = 5.198, p < .007$ ]. Post hoc comparisons (Fisher's PLSD) showed a significant difference in  $d'$  for the talker and rate conditions ( $p < .$

002) and for the rate and amplitude conditions ( $p < .039$ ). However, there was no difference in  $d'$  for the talker and amplitude conditions. These analyses suggest that speaking rate differences resulted in poorer old versus new discrimination overall than did talker and overall amplitude. In particular, listeners appeared to make substantially more errors when speaking rate differed across repetitions than when either talker or overall amplitude differed across repetitions. That is, speaking rate appeared to have a larger effect on repetition accuracy than did talker or overall amplitude. This same pattern of results was observed in a comparison of the difference scores (% correct old recognition for *same* trials—% correct old recognition for *different* trials) across each of the three conditions. A two-factor ANOVA on these difference scores showed a significant main effect of condition [ $F(2,351) = 21.080, p < .001$ ], but no main effect of lag. Post hoc comparisons (Fisher's PLSD) showed significantly larger difference scores for the rate condition than for either of the other two conditions, suggesting that rate variation had a larger effect on repetition accuracy than did either talker or amplitude variation.

In summary, as expected, the results of Experiment 1 demonstrate that same-talker trials were recognized better than different-talker trials across lags. As previously shown, talker information appears to influence both short- and long-term retention of spoken words. More importantly, we found that same-speaking-rate trials were recognized better than different-speaking-rate trials across lags. These findings constitute one of the first demonstrations that an intratalker source of variation appears to be retained in long-term lexical representations (see also Church & Schacter, 1994, for findings with intonation and fundamental frequency). These results suggest that the memory representations for spoken words preserve the detailed changes in speaking rate, and that repetition of these details can influence recognition of spoken words both in short-term (short lags) and long-term (long lags) memory.

In contrast to the effects of talker and rate variability, no difference was found in recognition memory for same- and different-amplitude trials. Thus, information about the talker and speaking rate appeared to affect both short-term processing of word identity and long-term memory retention, whereas no evidence was found that information about the overall amplitude of a spoken word affected either short- or long-term retention of word identity.

The finding that sources of stimulus variation such as talker and speaking rate are encoded and retained in long-term memory has several important implications for abstractionist theories of lexical representation. According to abstractionist approaches, if surface form is discarded during the process of spoken word recognition, then surface characteristics of individual words should not necessarily affect recognition memory. However, effects of talker and speaking rate on recognition accuracy suggest that the representations used to perform this task include access to considerable perceptual detail. Not only are individual talker characteristics preserved in memory, as has previously been shown, but also, a form of intratalker variation, speaking rate, is preserved.

The finding that all sources of variation may not be preserved in long-term memory to the same extent has implications for strong versions of episodic or exemplar-based accounts of lexical representation (see Goldinger, 1996; Pisoni, 1997). According to exemplar-based theories of word recognition, if lexical representations are based on collections of episodic traces that preserve perceptual detail, then all salient details of surface form should be included in memory representations. However, Experiment 1 shows that only certain types of surface form appear to be preserved, whereas other stimulus variations such as overall amplitude do not appear to affect recognition memory. One explanation may be, as Sommers et al. (1994) and Nygaard et al. (1995) have suggested, that only sources of variability that are linguistically relevant are retained in long-term memory representations

of spoken words. If this is so, episodic theories may need to include an attention mechanism that selectively represents only salient or relevant aspects of the surface form of spoken words.

The possibility remains, of course, that overall amplitude information may be retained in memory, but that when listeners are instructed to explicitly recognize the item as “new” or “old” they are unable to use this information as an implicit retrieval cue in this task. In order to evaluate this alternative, a second experiment was carried out.

## EXPERIMENT 2

This experiment was designed to investigate whether listeners can explicitly recognize changes in talker, rate, and amplitude for a repeated word. Whereas in Experiment 1, listeners were not required to pay explicit attention to the talker, rate, or amplitude of the test item, in Experiment 2, listeners were required to make an explicit judgment regarding a change in talker, rate, or amplitude. We hypothesized that this task would provide a more direct test of the extent to which detailed information about the instance-specific characteristics of a spoken word are encoded in long-term memory. Specifically, we were interested in investigating the possibility that listeners are able to detect and encode changes in overall amplitude even though overall amplitude did not affect item recognition accuracy in Experiment 1.

On the basis of previous research by Craik and Kirsner (1974) and Palmeri et al. (1993), we expected that listeners would be able to explicitly judge whether a word was repeated by the same or a different talker. Both studies showed that listeners were able to judge talker repetitions over lags of up to 32 intervening items. Likewise, we predicted that listeners would be accurate at explicitly judging same- versus different-speaking-rate repetitions. On the basis of the effects of speaking rate in the first experiment, we predicted that explicit judgments of rate would present no problem to our listeners. Finally, we predicted that listeners would not be able to judge whether words were repeated at the same or a different overall amplitude. Little effect of overall amplitude has been found in a variety of perceptual and memory tasks, so it was assumed that no effects would be uncovered in the present paradigm. However, the experiment was designed to test the possibility that overall amplitude was retained in memory to some extent, so we hypothesized that overall amplitude might have an effect with different task demands. Thus, it might be that when listeners are specifically asked to attend to overall amplitude, evidence for the retention of amplitude in long-term memory might be found.

### Method

**Listeners**—One hundred and nineteen students enrolled in undergraduate introductory psychology courses at Indiana University served as listeners. All listeners received partial course credit for their participation. All were native speakers of American English and reported no history of speech or hearing disorder at the time of testing. None of the listeners used in this experiment had participated in the previous experiment.

**Stimuli and Procedure**—The stimulus materials for Experiment 2 were identical to those used in Experiment 1. All aspects of the stimulus presentation and test conditions were identical except that in this experiment, listeners were given three response categories rather than two. In Experiment 2, after hearing the spoken word, listeners had 5 sec to identify the word as “new” if it had not occurred in the list before, as “old–same” if it had occurred before and was repeated by the same talker (Condition 1), rate (Condition 2), or amplitude (Condition 3), or as “old–different” if it was repeated by a different talker (Condition 1), rate (Condition 2), or amplitude (Condition 3). Thus, in Experiment 2, in addition to recognizing

a word as old or new, listeners were also required to make an explicit judgment for the items recognized as old regarding whether the talker, rate, or amplitude changed from the first to second repetition of the word. A group of 33 listeners participated in the talker condition. For the rate condition, a group of 21 listeners was tested on stimuli spoken by the male talker, and a separate group of 21 listeners was tested on stimuli spoken by the female talker. For the amplitude condition, a separate group of 22 listeners was tested on each of the two stimulus sets (one from the male talker, one from the female talker).

**Results and Discussion**—Figure 2 shows the overall percentage of correct old item recognition responses for the talker condition (Figure 2a), the rate condition (Figure 2b), and the amplitude condition (Figure 2c). The accuracy scores shown in this figure represent all cases of correct old item recognition regardless of accuracy on the *same–different* judgment of surface form. This analysis allowed us to compare the pattern of results on the item recognition task across Experiments 1 and 2.

As shown in Figure 2, same-talker trials were recognized better than different-talker trials and same-rate trials were recognized better than different-rate trials, but there was no difference in recognition accuracy for same- and different-amplitude trials. This pattern of results is consistent with the results of Experiment 1. For the talker condition, a two-factor ANOVA with repetition (same talker or different talker) and lag (2, 8, 16, 32) as factors showed main effects of both factors. Same-talker trials were better recognized than different-talker trials [ $F(1,256) = 4.541, p = .0340$ ], and recognition accuracy decreased with increasing lags [ $F(3,256) = 13.258, p < .0001$ ]. The twoway interaction was not significant.

For the rate condition, a three-factor repeated measures ANOVA with repetition (same rate or different rate), lag (2, 8, 16, 32), and talker (male vs. female) as factors showed main effects of all three factors. Same-rate trials were better recognized than different-rate trials [ $F(1,160) = 26.973, p < .0001$ ], recognition accuracy decreased with increasing lags [ $F(3,160) = 20.906, p < .0001$ ], and recognition accuracy was slightly better for tokens produced by the male talker than for those produced by the female talker [mean difference = 2.94%,  $F(1,160) = 4.815, p = .0297$ ]. None of the interactions involving talker as a factor was significant, indicating that the pattern of decreasing recognition accuracy with increasing lags, and across same-rate and different-rate trials, was consistent across both talkers. Similarly, the two-way interaction between repetition and lag was not significant.

For the amplitude condition, a three-factor repeated measures ANOVA with repetition (same amplitude vs. different amplitude), lag (2, 8, 16, 32), and talker (male vs. female) as factors showed a main effect of lag, but no main effects of repetition or talker. None of the interactions was significant. As expected from the results of Experiment 1, recognition accuracy decreased with increasing lags [ $F(3,168) = 48.820, p < .0001$ ], but there was no same-amplitude advantage relative to different-amplitude trials. This pattern of results replicates the main findings of Experiment 1 by providing evidence that information regarding the talker and rate of speech is encoded in long-term memory along with the symbolic linguistic information about a spoken word. In contrast, once again, we found no evidence that information about overall amplitude was retained in long-term memory.

The similarity between the patterns of item recognition accuracy scores for the two experiments indicates that requiring an additional response for Experiment 2 did not alter the main effects of lag and repetition on item recognition accuracy. In order to assess directly the effect of the additional response category, separate repeated measures ANOVAs for each of the three conditions with experiment (1 or 2) as the repeated measure were performed. For the talker condition, the analysis showed the expected main effects of lag [ $F(3,256) = 33.364, p < .0001$ ] and repetition [ $F(1,256) = 8.552, p = .0038$ ]. The two-way interaction

between lag and repetition was not significant. There was also a significant main effect of experiment [ $F(1,256) = 12.059, p = .0006$ ] due to generally higher accuracies for Experiment 1 than for Experiment 2 (means = 88.55% and 84.36%, respectively). None of the interactions involving the experiment factor were significant, indicating that the patterns of decreasing accuracy with increasing lag, and of higher accuracy for same-talker repetitions, were consistent across both experiments.

For the rate condition, main effects of lag [ $F(3,344) = 40.025, p < .0001$ ] and repetition [ $F(1,344) = 73.220, p < .0001$ ] were observed, but there was no effect of experiment and none of the interactions were significant. Finally, for the amplitude condition, the main effect of lag was significant [ $F(3,304) = 76.150, p < .0001$ ], as well as the main effect of experiment [ $F(1,304) = 7.398, p = .0069$ ], but there was no main effect of repetition. As for the talker condition, the effect of experiment for the amplitude condition was due to generally higher accuracies for Experiment 1 than for Experiment 2 (means = 90.21% and 87.82%, respectively). Thus, requiring the listeners to make an additional response in Experiment 2 resulted in slightly lower overall recognition accuracy scores for the talker and amplitude conditions. However, across all three conditions, the same general pattern of results for the two experiments was consistent in showing a same-talker and same-rate advantage relative to different-talker and different-rate trials, respectively. Similarly, both experiments failed to reveal a same-amplitude advantage relative to different-amplitude trials.

Finally, as in Experiment 1, a comparison of the difference scores (% correct old recognition for *same* trials—% correct old recognition for *different* trials) across each of the three conditions showed significantly greater difference scores for the rate condition than for either of the other two conditions. A two-factor ANOVA on these difference scores showed a significant main effect of condition [ $F(2,348) = 131.137, p < .001$ ], but no main effect of lag. Post hoc comparisons (Fisher's PLSD) confirmed that rate variation had a larger effect on repetition accuracy than either talker or amplitude variation.

In order to determine whether listeners can explicitly recognize variation in talker, rate, and amplitude for items that were correctly identified as old,  $d'$  scores were calculated for each condition at each lag. In this analysis, a *hit* was defined as a response of “old/same” to a stimulus that was repeated by the same talker, rate, or amplitude. A *false alarm* was defined as a response of “old/same” to a stimulus that was repeated by a different talker, rate, or amplitude. Using this measure, we were able to determine whether listeners can discriminate changes in talker, rate, and amplitude, and thus establish whether detailed information along each of these stimulus dimensions was encoded and retained in memory.

Figure 3 shows the  $d'$  scores for all three conditions as a function of lag. Two main findings are shown here. First, for all three conditions at all lags, the  $d'$  scores differed significantly from zero, indicating that listeners were able to discriminate “old/same” from “old/different” trials in all cases. One-sample  $t$  tests for each condition at each lag confirmed that these  $d'$  scores were all significantly different from zero at the  $p < .0001$  level. This finding suggests that, regardless of whether the instance-specific information affected recognition memory accuracy for items in the “old–new” task, listeners do retain highly detailed information in memory to the extent that variability along each of the three dimensions was explicitly detected. Second, variability along each of the three dimensions was discriminated with a different degree of accuracy: Talker variability was detected better than rate variability, which was detected better than amplitude variability. A two-factor ANOVA with condition (talker, rate, amplitude) and lag (2, 8, 16, 32) as factors showed main effects for both factors [condition,  $F(2,476) = 45.459, p < .0001$ ; lag,  $F(3,476) = 110.988, p < .0001$ ]. The two-way interaction was also significant [ $F(6,476) = 3.264, p = .0037$ ]. This finding suggests that,

although fine details of the stimulus dimensions are retained in memory, certain stimulus dimensions represent more perceptually salient characteristics than others and thus may produce more substantial effects on speech perception and spoken word recognition performance in different tasks.

The main goal of Experiment 2 was to investigate whether listeners were able to explicitly discriminate changes in talker, rate, or amplitude for items that they had correctly recognized as repeated items (i.e., old items). In particular, we were interested in the results of this task for the amplitude condition, in which differences in amplitude information did not affect recognition memory performance. The results showed that listeners were indeed able to explicitly detect changes in talker, rate, and amplitude. Thus, this task provided evidence that, even though all sources of variability do not function identically with respect to spoken word recognition, detailed stimulus information about the instance-specific characteristics of a spoken word is retained in memory along with the more abstract symbolic linguistic content of the word. These highly detailed memory representations even include information along an apparently linguistically irrelevant dimension, such as overall amplitude.

## GENERAL DISCUSSION

The goal of this study was to investigate the extent to which the neural representation of spoken words in memory encodes detailed, instance-specific information. The results that emerged from this study complement and extend the findings of earlier studies that have investigated the effects of talker, rate, and amplitude variability on speech perception and memory for spoken words. The general pattern of results that has emerged from this set of experiments (summarized in Table 2) suggests that detailed stimulus information about all sources of variability is retained to some degree in long-term memory. However, the extent to which different sources of variability are retained seems to depend on the specific source of stimulus variability as well as the task and encoding conditions.

More specifically, a comparison of the effects of talker, rate, and amplitude variability on the tasks listed in Table 2 reveals a hierarchy in which amplitude, rate, and talker variability have increasingly stronger effects on speech perception and memory for spoken words. The relatively weak effect of amplitude variability is seen by the fact that experiments using all three tasks (word identification, serial recall, and continuous recognition) failed to show an effect of trial-to-trial changes in signal level. In fact, the only evidence that overall amplitude information is retained in long-term memory comes from the task in which listeners were specifically asked to *explicitly* identify variability along this dimension (present study, Experiment 2).

In contrast, the stronger effect of speaking rate variability was evident across all three tasks. In terms of perception and encoding, trial-to-trial changes in speaking rate resulted in decreased performance relative to trials with no change in speaking rate. For instance, word lists in which each word was spoken at a constant speaking rate were better identified when embedded in noise than in identical lists spoken with multiple speaking rates (Sommers et al., 1994). Similarly, at fast presentations rates, single-speaking-rate word lists were more accurately recalled than multiple-speaking-rate lists (Nygaard et al., 1995). Finally, the present experiments demonstrated that speaking rate is also retained in long-term memory representations. Words repeated at the same speaking rate were better recognized in a continuous recognition memory task than were words repeated at a different speaking rate (present study, Experiment 2).

Overall, the effects of rate variability are comparable to the effects of talker variability. However, a difference between the two sources of variability did emerge in the serial recall

task with long interstimulus intervals (ISIs) (Nygaard et al., 1995). When listeners were given enough time, information about the talker's voice was apparently encoded in long-term memory representation of the spoken words and thus served as a distinctive identifying feature of the words. In this manner, the talker's voice functioned as a retrieval cue and aided the listener in the serial recall task to the extent that multiple-talker lists were *better* recalled than single-talker lists. In contrast, at long ISIs, the detrimental effect of multiple speaking rates was diminished only to the extent that multiple-rate lists were recalled as well as single-rate lists. Thus, one way in which to describe the overall pattern of results across experiments and studies is to conclude that talker, speaking rate, and amplitude constitute a hierarchy of effects on speech perception and memory for spoken words, with talker variability having the most pervasive effects, speaking rate variability having intermediate effects, and amplitude variability having the weakest effects.

At this point, we can speculate as to the mechanism that underlies the effects for these different sources of variability. It is possible that the differences in the effects of talker, rate, and amplitude variability reflect differences in the complexity of the acoustic correlates of changes along these dimensions. In all of the experiments summarized in Table 2 that investigated the effects of amplitude variability, a change in amplitude was achieved by simply setting the maximum level for each waveform to a specified value and then rescaling the remaining amplitude levels relative to that maximum. Thus, amplitude variability was a constant, unidimensional adjustment related to gain. In contrast, rate variability was more naturally achieved and was thus variable and multidimensional in its acoustic correlates. Rate variability within a given speaker is not achieved by a constant “stretching” or “shrinking” of the acoustic waveform in the temporal domain. Rather, certain acoustic segments are more dramatically reduced in duration than others when overall speaking rate is increased, and various other acoustic/phonetic changes (e.g., vowel reduction) occur in response to changes in speaking rate (see, e.g., Klatt, 1973, 1976; Lehiste, 1972; Picheny, Durlach, & Braida, 1986, 1989; Port, 1981; Uchanski, Choi, Braida, Reed, & Durlach, 1996). Thus, an increase or decrease in speaking rate is clearly a dynamic, multidimensional transformation of the speech signal. Similarly, a change in talker leads to a wide variety of acoustic/phonetic changes. Not only do talkers differ in vocal tract shape and size, which leads to different spectrotemporal characteristics, but talkers also differ in articulatory “style” (including speaking rate, dialect, and other idiosyncratic differences) which can lead to large differences in the acoustic waveform of a given word across various talkers (see, e.g., Fant, 1973; Joos, 1948; Peterson & Barney, 1952).

Thus, the varying degrees to which talker, rate, and amplitude variability affect speech perception and memory for spoken words appear to be directly related to the complexity of the acoustic correlates that result from these sources of variability. From the listener's point of view, then, it is possible that the simpler the acoustic transformation related to a given source of variability, the fewer the processing resources required to compensate for that variability, and consequently the lower the impact of this variability on speech perception and memory for spoken words. Certainly, this explanation would be consistent with theories proposing abstract representations. Rather than speculating that talker and rate characteristics are preserved in memory representations per se, effects of variation on recognition memory could be due to the retention of the compensatory procedures that are used to abstract the linguistic identity of each word (Kolers, 1976; Kolers & Ostry, 1974). Thus, speaking rate and talker might have greater effects on memory performance than overall amplitude simply because these sources of variability require more extensive processing operations at the time of initial encoding that, once learned, would more greatly facilitate processing when repeated. Thus, the difference between recognition of words repeated with same versus different surface characteristics would be the result of the

retention in procedural memory for the normalization procedures specific to each source of variability.

Another explanation for the differential effects of each source of variability on speech perception and memory for spoken words takes into account the relevance of each source of variability for the perception of phonetic contrasts. Variability in talker characteristics has been shown to have a significant impact on speech perception. For example, Ladefoged and Broadbent (1957) found that vowel identification could be altered depending on the perceived talker characteristics of a precursor phrase, and Johnson (1990) showed that perceived speaker identity plays an important role in the  $F_0$  normalization of vowels. Similarly, several studies have demonstrated the rate dependency of phonetic processing for both vowels and consonants (e.g., Miller, 1987; Miller & Volaitis, 1989; Port, 1981; Summerfield, 1981). In contrast, overall amplitude variability does not, by itself, signal phonetic contrasts, and there does not appear to be an amplitude dependency in speech perception that is comparable to talker- and rate-dependent phonetic processing. Thus, it is possible that the observed differences in perception and memory for spoken words as a function of talker and rate variability, on the one hand, and amplitude variability, on the other, are due to differences in their phonetic relevance to the listener.

An additional consideration, however, emerges from the differences among sources of variability in Experiments 1 and 2 of the present study. Interestingly, the extent to which each source of variability affected listeners' performance differed depending on the task. For example, although listeners in Experiment 2 were poorer at recognizing repeated speaking rate characteristics than repeated talker characteristics, speaking rate had larger effects than did talker on recognition memory for word identity in Experiments 1 and 2. Similarly, overall amplitude had virtually no effect on the recognition of word identity, but was at least modestly recognizable when listeners were explicitly asked to attend to overall amplitude. It is possible that differences in the retention and use of different types of surface form may result not only from intrinsic differences in saliency or relevance of each surface form, but also from the extent to which unique aspects of each surface form interact with particular task demands. That is, perhaps the extent to which a given source of variability affects memory is tied to the extent to which a particular task requires attention to that source of variation.

The explanations suggested earlier concerning linguistic relevance and task constraints are consistent with recent exemplar-based approaches to lexical representation (Goldinger, 1996). Although a strong version of this view would predict encoding and retention of all aspects of surface form, differential effects could be found if lexical traces were assumed to include only some selected aspects of surface form. That is, representations of individual instances might not represent every perceptual event veridically; rather, some surface details might be included in lexical instance-based representations based on the saliency or relevance of individual dimensions to the linguistic episode and/or on the extent to which the task focuses attention on a particular aspect of surface form.

Of course, a wider range of sources of variability and perceptual and memory tasks needs to be investigated in order to provide conclusive evidence for the alternative explanations for the effects of different sources of stimulus variability. For example, it might be enlightening to investigate the effects of variations in dialect, vocal effort, speaking style, emotional state, and other such paralinguistic or indexical characteristics of speech, as well as the effects of nonlinguistic factors such as filtering characteristics due to different microphones or recording conditions. Similarly, a variety of other perceptual tasks need to be studied to determine how the relevance and/or saliency of a particular surface form might interact with attentional and task constraints. The present results suggest that all instance-specific



stimulus attributes are encoded and retained in memory to the extent that listeners are able to detect such changes. There is now a rapidly growing body of converging evidence demonstrating that the processes of speech perception and spoken word recognition operate in the context of highly detailed representations of the acoustic speech signal, rather than on idealized abstract symbolic representations of abstract linguistic information. We believe these are important new observations about speech and spoken language processing that have broad implications for future research and theory about speech perception, word recognition, and lexical access.

## Acknowledgments

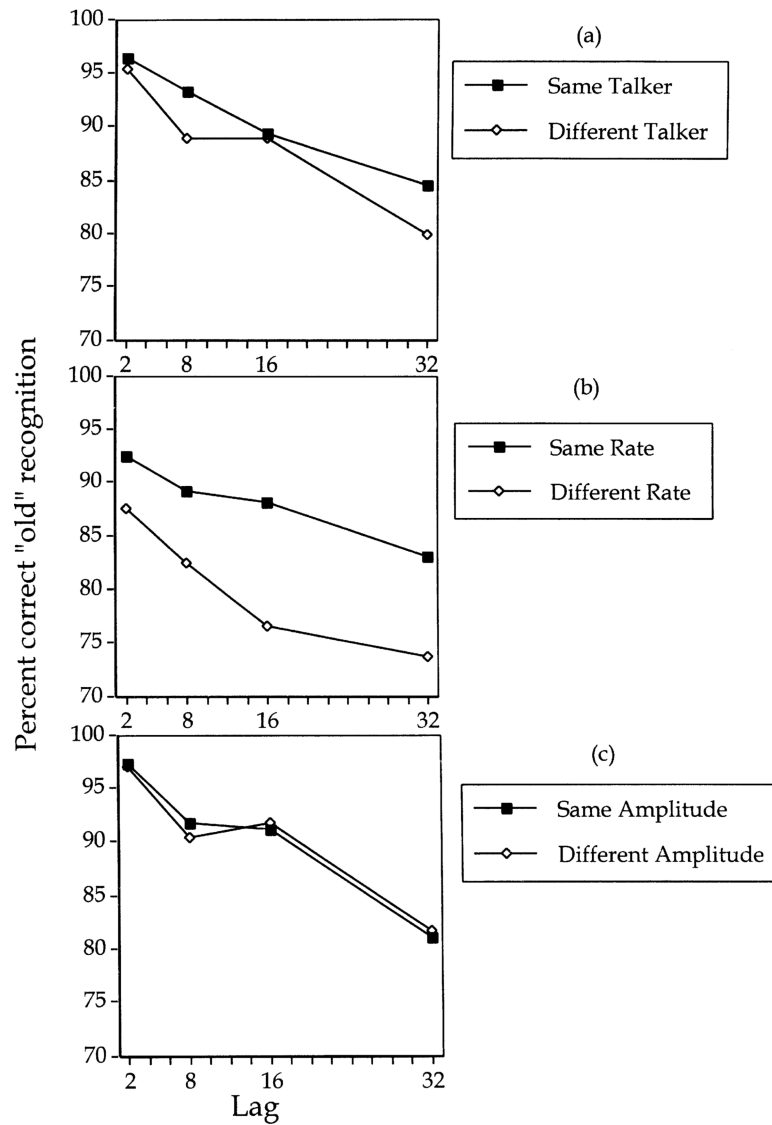
This research was supported by NIDCD Training Grant DC-00012 and NIDCD Research Grant DC-00111 to Indiana University. We are grateful to Luis Hernandez for technical support and to Thomas Palmeri for programming assistance.

## REFERENCES

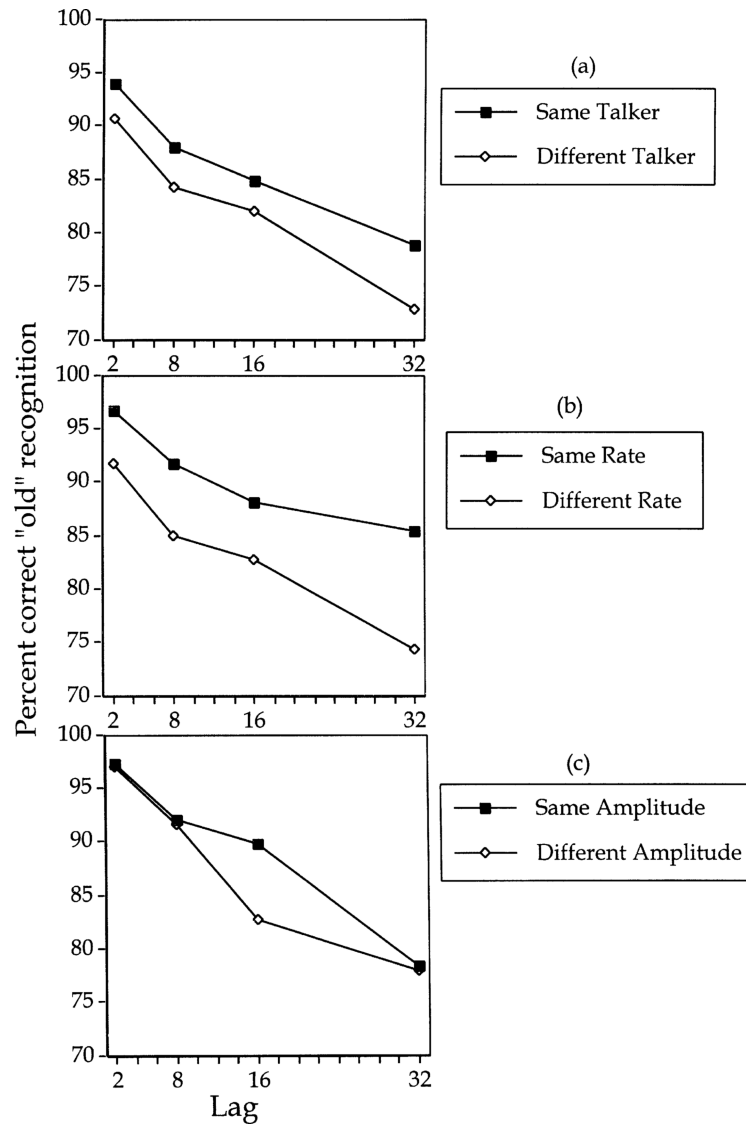
- American National Standards Institute. Method for measurement of monosyllabic word intelligibility. Author; New York: 1971. (American National Standards S3.2-1960 [R1971])
- Church BA, Schacter DL. Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, & Cognition*. 1994; 20:521–533.
- Cole R, Coltheart M, Allard F. Memory of a speaker's voice: Reaction time to same- or different-voiced letters. *Quarterly Journal of Experimental Psychology*. 1974; 26:274–284.
- Craik FIM, Kirsner K. The effect of speaker's voice on word recognition. *Quarterly Journal of Experimental Psychology*. 1974; 26:274–284.
- Creelman CD. Case of the unknown talker. *Journal of the Acoustical Society of America*. 1957; 29:655.
- Eich JM. A composite holographic associative recall model. *Psychological Review*. 1982; 89:627–661.
- Fant, G. *Speech sounds and features*. MIT Press; Cambridge, MA: 1973.
- Garner, WR. *The processing of information and structure*. Erlbaum; Potomac, MD: 1974.
- Geiselman RE. Inhibition of the automatic storage of speaker's voice. *Memory & Cognition*. 1979; 7:201–204.
- Geiselman RE, Bellezza FS. Long-term memory for speaker's voice and source location. *Memory & Cognition*. 1976; 4:483–489.
- Geiselman RE, Bellezza FS. Incidental retention of speaker's voice. *Memory & Cognition*. 1977; 5:658–665.
- Geiselman RE, Crawley JM. Incidental processing of speaker characteristics: Voice as connotative information. *Journal of Verbal Learning & Verbal Behavior*. 1983; 22:15–23.
- Gerstman LJ. Classification of self-normalized vowels. *IEEE Transactions on Audio & Electroacoustics*. 1968; 16:78–80.
- Goldinger SD. Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Human Perception & Performance*. 1996; 22:1166–1183.
- Goldinger SD, Pisoni DB, Logan JS. On the nature of talker variability effects on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition*. 1991; 17:152–162.
- Halle, M. Speculations about the representation of words in memory.. In: Fromkin, VA., editor. *Phonetic linguistics*. Academic Press; New York: 1985. p. 101-114.
- Hintzman DL. "Schema abstraction" in a multiple-trace memory model. *Psychological Review*. 1986; 93:411–428.
- Johnson K. The role of perceived speaker identity in *F0* normalization of vowels. *Journal of the Acoustical Society of America*. 1990; 88:642–654. [PubMed: 2212287]

- Joos MA. Acoustic phonetics. *Language*. 1948; 24:1–136.
- Kewley-Port D. Time-varying features as correlates of place of articulation in stop consonants. *Journal of the Acoustical Society of America*. 1983; 73:322–335. [PubMed: 6826902]
- Klatt DH. Interaction between two factors that influence vowel duration. *Journal of the Acoustical Society of America*. 1973; 54:1102–1104. [PubMed: 4757455]
- Klatt DH. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*. 1976; 59:1208–1221. [PubMed: 956516]
- Kolers PA. Reading a year later. *Journal of Experimental Psychology: Human Learning & Memory*. 1976; 2:554–565.
- Kolers PA, Ostry D. Time course of loss of information regarding pattern analyzing operations. *Journal of Verbal Learning & Verbal Behavior*. 1974; 13:599–612.
- Ladefoged P, Broadbent DE. Information conveyed by vowels. *Journal of the Acoustical Society of America*. 1957; 29:98–104.
- Laver, J. Cognitive science and speech: A framework for research.. In: Schnelle, H.; Bernsen, NO., editors. *Logic and linguistics (Research Directions in Cognitive Science: European Perspectives)*. Vol. 2. Erlbaum; Hillsdale, NJ: 1989. p. 37-70.
- Laver, J.; Trudgill, P. Phonetic and linguistic markers in speech.. In: Scherer, KR.; Giles, H., editors. *Social markers in speech*. Cambridge University Press; Cambridge: 1979. p. 1-32.
- Lehiste I. The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America*. 1972; 51:2018–2024.
- Luce, PA.; Carrell, TD. Creating and editing waveforms using WAVES. Indiana University, Speech Research Laboratory; Bloomington: 1981. (Research in Speech Perception, Progress Report No. 7)
- Martin CS, Mullennix JW, Pisoni DB, Summers WV. Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition*. 1989; 15:676–684.
- Miller, JL. Rate-dependent processing in speech perception.. In: Ellis, A., editor. *Progress in the psychology of language*. Erlbaum; Hillsdale, NJ: 1987. p. 119-157.
- Miller JL, Liberman AM. Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics*. 1979; 25:457–465. [PubMed: 492910]
- Miller JL, Volaitis LE. Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*. 1989; 46:505–512. [PubMed: 2587179]
- Mullennix JW, Pisoni DB. Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics*. 1990; 47:379–390. [PubMed: 2345691]
- Mullennix JW, Pisoni DB, Martin CS. Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America*. 1989; 85:365–378. [PubMed: 2921419]
- Nearey T. Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America*. 1989; 85:2088–2113. [PubMed: 2659638]
- Nosofsky RM. Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception & Performance*. 1991; 17:3–27. [PubMed: 1826320]
- Nygaard LC, Sommers MS, Pisoni DB. Speech perception as a talker-contingent process. *Psychological Science*. 1994; 5:42–46. [PubMed: 21526138]
- Nygaard LC, Sommers MS, Pisoni DB. Effects of stimulus variability on perception and representation of spoken words in memory. *Perception & Psychophysics*. 1995; 57:989–1001. [PubMed: 8532502]
- Palmeri TJ, Goldinger SD, Pisoni DB. Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, & Cognition*. 1993; 19:309–328.
- Peterson GE, Barney HL. Control methods used in the study of vowels. *Journal of the Acoustical Society of America*. 1952; 24:175–184.
- Picheny MA, Durlach NI, Braida LD. Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech & Hearing Research*. 1986; 29:434–446. [PubMed: 3795886]

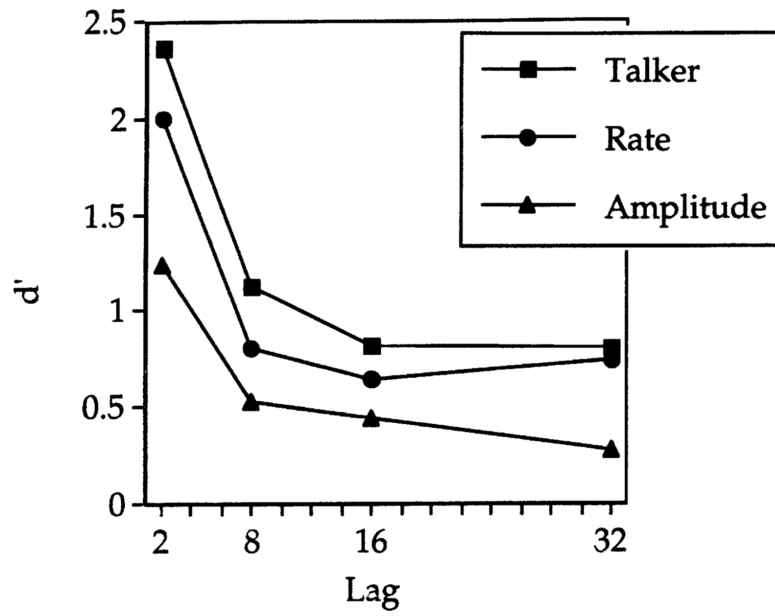
- Picheny MA, Durlach NI, Braida LD. Speaking clearly for the hard of hearing III: An attempt to determine the contribution of speaking rate to differences in intelligibility between clear and conversational speech. *Journal of Speech & Hearing Research*. 1989; 32:600–603. [PubMed: 2779204]
- Pisoni DB. Long-term memory in speech perception: Some new findings on talker variability, speaking rate and perceptual learning. *Speech Communication*. 1993; 13:109–125. [PubMed: 21461185]
- Pisoni, DB. Some thoughts on “normalization” in speech perception.. In: Mullennix, J.; Johnson, KA., editors. *Talker variability in speech processing*. Academic Press; New York: 1997. p. 9-32.
- Port RF. Linguistic timing factors in combination. *Journal of the Acoustical Society of America*. 1981; 69:262–274. [PubMed: 7217524]
- Schacter DL, Church BA. Auditory priming: Implicit and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory, & Cognition*. 1992; 18:915–930.
- Sheffert SM. Contributions of surface and conceptual information to recognition memory. *Perception & Psychophysics*. 1998a; 60:1141–1152. [PubMed: 9821776]
- Sheffert SM. Voice-specificity effects on auditory word priming. *Memory & Cognition*. 1998b; 26:591–598.
- Sheffert SM, Fowler CA. The effects of voice and visible speaker change on memory for spoken words. *Journal of Learning & Memory*. 1995; 34:665–685.
- Sommers MS, Nygaard LC, Pisoni DB. Stimulus variability and spoken word recognition. I. Effects of variability in speaking rate and overall amplitude. *Journal of the Acoustical Society of America*. 1994; 96:1314–1324. [PubMed: 7962998]
- Stevens KN. Understanding variability in speech: A requisite for advances in speech synthesis and recognition. *Journal of the Acoustical Society of America*. 1996; 100:2634.
- Stevens KN, Blumstein SE. Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America*. 1978; 64:1358–1368. [PubMed: 744836]
- Summerfield Q. On articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception & Performance*. 1981; 7:1074–1095. [PubMed: 6457109]
- Tenpenny PL. Abstractionist versus episodic theories of repetition priming and word identification. *Psychonomic Bulletin & Review*. 1995; 2:339–363.
- Uchanski RM, Choi S, Braida LM, Reed CM, Durlach NI. Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate. *Journal of Speech & Hearing Research*. 1996; 39:494–509. [PubMed: 8783129]
- Volaitis LE, Miller JL. Phonetic prototypes: Influences of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America*. 1992; 92:723–735. [PubMed: 1506527]



**Figure 1.** Item recognition accuracy scores as a function of lag from Experiment 1 for (a) the talker condition, (b) the rate condition, and (c) the amplitude condition.



**Figure 2.** Item recognition accuracy scores as a function of lag from Experiment 2 for (a) the talker condition, (b) the rate condition, and (c) the amplitude condition.



**Figure 3.**  
 $d'$  scores for all three conditions of Experiment 2 as a function of lag.

**Table 1**

## False Alarm Rates for Experiment 1

Condition	Hit Rate (%)	False Alarm Rate (%)	$d'$
Talker	81.0	11.6	2.17
Rate	80.3	16.4	1.91
Amplitude	81.5	13.5	2.08

Table 2

Summary of Findings Regarding the Effect of Talker, Rate, and Amplitude Variability on Speech Perception and Memory for Spoken Words

Source of Variation	Word Identification		Serial Recall		Recognition Memory	
	Short ISIs	Long ISIs	Item	Attribute	Item	Attribute
Talker	Single > Multi <sup>1,2</sup>	Single > Multi <sup>3,4,5</sup>	Multi > Single <sup>4,5</sup>	Same > Different <sup>6,7</sup>	Yes <sup>6,7</sup>	Yes <sup>6,7</sup>
Rate	Single > Multi <sup>2</sup>	Single > Multi <sup>5</sup>	Multi = Single <sup>5</sup>	Same > Different <sup>7</sup>	Yes <sup>7</sup>	Yes <sup>7</sup>
Amplitude	Single = Multi <sup>2</sup>	Single = Multi <sup>5</sup>	Multi = Single <sup>5</sup>	Same = Different <sup>7</sup>	Yes <sup>7</sup>	Yes <sup>7</sup>

<sup>1</sup>Mullennix et al., 1989

<sup>2</sup>Sommers et al., 1994

<sup>3</sup>Martin et al., 1989

<sup>4</sup>Goldinger et al., 1991

<sup>5</sup>Nygaard et al., 1995

<sup>6</sup>Palmeri et al., 1993

<sup>7</sup>Present study; ISI, interstimulus interval.