Journal of Experimental Psychology:
Learning, Memory, and Cognition
1991, Vol. 17, No. 1, 152-162

0278-7393/91/$3.00

On the Nature of Talker Variability Effects on Recall of Spoken Word Lists

Stephen D. Goldinger, David B. Pisoni, and John S. Logan
Indiana University

In a recent study, Martin, Mullennix, Pisoni, and Summers (1989) reported that subjects' accuracy in recalling lists of spoken words was better for words in early list positions when the words were spoken by a single talker than when they were spoken by multiple talkers. The present study was conducted to examine the nature of these effects in further detail. Accuracy of serial-ordered recall was examined for lists of words spoken by either a single talker or by multiple talkers. Half the lists contained easily recognizable words, and half contained more difficult words, according to a combined metric of word frequency, lexical neighborhood density, and neighborhood frequency. Rate of presentation was manipulated to assess the effects of both variables on rehearsal and perceptual encoding. A strong interaction was obtained between talker variability and rate of presentation. Recall of multiple-talker lists was affected much more than single-talker lists by changes in presentation rate. At slow presentation rates, words in early serial positions produced by multiple talkers were actually recalled more accurately than words produced by a single talker. No interaction was observed for word confusability and rate of presentation. The data provide support for the proposal that talker variability affects the accuracy of recall of spoken words not only by increasing the processing demands for early perceptual encoding of the words, but also by affecting the efficiency of the rehearsal process itself.

The perception of spoken language is a skill that requires the listener to extract stable linguistic content from a physical signal that is notoriously unstable. The acoustic realization of speech is simultaneously modulated by numerous variable phonetic, prosodic, and semantic characteristics inherent to the particular message. The speech signal is further modulated by variable source characteristics. Indeed, it has frequently been suggested that if the perceptual system is required to extract canonical units of meaning from spoken language, various idiosyncracies related to different talkers must be somehow "normalized" during an early stage of processing the sensory input (e.g., Joos, 1948). Such talker-specific sources of variability include differing individual dialects, vocal tract sizes and shapes, and speaking rates (Mullennix, Pisoni, & Martin 1988; Verbrugge, Strange, Shankweiler, & Edman, 1976).

Sources of variability in speech, whether from phonetic context, spoken stress, or talker variations, have typically been considered "perceptual problems" to be solved by listeners, just as they must be solved for the design of adequate speech-recognition systems (e.g., Gerstman, 1968; Shankweiler, Strange, & Verbrugge, 1976). Although the communicative importance of talker-specific information has long been recognized (e.g., Ladefoged & Broadbent, 1957), only recently have questions regarding the encoding of talker-specific information been considered alongside questions regarding the perception of the phonetic content of speech. By assessing the effects of talker variability on higher-level cognitive processing, such as memory and recall, we may gain a more complete understanding of the relation between the perception of the linguistic message and the processing of talker-specific information in the speech signal.

Perhaps because of the efficiency of talker normalization processes, our everyday observation of language performance leaves the impression that listeners can attend to several different voices in succession with virtually no perceptual costs or consequences. Indeed, it has even been argued in the literature that our standard characterization of "normalization" per se may be unnecessarily rigid, portraying human listeners too much like speech recognition systems rather than as animate, event-oriented perceivers (see, e.g., Verbrugge et al., 1976). On the basis of vowel identification data, Verbrugge et al. suggested that the "problem of normalization" may be far less problematic for listeners than current speech recognition systems would lead one to believe. Nevertheless, several studies have shown that speech perception and spoken word recognition are affected by talker variability.[1] For example, Summerfield and Haggard (1973) showed that talker varia-

This research was supported by National Institutes of Health Research Grant DC-00111-13 to Indiana University, Bloomington, IN.

We thank Chris Martin and John Mullennix for their helpful comments, Ralph Geiselman and two anonymous reviewers for their insightful reviews, and Kim Spence for her help in collecting and scoring data.

Correspondence concerning this article should be addressed to David B. Pisoni, Department of Psychology, Indiana University, Bloomington, Indiana 47405.

[1] Throughout this article, the general terms *talker variability*, *talker manipulation*, and *talker condition* are used. Although the term could mean several kinds of stimulus variability, we are using it here only to denote situations in which spoken items are produced by different talkers from trial to trial in an experiment. We do not intend the term to denote variability of vocal quality within any given talker.

bility impairs vowel perception, as indexed by reaction time measures. More recently, Mullennix et al. (1988) showed that talker variability impairs subjects' word recognition performance in a variety of tasks. Mullennix et al. found that when subjects listened to words that were spoken by multiple talkers in succession, recognition was less accurate in a perceptual identification task and was slower in a naming task than when subjects listened to words produced by only a single talker. These findings led Mullennix et al. to propose that some resource-demanding mechanisms or processes are used by listeners to compensate for variations in speakers' voices (for a discussion of current theories of speaker and vowel normalization, see Johnson, 1990).

Other lines of evidence suggest that talker variability affects not only speech perception but memory processes as well. Recent experiments conducted by Martin, Mullennix, Pisoni, and Summers (1989) and by Logan and Pisoni (1987) have shown that spoken word lists produced by multiple talkers are more difficult to recall than the same word lists produced by a single talker. In a series of experiments, Martin et al. found that serial-ordered recall of spoken word lists was worse for multiple-talker lists than for single-talker lists, but only for items from early list positions. In addition, Martin et al. found that recall of visually presented digits presented before the spoken lists was worse if the subsequent lists were multiple-talker lists than if they were single-talker lists, and that the differences in recall of items from the primacy portion of the curve were unaffected by a postperceptual distractor task. From these findings, Martin et al. suggested that spoken word lists produced by multiple talkers may require greater processing resources for encoding and rehearsal in working memory than word lists produced by a single talker.

The account offered by Martin et al. was based on demonstrations of the obligatory and attention-demanding nature of voice information in speech perception. One such demonstration was provided by the recent findings of Mullennix and Pisoni (1990), who had subjects perform a Garner (1973) speeded-classification task. Subjects were required to classify spoken words according to either phonetic identity (/b/ vs. /p/) or speaker gender (male vs. female), while selectively ignoring variations along the irrelevant dimension. Mullennix and Pisoni obtained a pattern of results suggesting that information about talkers' voices is processed along with information about the phonetic content of words in an integral manner. That is, subjects were unable to selectively ignore voice information and attend only to phonetic information, even when the changes in voice slowed performance in the primary, phonetic classification task. In fact, subjects were better able to selectively ignore phonetic variability than talker variability. Apparently, talker variability across trials in these kinds of speech perception tasks requires continual reallocation of selective attention.

Similar findings have been reported by Geiselman and Bellezza (1976), who found that talker-specific voice information for spoken material is retained incidentally, even when subjects receive no specific instructions to attend selectively to voice characteristics. On the basis of these findings, Martin et al. suggested that when listeners are required to memorize lists of words spoken by multiple talkers, the variability introduced by changing talkers' voices requires allocation of processing resources that are also needed for the efficient rehearsal and transfer of list items to long-term memory. As a consequence, recall of early list items is impaired. This explanation is similar to a proposal such as that of Baddeley and Hitch (1974), that working memory can be characterized as a limited-capacity articulatory loop that can store only a small number of items before they need to be transferred to long-term storage. As more variability is imposed on this articulatory loop, greater demands are placed on the central processors that transfer items to long-term storage.

Although the processing capacity-based explanation offered by Martin et al. is consistent with their data, the authors noted that there are actually at least two possible ways that talker variability could affect rehearsal processes, and that either one or both may be responsible for the observed decrements in serial recall. The first possibility, following the findings reported earlier by Mullennix et al. (1988), is that the locus of the effects is confined exclusively to early perceptual encoding. That is, the extra time and resources required to "normalize" each token in a multiple-talker list are time and resources taken away from higher processing systems. Accordingly, the available rehearsal capacity for each word is reduced, and early list performance is attenuated. The second possibility considered by Martin et al. is that talker variability may also influence rehearsal processes themselves in a more direct way through changes in attention. That is, in addition to the early perceptual costs associated with encoding multiple-talker word lists, it is possible that the variability in voice information simply makes individual list items more difficult to rehearse and transfer into long-term memory because the rehearsal mechanisms have to accommodate increased variability for each new voice.

The present study was conducted to examine these two alternate explanations more closely. In particular, we wanted to determine whether talker variability affects perceptual encoding and only indirectly affects rehearsal, or if it is reasonable to assert that talker variability affects both perceptual encoding *and* subsequent rehearsal processes more directly. To explore this question, we replicated the first Martin et al. recall experiment, but we also manipulated two additional variables: word confusability and rate of presentation. The word confusability manipulation was selected because of its well-known influence on perception. Each word list in the present experiment consisted of 10 words; half of the lists contained *easy* words and half of the lists contained *hard* words. The word confusability measure used here was based on a combined metric of three factors known to influence spoken word recognition: word frequency, lexical neighborhood density, and neighborhood frequency. The first measure is simply the classic word frequency index based on the word frequency count of Kučera and Francis (1967). The second and third measures are based on analyses of *similarity neighborhoods* (e.g., Coltheart, Davelaar, Jonasson, & Besner, 1977; Luce, 1986). A similarity neighborhood is defined as a collection of words that sound similar to a given word. One characteristic of similarity neighborhoods that has proven important in word recognition studies is the number of neighbors that a word has; some words have many similar-sounding

neighbors, whereas other words have relatively few neighbors. For spoken words, Luce (1986) showed that word recognition is slower and less accurate for words selected from dense neighborhoods than for words selected from sparse neighborhoods, and that words with many neighbors of equal or higher frequency are recognized less easily than words with few neighbors of equal or higher frequency (see also Pisoni, Nusbaum, Luce, & Slowiaczek, 1985). In the present study, *easy* words were defined as high frequency words selected from sparse, low frequency neighborhoods, whereas *hard* words were defined as low frequency words selected from dense, high frequency neighborhoods.[2]

Talker variability was manipulated in the present investigation in the same manner as it was in the Martin et al. study: Talker variability was treated as a between-subjects variable; half of the subjects heard only single-talker lists, and the other half heard only multiple-talker lists. The confusability manipulation was a within-subjects variable; half of the lists presented to each group consisted of easy words, and half consisted of hard words, as defined by the criteria previously enumerated.

These two stimulus dimensions were manipulated in the present study because they have been shown to be perceived in fundamentally different ways. There are several qualitative differences between these two dimensions. First, information about the talker's voice is *concrete*: It is manifested by physical variations in the acoustic forms of spoken words. Conversely, information about word frequency and neighborhood confusability is of a far more abstract and derived nature, denoting relationships among words in lexical memory. Second, the findings of Mullennix and Pisoni (1990) and Geiselman and Bellezza (1976) suggest that information regarding the source of an utterance is not only attention demanding, but is potentially useful and ecologically relevant as well. Voices convey important indexical information about a speaker's gender, age, and emotional state (Geiselman & Bellezza, 1976). In contrast, if word confusability conveys any information at all, it could only be about the average listener's internal lexicon and the relative accessibility of its component words. Finally, talker variability and word confusability differ in perceptual salience. Although it may be readily apparent to a listener that two successive words are spoken by two different talkers, it may not be apparent that two successive words differ in frequency or confusability.

To assess the effects of both talker variability and word confusability on list rehearsal, we examined the accuracy of recall for spoken word lists varying along both stimulus dimensions across five levels of a third experimental manipulation. Specifically, because we were interested in a manipulation that would primarily affect rehearsal processes, we varied the rate of presentation of words in the lists presented to subjects. In earlier studies, Murdock (1962), Jahnke (1968), and Rundis (1971) all showed that rate of presentation affects the accuracy of recall of early list items, and they argued that shorter presentation rates may not provide enough time for adequate rehearsal and transfer of items to long-term storage. In the present study, words were presented at one of five rates, with interword intervals of either 250, 500, 1,000, 2,000, or 4,000 ms.

The logic and predictions of the manipulations used in this experiment rested upon two assumptions. The first assumption was that variations in speaker identity contain more salient and usable information than variations in word confusability. The second assumption was that subjects' rehearsal strategies would be affected by the extra information presented by each voice, but this would depend on the amount of rehearsal time available. If talker variability affects rehearsal efficiency above and beyond simply disrupting perceptual encoding, we would expect a strong interaction of talker and presentation rate, perhaps even to the degree that recall of multiple-talker lists could exceed recall of single-talker lists at slow rates. Changes in rate should affect recall of words from multiple-talker lists more than words from single-talker lists, because there is attention-demanding, distinctive information contained in the multiple-talker lists. Furthermore, if word confusability only affects perceptual encoding but leaves rehearsal processes relatively unaffected, no interaction should be observed between confusability and presentation rate. Changes in rate should affect recall of words from easy and hard lists equivalently, because abstract word confusability information may be more intrinsically difficult to exploit by any overt, elaborative rehearsal or retrieval strategies. By examining the interactions of these stimulus variables with presentation rate, we can better assess the degree to which talker variability affects rehearsal processes above and beyond perceptual encoding.

Method

Subjects

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

Stimuli

The stimuli were obtained from a large digitized database of spoken monosyllabic words recorded by several different talkers. This was the same source used by Martin et al. (1989). The original words came from the vocabulary used in the Modified Rhyme Test (House, Williams, Hecker, & Kryter, 1965). In the present experiment, only a subset of the original 300 words were used. The words selected for the present experiment satisfied several constraints: First, the words were ranked according to their frequency of occurrence according to the Kučera and Francis (1967) norms. Second, the words were ranked according to their neighborhood densities as determined by a one-phoneme substitution, addition, and deletion metric (Luce, 1986). Third, the words were ranked according to their neighborhood frequencies, a measure of the average frequency of the words' neighbors.

[2] Throughout the remainder of this article, for ease of composition and comprehension, the following terminology is used: The stimulus dimension relating to single versus multiple takers is referred to as the *talker* variable or manipulation. Similarly, the stimulus dimension relating to *easy* versus *hard* words is referred to as the *confusability* variable or manipulation.

Using these three criteria, two sets of words were selected for use in the present experiment. One set, the easy words, consisted of high frequency words from low density, low frequency neighborhoods. The other group of words, the hard words, consisted of low frequency words from high density, high frequency neighborhoods. A final criterion used in selection was subjective familiarity; all of the words chosen for use in the experiment were rated as highly familiar by subjects in an earlier experiment conducted by Nusbaum, Pisoni, and Davis (1984). After the words were divided into easy and hard sets according to these four criteria, each condition contained 50 items. These 100 words were then used to generate 10 lists of 10 words each. Five of the lists contained all easy words, and 5 contained all hard words.

Once the words had been selected, digitized files containing tokens of each word were obtained from the database. One set of tokens was chosen from utterances produced by a single male talker; these tokens were used for the single-talker conditions of the experiment. Another set of tokens was selected from the database so that every word in each list was spoken by a different talker; these tokens were used for the multiple-talker conditions. In the multiple talker conditions, the same 10 talkers, 5 men and 5 women, were used for all 10 list of words. The talkers used in the present experiment were the same talkers used in the Martin et al. (1989) study.[3] All of the stimuli were originally recorded on audiotape and digitized with a 12-bit analog-to-digital converter using a PDP 11/34 computer. The mean root-mean-square amplitude of all stimulus tokens was equated by using a signal processing package. All stimulus tokens used in the present experiment had been tested for intelligibility in previous experiments that made use of this database (e.g., Martin et al., 1989), and were found to be equally intelligible to subjects.

Procedure

Subjects were tested in groups of 6 or fewer in a quiet testing room used for speech perception experiments. Stimuli were presented over matched and calibrated TDH-39 headphones at 75 dB (SPL). A PDP 11/34 computer was used to present the stimuli and to control the experimental procedure in real time. The digitized stimuli were reproduced using a 12-bit digital-to-analog converter and were low-pass filtered at 4.8 kHz.

All subjects were tested under the same conditions. Subjects first heard a 500-ms, 1000-Hz warning tone indicating that a list of words was about to be presented. Then, a list of 10 words was presented at one of five rates: one word was presented either every 250, 500, 1,000, 2,000, or 4,000 ms. The presentation rate selected was held constant for any given group of subjects for an entire session. After each list of words, another tone was presented, indicating the beginning of the recall period. Subjects had 60 s to recall all the words they could. The end of the recall period was indicated by the presentation of a third tone. Subjects were instructed to recall the words in the exact order of their presentation in the lists. Subjects wrote their responses in specially prepared answer booklets using pen or pencil.

Presentation rate and talker condition were between-subjects variables; word confusability was a within-subjects variable. Thirty-two subjects were tested at each rate of presentation. Half of the subjects tested at each rate received single-talker lists and half received multiple-talker lists. The same words were heard by all subjects; only the number of talkers and the presentation rates varied between subjects. The order of presentation of words within each list varied randomly across sessions. The lists themselves were presented in the same order in all conditions of the experiment; the presentation of lists for each group alternated between those lists containing easy words and those containing hard words.

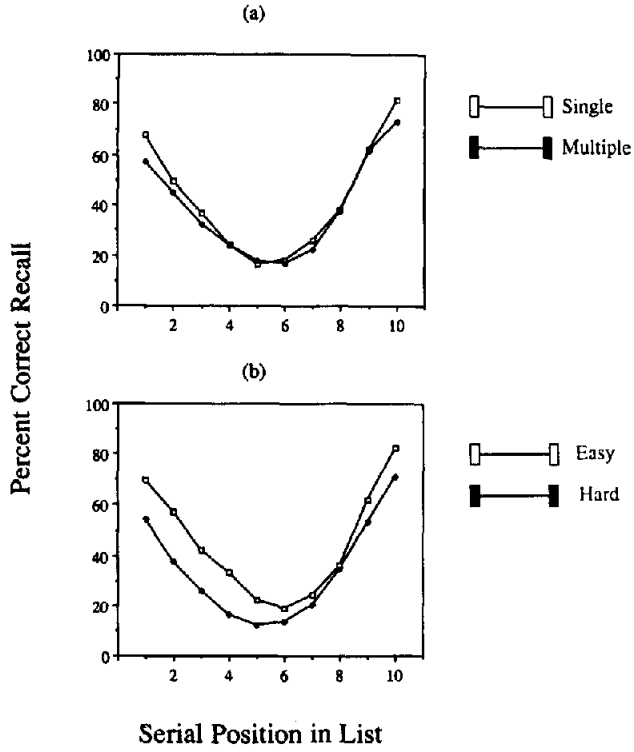

Figure 1. Mean percentages of correctly recalled words as a function of (a) serial position and talker condition, collapsed across presentation rate and word confusability, and (b) serial position and word confusability, collapsed across presentation rate and talker.

Results

Subjects' responses were scored as correct only if the target word or some phonetically equivalent spelling of the target word was recalled in the same serial position as the word presented in the list. Panel a of Figure 1 displays the percentage of correctly recalled words as a function of serial position

[3] Because the premise of this study rested upon the working assumption that both word confusability and talker variability affect spoken word recognition approximately equivalently, we conducted a pilot experiment with the stimulus items selected for the recall experiment. This experiment was conducted to ensure that both the confusability and talker manipulations used in the recall experiment would produce deficits in perceptual encoding time of comparable magnitude. Forty subjects participated in a naming task, in which they were presented the stimulus tokens and were asked to repeat the words as quickly and accurately as possible. Replicating previous work (e.g., Mullennix et al., 1988), we observed main effects of both word confusability and talker. The latencies to respond were longer for both hard words and words from multiple-talker lists. The average magnitudes of these effects were 67.80 and 79.60 ms, respectively ($p < .02$, both main effects), and no significant interaction of the factors was observed ($p = .7561$). Because the perceptual effects of both the talker and confusability manipulations were equivalent in the naming task, we assumed comparisons of their respective interactions with presentation rate would not be confounded by possible differences in the underlying psychological scales for the two variables.

and talker condition, collapsed across presentation rate and word confusability. Panel b displays the percentage of correctly recalled words as a function of serial position and word confusability, collapsed across presentation rate and talker.

A four-way analysis of variance (ANOVA) (Talker × Word Confusability × Serial Position × Presentation Rate) was conducted on the percentage of correct responses. In the ANOVA, talker and presentation rate were treated as between-subjects variables; word confusability and serial position were treated as within-subjects variables. As expected, a significant main effect of talker was observed, $F(1, 150) = 3.98$, $MS_e = 22.05$, $p < .05$. (All results reported are statistically reliable at the $p < .05$ level or beyond, except for specifically reported null findings.) Words from single-talker lists were recalled more accurately than words from multiple-talker lists. In addition to the effect of talker, Figure 1 also shows a strong main effect of serial position, $F(9, 1350) = 267.00$, $MS_e = 5.52$, reflecting the usual U-shaped function obtained in recall tasks. A significant two-way interaction of talker and serial position was also obtained, $F(9, 1350) = 2.05$, $MS_e = 5.52$. The differences in recall between the single-talker and multiple-talker lists tended to be larger at earlier list positions. Post hoc Tukey's honestly significant difference (HSD) analyses were performed on the percentage of correctly recalled words at each serial position. These analyses indicated that the recall functions for single- and multiple-talker lists were significantly different only at serial positions 1 and 10. The magnitude of the effects of talker obtained here was smaller than the effects reported by Martin et al. (1989). However, the major effects of the talker manipulation are obscured by averaging over the five presentation rate conditions, as discussed later.

A significant main effect of word confusability was obtained, $F(1, 150) = 147.70$, $MS_e = 5.16$. Recall of easy words was more accurate than recall of hard words at most serial positions of the lists. There was, however, a significant two-way interaction of word confusability and serial position, $F(9, 1350) = 8.29$, $MS_e = 3.14$, reflecting the larger differences between the recall functions for easy and hard words at early list positions. Post hoc Tukey's HSD analyses showed that accuracy of recall for easy and hard words was significantly different at serial positions 1, 2, 3, 4, 5, 9, and 10.

Figure 2 shows data for the single- and multiple-talker conditions as a function of serial position and presentation rate collapsed across word confusability. Panel a displays the recall functions for lists spoken at the fastest of the five rates. Panels b to e display the recall functions for lists presented at each of the remaining four rates in decreasing order, with the slowest of the five rates displayed in Panel e.

The ANOVA revealed a significant main effect of presentation rate, $F(4, 150) = 22.56$, $MS_e = 22.05$. Overall, word recall improved as the rate of presentation became slower. More important, however, was the significant three-way interaction of talker, presentation rate, and serial position, $F(36, 1350) = 2.23$, $MS_e = 5.52$. This interaction demonstrates that rate of presentation affected recall of items from the primacy portions of multiple-talker lists more than single-talker lists.

At the faster rates of presentation, the accuracy of recall of items from single-talker lists was better than recall of items from multiple-talker lists, especially in the primacy portion of the curves. As the rate of presentation decreased, however, the differences between the two recall functions diminished and then reversed at the slower rates, as shown in Panels d and e of Figure 2. Indeed, at the slowest rate, recall for early

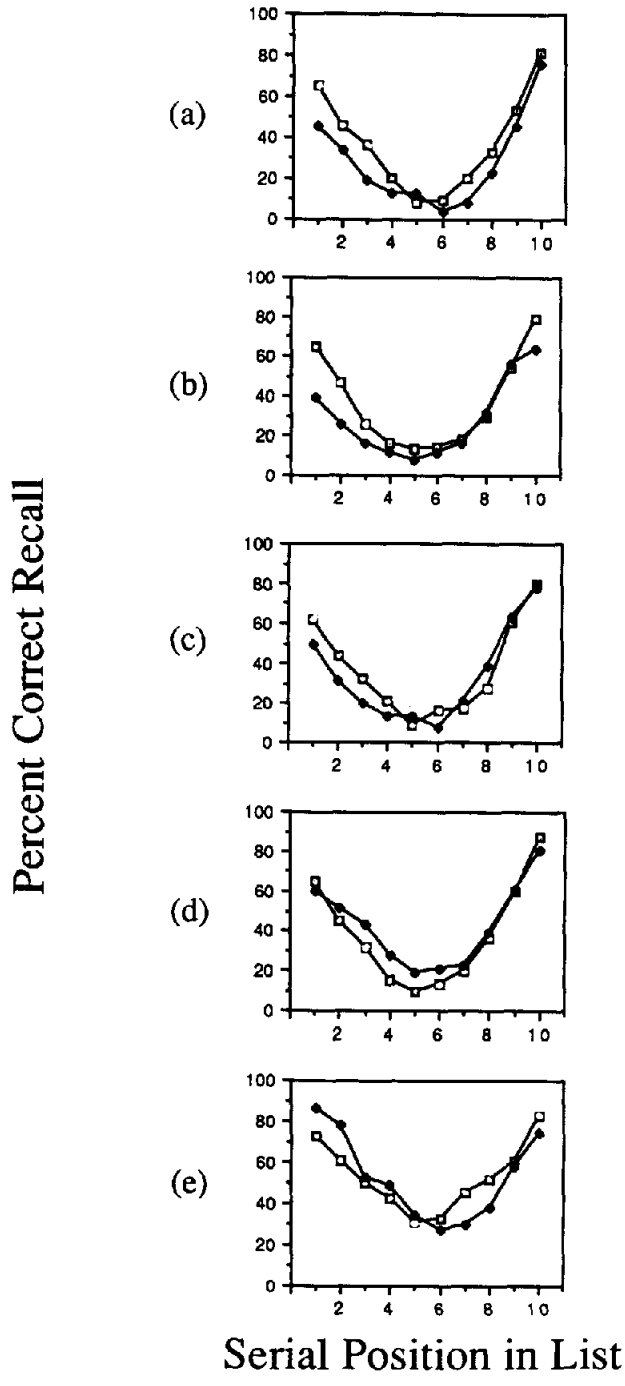

Figure 2. Mean percentages of correctly recalled words for both the single- and multiple-talker lists as a function of serial position and presentation rate, collapsed across word confusability. The five panels display the results at each rate of presentation, one word every (a) 250 ms, (b) 500 ms, (c) 1,000 ms, (d) 2,000 ms, and (e) 4,000 ms. Open squares represent single-talker lists; filled squares represent multi-talker lists.

list items from the multiple-talker lists was actually better than recall for the single-talker lists. Post hoc Tukey's HSD analyses were conducted to compare the recall functions at all serial positions. These analyses showed that in all conditions, with the exception of the 2,000-ms condition (shown in Panel d), the differences obtained in the early list positions were statistically reliable. In the 2,000-ms condition, significant differences in recall were observed only at serial positions 3, 4, 5, and 6. In the 4,000-ms condition (shown in Panel e), words from early serial positions of the multiple-talker lists were recalled better than words from single-talker lists. The post hoc analyses showed that recall of items from early list positions significantly improved when the presentation' rate was changed from 1,000 to 2,000 ms and when the rate was changed from 2,000 to 4,000 ms. This crossover effect is responsible for the three-way interaction observed between talker, presentation rate, and serial position previously noted.

Figure 3 shows recall of both the easy and hard word lists as a function of serial position and presentation rate, collapsed across talker. Panel a displays the recall functions for lists presented at the fastest rate. Panels b to e display the recall functions for lists presented at each of the remaining four rates in decreasing order, with the slowest rate displayed in Panel e.

As shown in Figure 3, accuracy of word recall improved uniformly as presentation rates slowed. However, although it is clear that the recall functions for both easy and hard words were affected by the presentation rate manipulation, the critical three-way interaction of word confusability, presentation rate, and serial position was not significant in the analysis, $F(36, 1350) = 1.14$, $MS_e = 3.13$, $p = .2660$. Thus, unlike the finding obtained for the talker manipulation, changes in rate of presentation did not differentially affect the recall of easy and hard words. Examination of the recall functions for easy and hard words reveals that the manipulation of presentation rate had comparable effects on both kinds of words; the improvement of word recall with slower rates was equivalent for easy and hard lists. These results are in marked contrast to the data shown in Figure 2, which showed that slower presentation rates affected recall of multiple-talker lists much more than single-talker lists. Post hoc Tukey's HSD analyses revealed significant differences in recall performance for easy and hard words at most of the early list positions and at several terminal positions as well. Accuracy of recall was consistently better for easy words than for hard words at all presentation rates.

In addition to the null finding previously reported, several additional null findings deserve mention. First, no significant interaction was observed between talker variability and word confusability, $F(1, 150) = .07$, $MS_e = 5.16$, $p = .7917$, implying that the word confusability differences were unaffected by talker condition. Second, no significant interaction was observed between word confusability, talker, and presentation rate, $F(4, 150) = 1.39$, $MS_e = 5.16$, $p = .2390$, implying that the changes observed in recall of multiple-talker lists following rate changes were independent of word confusability condition. Finally, the four-way interaction of word confusability, talker variability, presentation rate, and serial position was not significant, $F(36, 1350) = 1.17$, $MS_e = 3.13$, $p = .2253$. It is easy to determine why the four-way interaction did not reach significance, as several of the experimental conditions showed no differences. Figures 2 and 3 clearly illustrate that presentation rate did not differentially affect recall of easy versus hard words, and only differentially af-

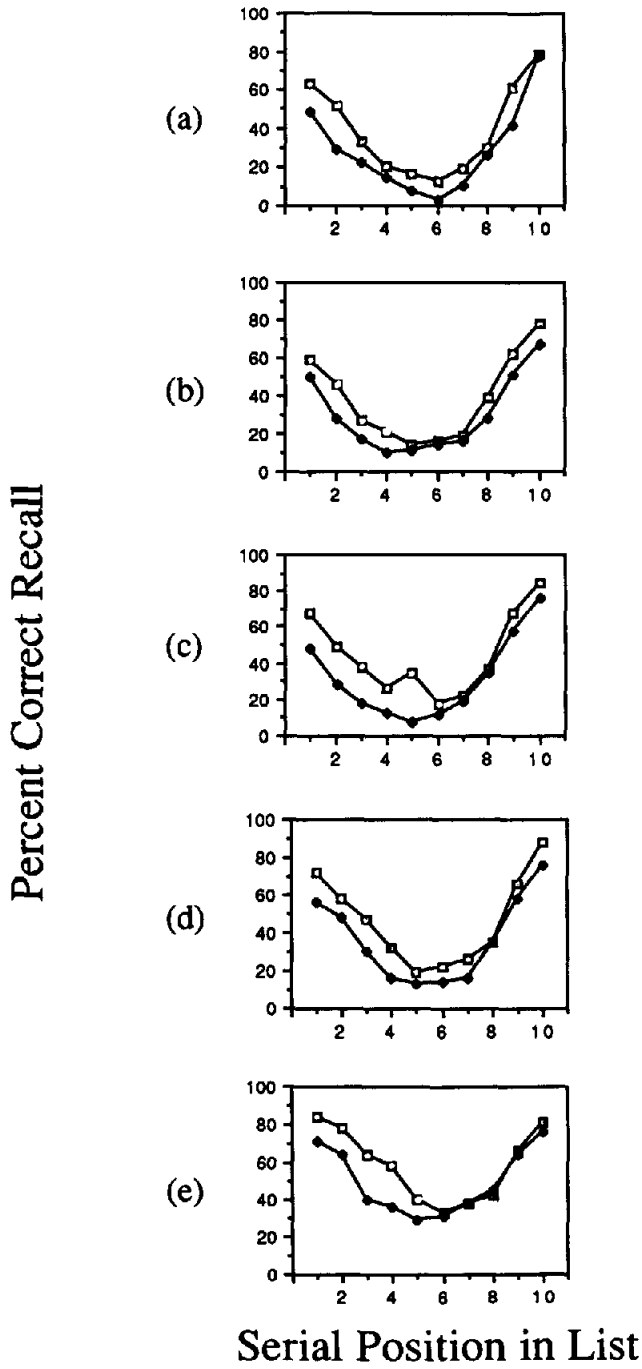

Figure 3. Mean percentages of correctly recalled words for both the easy and hard word lists as a function of serial position and presentation rate, collapsed across talker. The five panels display the results at each rate of presentation, one word every (a) 250 ms, (b) 500 ms, (c) 1,000 ms (d) 2,000 ms, and (e) 4,000 ms. Open squares represent easy lists; filled squares represent hard lists.

fected recall of words spoken by single versus multiple talkers at the slower presentation rates.

Discussion

The present investigation was conducted to further examine the effects of talker variability on recall of spoken word lists. Specifically, this study was designed to evaluate two alternate explanations suggested by Martin et al. (1989) for their recent finding that recall was better in early list positions for lists spoken by a single talker than for lists spoken by multiple talkers. One possible explanation, an encoding account, is based exclusively on the perceptual consequences of talker variability. This view suggests that the initial delays that occur in word recognition when talkers' voices change from trial to trial simply "cascade up the system," passively reducing the time and processing resources available for rehearsal of these items. The second explanation is a more direct, rehearsal-based account, suggesting that, in addition to the perceptual costs associated with talker variability, changes in voice from item to item in a word list also affect the speed and/or efficiency both of the rehearsal processes required to transfer items from working memory into long-term memory.

To distinguish between these alternate explanations, we examined the effects of presentation rate on recall of word lists spoken by single and multiple-talker, and we compared these results to the effects of presentation rate on recall of lists of easy and hard words. Previous evidence (Mullennix et al., 1988) and our own naming data (see Footnote 3) show that the influences of these two manipulations on word recognition are of comparable magnitudes. Therefore, if simple encoding deficits were responsible for the effects of talker variability on recall observed by Martin et al., then both talker variability and word confusability should have affected the accuracy of recall in the same way. Both variables should have interacted equivalently with the rate manipulation and, therefore, recall for both multiple-talker lists and lists of hard words should have changed equivalently with respect to their appropriate counterparts as presentation rate changed. This pattern of results was not obtained in the present study; instead, we found that recall of multiple-talker lists changed much more than recall of single-talker lists when presentation rate was varied, whereas recall of lists of hard words changed no more than recall of easy words. Indeed, we observed a surprising sensitivity of multiple-talker lists to the presentation rate changes; at the slowest rate of one word every 4 s, recall of early list items for words spoken by multiple talkers was actually better than recall of words spoken by a single talker.[4]

The differential interactions of the talker and confusability factors with presentation rate suggest that talker variability affects not only early perceptual encoding but rehearsal processes as well. One final observation from the present data provides further support for this conclusion. If both the word confusability and talker variability manipulations affected recall by impairing only perceptual encoding, one might expect to find an additive effect of the two factors when presentation rate is manipulated. That is, if only perceptual encoding deficits were responsible for the interaction of talker and presentation rate, then the effect should be even greater for hard words spoken by multiple talkers, implying a three-way interaction of word confusability, talker, and presentation rate. However, the three-way interaction was not significant in the present experiment.

Interpretations of the present results should not be overstated, however. Although the data provide support for the rehearsal-based account of the recall data collected by Martin et al. (1989), it is clear that results of talker variability *interfering* with recall remain potentially ambiguous. There is ample evidence that both word confusability and talker variability adversely affect word recognition speed and accuracy (Mullennix et al., 1988; see also Footnote 3). Given that talker variability reliably and strongly interferes with word recognition, and that word recognition is the necessary first step in the recall process, it may not be possible to directly assess the independent effects of perceptual encoding deficits and rehearsal deficits in reducing the serial recall of multiple-talker lists. Furthermore, to the extent that perceptual encoding and rehearsal can be considered separate but dependent stages, as in a cascade model (e.g., McClelland, 1979), it is apparent that any variable that affects encoding speed will at least indirectly affect rehearsal as well.

Rather than attempt to deny the perceptual encoding side of the issue, therefore, we have attempted to marshall further evidence that talker variability does not *only* affect perceptual encoding, but that voice information remains as an integral component of the memory representations of all items. We suggest that this "extra" information provided by talker variability may be either harmful to subjects' performance in tasks that are highly resource demanding or helpful to performance when task restrictions allow for more thorough, elaborative processing to proceed. In brief, we suggest that voice information may directly affect list rehearsal above and beyond initial perceptual effects.

Given this suggestion, it is appropriate to briefly review some of the available evidence demonstrating that talker-related information remains available to the listener and directly influences rehearsal processes. Consider first several of the results reported by Martin et al. (1989). As described in the introduction, Martin et al. found that recall of visually presented digits was reduced when subsequent lists of to-be-remember words were spoken by multiple talkers compared with when they were spoken by a single talker. Effects of this sort have typically been considered evidence that a common

---

[4] Upon examination of Figure 2, the reader may notice that no appreciable changes occurred between the single- and multiple-talker recall functions across Panels a, b, and c. When considering these results, it is important to bear in mind that the presentation rate delays doubled with each condition, so one should not expect the changes in recall of multiple-talker lists to be commensurate with condition number. The data seem to suggest that if longer presentation delays benefit recall at all, it is only when the delays between words are considerably longer than the time needed to recognize the words (i.e., delays longer than about 1,000 ms). This finding may be considered a further implication that the rate manipulation improves recall primarily by affecting some stage of processing that operates only after perceptual encoding is complete.

pool of attentional resources must be simultaneously dedicated to maintaining the first items in working memory and processing the subsequent items as well (Baddeley & Hitch, 1974). If talker variability merely affected word recognition and voice information could be subsequently ignored, such effects would not be expected. In an earlier article, Martin et al. (1987) also reported that effects of talker variability on list recall are observed only in serial-ordered recall but not in free recall, suggesting that, in the more difficult task, talker-specific information and list-order information compete for limited attentional resources in working memory. Finally, consider again the data provided by the present study; talker variability interacts with presentation rate whereas word confusability does not. Interactions with presentation rate are typically assumed to reflect factors involved with rehearsal or other attentional processing (Glanzer & Cunitz, 1966; Murdock, 1962), further suggesting that talker variability affects more than simply perceptual encoding.

Related results are reported in recent study conducted by Lightfoot (1989). Using stimulus tokens from the same database used in the present experiment, Lightfoot trained subjects over a period of 9 days to correctly recognize the voices of five male and five female talkers in the database, and to associate the voices with fictitious names, such as Brad, Bill, Jane, Mary, and so on. Following the training period, both trained and untrained subjects participated in a serial recall task. The stimulus words presented in the recall task were the same tokens used in the present experiment, and were never presented during training. The trained subjects in Lightfoot's study displayed the same multiple-talker list advantage that we observed in the present investigation. However, whereas our subjects required a slow, 4,000-ms presentation rate, Lightfoot's trained subjects required only a 1,500-ms presentation rate to show the reversal. Although the training procedure presumably facilitated subjects' perceptual encoding of the list items, training only marginally diminished the effects of the word confusability variable. Lightfoot concluded that training subjects to process the stimulus voices more efficiently improved recall primarily by reducing rehearsal demands and enhancing retrieval cues, rather than by merely facilitating perceptual encoding.

Returning to the present results, it should be noted that, although word confusability did not interact with presentation rate, there is a suggestion in the present data that word confusability may differentially affect long- and short-term memory. As Figures 1 and 3 show, the differences in recall between easy and hard lists were consistently larger in early list positions than in late list positions. Similar findings have been reported by Sumby (1963), who found that recall of high frequency words was better than recall of low frequency words, especially from early list positions. Sumby suggested that it may be easier to rehearse more familiar words and encode them into long-term memory.

Other explanations for the effects of word confusability are available as well. For example, the perceptual encoding deficits associated with low frequency words may provide an adequate explanation for the larger primacy effects of word confusability. Although the explanation now seems to simplistic to account for the effects of talker variability on recall, it is possible that the early perceptual problems caused by hard words indirectly affect rehearsal processes merely by usurping processing time. Another possible explanation of the larger differences between easy and hard words in early list positions may be differential retrieval of items from long-term memory. Because hard words in this study were low frequency words with many higher frequency neighbors, it follows that ambiguity and perceptual confusions during retrieval might produce more errors for hard words than for easy words. Whether the differential effects of word confusability on primacy and recency imply interactions with rehearsal or retrieval processes is not the issue here, however. The important point is that, although there is some suggestion that word confusability *may* affect rehearsal processes, there is now strong evidence that talker variability *does* directly affect rehearsal processes.

Indeed, the observed advantage for recall of lists spoken by multiple talkers at the slowest rate suggests that voice information may be incidentally retained in the representations of list items and may remain available to subjects throughout the task, facilitating retrieval of words from long-term memory. Detailed information about the talker's voice for each word may be used as an additional cue for encoding temporal order, for example, but only if subjects are given sufficient time to use this information during rehearsal. Given the salient nature of voice information for spoken words, one might speculate that at faster presentation rates, talker variability simply overwhelms subjects, distracting their attention and increasing the total amount of information to rehearse per unit time. This consideration of the present investigation bears a close resemblance to an account of recent data collected by Aldridge, Garcia, and Mena (1987). Aldridge et al. had subjects recall one target word from each of a series of lists of visually presented words. Subjects were provided either 10 or 60 s to rehearse the target item, and the lists were presented either with or without the distraction of irrelevant visual and auditory events. The data showed that amount of rehearsal time only affected recall in the distracting presentation conditions. Aldridge et al. suggested that the distracting events prevented subjects from habituating to the experimental task, and that habituation is a necessary condition for efficient maintenance rehearsal in working memory. In the present study, talker variability at fast presentation rates may similarly prevent subjects from habituating to the recall task and thereby preclude efficient rehearsal.

Conversely, at slower presentation rates, listeners may attend more completely to the pairings of specific words and talkers, and can use this distinctive information as an elaborative temporal cue to make list items more discriminable, and thereby improve recall of both item and order information. With either increased processing time or familiarity with voices (Lightfoot, 1989), talker variability appears to change from "noise" that subjects cannot ignore to useful information they can exploit. In this connection, it is also interesting to note that Craik and Kirsner (1974) found that talker variability improved subjects' recognition memory for lists of words. Because the recognition memory task requires less processing effort than the serial recall task, this improvement for multiple-talker lists may be considered analogous to the improve-

ment observed at slower presentation rates in the present investigation.

The present finding that talker variability not only affects early perceptual encoding but also higher level processes has precedents in the literature. As Mullennix and Pisoni (1990) and Martin et al. (1989) suggested, voice information appears to be processed in an obligatory manner, in the sense of Fodor (1983). When subjects listen to words spoken by many different talkers in succession, information provided by the different voices demands attention and processing capacities. It is not clear, however, whether the bulk of the processing efforts in the present task are dedicated to intentionally ignoring or attenuating voice information, or if they are allocated to encoding voice information as a useful cue for preserving temporal order information in recall. At slower presentation rates, the latter explanation seems more appropriate, given the advantage observed for recall of multiple-talker lists. At faster rates, either of these explanations could account for the decrease of recall in primacy.

Other researchers have reported similar findings about the obligatory nature of voice information in speech perception. For example, Cole, Coltheart, and Allard (1974) and Allard and Henderson (1976) reported that reaction times to report "same word" are faster in a same–different task when the words are spoken by a single talker than when the words are spoken by different talkers. Furthermore, Craik and Kirsner (1974) found that voice information remains in memory and affects word recognition for at least 2 min. Several studies conducted by Geiselman and his colleagues (Geiselman & Bellezza, 1976, 1977; Geiselman & Crawley; 1983; Geiselman & Glenny, 1977) also examined the incidental storage of talkers' voices during the processing of linguistic information, and found that a talker's voice is encoded into long-term memory even when subjects are not specifically instructed to attend to voices.[5] Finally, Kosslyn and Matt (1977) found that subjects' knowledge of a writer's speaking rate can affect how quickly they read his or her prose.

In summary, the present data provide support for the explanation suggested recently by Martin et al. (1989), that talker variability may affect recall of spoken words lists not only by slowing down initial perceptual encoding, but by reducing the efficiency of rehearsal processes as well. Speech perception theorists since the time of Joos (1948) have been guided by an assumption that sources of acoustic variability, such as talker variations, must be "removed" from the listener's percept so that only some canonical linguistic message remains (Studdert-Kennedy, 1974, 1976). A growing body of data now seems to imply that this notion of "perceptual normalization" as an information-reduction procedure may be incorrect. Normalization may well be an integral aspect of speech perception, but our characterization of the process should involve not only the extraction of linguistic meaning from the signal, but extraction of important speaker-dependent information as well (see Ladefoged & Broadbent, 1957). The example provided in this study may be regarded as a case in point; the memory deficits associated with talker variability apparently reveal more than just simple peripheral adjustments of the speech perception mechanisms that allow the words to be recognized and the voice to be discarded. Instead, processing talker-specific characteristics appears to be an integral aspect of speech perception, and, like so much of language and cognitive processing, it appears to require a complex interplay of perceptual, attentional, and memory processes.

[5] It is interesting to note here that in another of Geiselman's studies (Geiselman, 1979), it was reported that subjects can inhibit the automatic encoding of voice information when such information interferes with a primary cognitive task. The findings of the present study, as well as the Martin et al. (1989) study, do not support this claim. However, it is possible that the deficits in recall for multiple-talker lists reported here could result from inefficient use of voice information rather than from interfering effects of voice information per se. Geiselman's (1979) finding lends support to the notion that voice information may be incorporated in some way into the long-term memory representation for spoken words, not "stripped away" from the tokens during initial encoding as a traditional "perceptual normalization" hypothesis would imply.

References

Aldridge, J. W., Garcia, H. R., & Mena, G. (1987). Habituation as a necessary condition for maintenance rehearsal. *Journal of Memory and Language, 26*, 632–637.

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

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and memory, vol. 8* (pp. 47–90). New York: Academic.

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

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

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

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

Garner, W. R. (1973). *The processing of information and structure.* Potomac, MD: Erlbaum.

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

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

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

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

Geiselman, R. E., & Glenny, J. (1977). Effects of imagining speakers' voices on the retention of words presented visually. *Memory & Cognition, 5*, 499–504.

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

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

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.

Jahnke, J. C. (1968). Presentation rate and the serial-position effect of immediate serial recall. *Journal of Verbal Learning and Verbal Behavior, 7*, 608–612.

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

Joos, M. A. (1948). Acoustic phonetics. *Language, 24*, 1–136.

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

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

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

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

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

Luce, P. A. (1986). Neighborhoods of words in the mental lexicon. *Research on Speech Perception, Technical Report No. 6*. Bloomington, IN: Indiana University.

Martin, C. S., Mullennix, J. W., Pisoni, D. B., & Summers, W. V. (1987). Effects of talker variability on recall of spoken word lists. *Research on Speech Perception Progress Report No. 13*. Bloomington, IN: Indiana University.

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

McClelland, J. L. (1979). On the time-relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86*, 287–330.

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

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1988). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America, 85*, 365–378.

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

Nusbaum, H. C., Pisoni, D. B., & Davis, C. K. (1984). Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words. *Research on Speech Perception Progress Report No. 10*. Bloomington, IN: Indiana University.

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

Rundis, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology, 89*, 63–77.

Shankweiler, D. P., Strange, W., & Verbrugge, R. R. (1976). Speech and the problem of perceptual constancy. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing: Toward an ecological psychology* (pp. 315–346). Hillsdale, NJ: Erlbaum.

Studdert-Kennedy, M. (1974). The perception of speech. In T. A. Sebeok (Ed.), *Current trends in linguistics* (Vol. 12, pp. 2349–2385). The Hague: Mouton.

Studdert-Kennedy, M. (1976). Speech perception. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics* (pp. 243–293). New York: Academic Press.

Sumby, W. H. (1963). Word frequency and serial position effects. *Journal of Verbal Learning and Verbal Behavior, 1*, 443–450.

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

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

Appendix

Stimulus Materials Used in the Recall Task

Easy words		Hard words	
page	dust	cane	pill
book	shop	real	kick
pass	soil	pun	gale
sing	sang	peak	fame
sold	park	fin	scep
cut	beach	mat	sane
teach	bus	rip	tin
top	dark	dill	cake
peace	raw	lick	bail
rang	bent	kit	sack
big	took	buck	lace
hold	got	heal	cop
kid	told	ban	pip
shook	king	peat	den
save	just	bead	pit
mark	oil	lame	neat
west	cup	tack	din
went	did	kin	beak
gun	test	wick	lip
law	team	hip	tick
look	name	rake	ray
then	cook	pin	wed
cold	safe	peel	wit
bath	must	bat	pan
rest	path	bun	sill

Received November 21, 1989
Revision received June 8, 1990
Accepted June 20, 1990 ■