



Published in final edited form as:

J Exp Psychol Learn Mem Cogn. 1989 July ; 15(4): 676–684.

Effects of Talker Variability on Recall of Spoken Word Lists

C. S. Martin, J. W. Mullennix, D. B. Pisoni, and W. V. Summers

Indiana University

Abstract

Three experiments were conducted to investigate recall of lists of words containing items spoken by either a single talker or by different talkers. In each experiment, recall of early list items was better for lists spoken by a single talker than for lists of the same words spoken by different talkers. The use of a memory preload procedure demonstrated that recall of visually presented preload digits was superior when the words in a subsequent list were spoken by a single talker than by different talkers. In addition, a retroactive interference task demonstrated that the effects of talker variability on the recall of early list items were not due to use of talker-specific acoustic cues in working memory at the time of recall. Taken together, the results suggest that word lists produced by different talkers require more processing resources in working memory than do lists produced by a single talker. The findings are discussed in terms of the role that active rehearsal plays in the transfer of spoken items into long-term memory and the factors that may affect the efficiency of rehearsal.

The acoustic properties of speech vary dramatically as a function of context, speaking rate, and a number of talker-related factors such as vocal tract configuration, glottal characteristics, and dialect. Many theorists have argued that in order for spoken language to be perceived rapidly and efficiently, some sort of perceptual process must compensate for the acoustic differences between individual talkers (e.g., Joos, 1948; Verbrugge, Strange, Shankweiler, & Edman, 1976). Talker differences are typically assumed to be “normalized” at early stages of perceptual analysis so that linguistic units can be efficiently extracted from the speech waveform (Summerfield & Haggard, 1973). Although perceptual normalization of talker differences has been recognized as an important problem almost from the beginning of modern speech research, little is actually known about the nature of this type of perceptual compensation. Human listeners are able to perceive and understand speech produced by a wide variety of talkers and appear to display little, if any, additional effort or processing demands. An examination of the published literature reveals that almost all speech perception and memory research using natural speech has employed stimulus tokens produced by only a single talker. In the present study, we are interested in the effects of multiple talkers on perception and memory.

Some related research has been reported in the literature. Several experiments have shown that when the talker’s voice changes from trial to trial,¹ vowel perception becomes impaired (Assmann, Nearey, & Hogan, 1982; Strange, Verbrugge, Shankweiler, & Edman, 1976; Summerfield, 1975; Summerfield & Haggard, 1973; Verbrugge et al., 1976; Weenink, 1986). Effects due to talker variability have also been demonstrated at the lexical level

Copyright 1989 by the American Psychological Association. Inc

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

¹We use the expression “changes in the talker’s voice” throughout the article to refer to variability in the production of specific test items spoken by *different* talkers. Although the term is potentially ambiguous, we are concerned primarily in this research with variability between talkers rather than variability within a specific talker.

(Allard & Henderson, 1975; Cole, Coltheart, & Allard, 1974; Creelman, 1957). Several recent experiments in our laboratory have examined the effects of talker variability on spoken word recognition (Mullennix, Pisoni, & Martin, 1989). We observed poorer identification of words when the talker's voice changed from trial to trial compared with when the talker's voice remained the same. Reliable effects were obtained in both perceptual identification and naming tasks.

Although our earlier research has demonstrated that talker variability affects perceptual processing, little research has studied the effects of talker variability on memory processes. One study conducted by Craik and Kirsner (1974) examined the effects of talker variability on recognition memory performance. They found that recognition of words was faster and more accurate when words were repeated in the same voice as the original study items. Their results suggest that information about a talker's voice can be retained in memory and that talker-specific cues may be used to facilitate recognition memory for spoken words. In another study, Geiselman and Bellezza (1976, 1977) demonstrated that long-term memory for a talker's voice is retained automatically, even if subjects are not instructed to attend to voice characteristics (see also Geiselman, 1979; Geiselman & Crawley, 1983). Thus, information about a talker's voice can be retained and transferred into long-term memory along with the to-be-recalled items.

Tulving and Colotla (1970) examined trilingual subjects' recall of lists of words spoken in one language compared with three languages. They found the recall of early list items was better when the lists contained words from only one language. Watkins and Watkins (1980) examined free recall of word lists spoken by either one or two talkers. They found an advantage for the recall of early list items if the word lists were produced by a single talker. The focus of their research was on recency effects, and they did not discuss the single-talker advantage in primacy in any detail.

In a more recent study, Mattingly, Studdert-Kennedy, and Magen (1983) examined the effects of changing the talker's voice on the serial recall of spoken word lists. The results indicated that recall performance for early list items was worse when the items were produced by different talkers with different dialects compared with lists produced by a single talker or by different talkers with the same dialect. Mattingly et al. (1983) suggested that changes in dialect, but not in the talker's voice within a dialect, affected the encoding and/or rehearsal of list items.

These memory experiments demonstrate that recall measures can be used to examine the processing of different types of word lists. Specifically, recall of early list items might provide information about the capacity demands required for the processing of words produced by different talkers. Primacy effects are assumed to reflect a greater number of rehearsals or more elaborative rehearsal devoted to early list items (e.g., Rundus, 1971). A number of theorists have suggested that increased amounts of rehearsal leads to a higher probability that an item will be transferred to long-term memory (Atkinson & Shiffrin, 1968; Waugh & Norman, 1965) or leads to stored images of greater strength, which are then more easily retrieved from memory (Shiffrin, 1970). Thus, recall of items from the primacy portion could be used as an index of the amount and/or efficiency of rehearsal given these items. Recency effects, on the other hand, have been assumed to reflect the output of items from a short-term memory buffer (Crowder, 1976; Glanzer, 1972; Waugh & Norman, 1965).

It is now well accepted by most memory theorists that short-term memory is limited in its capacity to hold and process information (Shiffrin, 1976). Different amounts of processing resources are available for a particular task, depending on how much capacity or attentional resources are allocated to other tasks. In recall tasks, a limited amount of processing

capacity is available for the encoding and rehearsal of stimulus items (Baddeley & Hitch, 1974). If the processing of words produced by different talkers requires greater capacity demands, fewer resources should therefore be available for the rehearsal of these items compared with items produced by a single talker. Because the recall of early list items is affected by the amount and/or efficiency of rehearsal, any differences in the capacity demands required to process multiple-talker versus single-talker word lists should produce lower recall performance in the primacy portion of the serial position curve.²

In the first experiment, serial recall of word lists was investigated. Ten-item word lists were constructed from words spoken by a single talker, 10 talkers of the same gender, or 5 male and 5 female talkers. Two multiple-talker conditions were included to determine whether increased acoustic-phonetic variability due to gender differences would affect recall performance. We predicted that talker variability would produce greater capacity demands for the processing of list items and that these effects would cascade up the processing system to affect the amount of rehearsal given items and the retrieval of early list items from long-term memory at the time of recall. Thus, we expected that recall performance for early list items would be worse when the talker's voice changed from item to item compared with when the talker's voice remained the same within a list. In addition, we predicted that no differences would be observed for terminal list positions. Because of these a priori predictions, when interactions between talker condition and serial position were obtained, we analyzed the recall data separately for three portions of the list corresponding to early, middle, and late serial positions.

Experiment 1

Method

Subjects—Subjects were 112 undergraduate students at Indiana University who participated to fulfill a course requirement. Each subject participated in one hour-long session. All subjects were native speakers of English who reported no history of a speech or hearing disorder at the time of testing.

Stimuli—The stimuli consisted of five lists of 10 monosyllabic English words. Words were originally recorded in isolation on audiotape and digitized via a 12-bit analog-to-digital converter using a PDP 11/34 computer. All word lists were generated from digital files stored in the computer. Three versions of each word list were prepared. In the single-talker condition, all items in a list were spoken by one talker. In the multiple-talker same-gender condition, the 10 items in each list were spoken by 10 different talkers of the same gender. And in the multiple-talker different-gender condition, the 10 items in each list were spoken by 5 different male and 5 different female talkers.

Overall RMS (root mean squared) amplitude levels for all words were digitally equated. Stimuli were low-pass filtered at 4.8 kHz and played to listeners through a 12-bit digital-to-analog converter over matched and calibrated TDH-39 headphones at 80 dB (SPL). The presentation of the word lists was controlled by a PDP 11/34 computer. Words within a list differed from one another by at least two phonemes. All of the words used in the experiment

²In this connection, it is worth pointing out that the increased capacity demands may affect the rehearsal process in several ways, either directly or indirectly, depending on the precise locus of the effects of talker variability. In the direct case, talker variability affects only the speed and efficiency of the rehearsal process itself. Changes in processing capacity due to increased difficulty in encoding items at the time of input are not passed up the system because items are assumed to be identified correctly at the time they enter working memory. In the indirect case, talker variability is assumed to affect the early perceptual encoding process so that the increased capacity demands needed for encoding reduce the available resources needed for subsequent rehearsal of the items. Finally, it is possible that talker variability affects *both* encoding and rehearsal and that increased processing demands are incurred by both processes concurrently.

had been previously tested for intelligibility in a separate experiment using a different group of listeners. All items received identification scores of 95% correct or above when presented in isolation.

Procedure—On each trial, subjects were presented with a spoken list of 10 words. They were then given 60 s to recall the words in the exact serial position in which they were presented. Subjects were free, however, to output their responses in any order. Subjects recorded their responses by writing them on a response sheet. Each response sheet contained 10 lines, numbered 1 through 10, corresponding to the 10 items presented in each list. Subjects were told that items not recalled in the correct position would be scored as incorrect responses.

The interword interval for stimulus presentation was 1.5 s. Immediately before the presentation of each list, subjects heard a brief 1000-Hz warning tone over their headphones. Following presentation of each list, another tone signaled the end of the list and the beginning of the 60-s recall period.

The talker variable was manipulated in a between-subjects design. Subjects were randomly assigned to one of three conditions: single talker, multiple talker/same gender, or multiple talker/different gender. Identical word lists were used in each condition; the conditions differed only in the voices used to produce the words in each list. Each subject heard four blocks of the five word lists for a total of 20 list presentations. The order of lists within each block and the order of stimuli within each list were randomized. Two practice lists were presented at the beginning of the experimental session in order to familiarize subjects with the experimental procedure.

Results and Discussion

Figure 1 shows the percentage of words correctly recalled as a function of serial position and talker condition, averaged over all trials.³ To assess differences between the talker conditions, a two-way analysis of variance (ANOVA) was conducted for the factors of talker condition and serial position. We analyzed serial position as a variable within the following analyses, but do not report post hoc tests for effects of serial position unless they interact with another variable. A main effect of talker was not observed. A significant main effect of serial position was obtained, $F(9, 981) = 334.1, p < .001$. A marginally significant interaction of talker and serial position was also obtained $F(18, 981) = 1.57, p < .06$.

In order to investigate the interaction between talker and serial position, three separate two-way ANOVAs for the factors of talker and serial position were conducted for the primacy (List Positions 1–3), middle (List Positions 4–7), and recency (List Positions 8–10) portions of the serial position curve.

In the primacy portion, a main effect of talker was obtained, $F(2, 109) = 4.41, p < .02$. Newman-Keuls tests revealed that recall of items from single-talker lists was significantly better than recall of items from either of the multiple-talker conditions. The two multiple-talker conditions did not differ from one another. A significant main effect of serial position was also obtained, $F(2, 218) = 380.1, p < .001$. The interaction of talker and serial position was not significant.

³These results were originally analyzed for the factor of block. Although recall performance improved across blocks, the pattern of results between the single-talker and multiple-talker conditions did not change as a function of block. For ease of exposition, all analyses reported in this article were therefore collapsed across blocks.

In the middle portion, the main effect of talker was not significant, and no significant interactions were obtained. A significant main effect of serial position was obtained, $F(3, 327) = 24.1, p < .001$.

Finally, in the recency portion of the serial position curve, the main effect of talker was not significant, and no significant interactions were obtained. A significant main effect of serial position was obtained, $F(2, 218) = 427.2, p < .001$.

As expected, recall of early list items from the single-talker lists was better than recall of early list items from the multiple-talker lists. We suggest that the processing of words from multiple-talker lists requires more limited capacity resources in working memory than the processing of words from single-talker lists. As a result, fewer processing resources are available for rehearsal of items from multiple-talker lists, leading to a lower probability that they will be successfully transferred to and subsequently retrieved from long-term memory (Baddeley & Hitch, 1974; Luce, Feustel, & Pisoni, 1983). Differences in recall of early list items may thus reflect differential capacity demands for the processing of these items. Evidently some processing cost is associated with talker variability. In order to investigate the nature of the capacity demands involved in the recall of single-talker and multiple-talker word lists, a second experiment was conducted.

Experiment 2

In this experiment, processing demands in short-term memory were manipulated with a preload memory task (Baddeley & Hitch, 1974). A sequence of digits was presented visually on a cathode-ray tube (CRT) display prior to the auditory presentation of each word list. Subjects were required to recall the visual digits *and* the spoken word lists.

Three preload conditions were included: zero-digit preload, three-digit preload, and six-digit preload. The zero-digit preload condition provided an opportunity to replicate the results of Experiment 1. The digit recall data from the three-digit and six-digit preload conditions served as the primary experimental conditions to assess whether the rehearsal of items in working memory would be differentially affected by the processing demands of multiple-talker word lists. We predicted that more visually presented digits would be recalled in the single-talker condition than in the multiple-talker condition. This prediction was based on the assumption of a limited-capacity system in which both spoken items and visually presented digits share and compete for common processing resources in working memory.

Method

Subjects—Subjects were 72 volunteers from the Bloomington, Indiana, area. Subjects participated in one hour-long session and were paid \$4 for their participation. All subjects were native speakers of English who reported no history of a speech or hearing disorder at the time of testing.

Stimuli—The spoken words used in this experiment were a subset of the stimuli used in Experiment 1. Five lists of 10 monosyllabic words were used for each of two talker conditions: single talker and multiple talker. In the multiple-talker condition items were produced by five different male and five different female talkers. All aspects of the stimuli used in Experiment 2 remained the same as in the first experiment.

Procedure—The experimental procedure was identical to that used in Experiment I, with the exception that a memory preload task was included to increase capacity demands. Prior to the presentation of each spoken word list, subjects in the preload conditions saw either three or six digits presented sequentially on a CRT monitor located directly in front of them.

On each trial, digits were sampled without replacement from the digits 1 through 9. Each digit remained on the CRT screen for 2 s, with a 1-s interdigit interval. The placement of warning tones was the same as in Experiment 1, except that an additional tone was added in the three-digit preload and six-digit preload conditions to alert subjects to the beginning of the digit presentation. During the recall interval, subjects first wrote down the series of digits and then wrote down as many of the spoken words as possible. In order to ensure that subjects in the preload conditions actively maintained the digits in memory during presentation of the word lists, they were explicitly told that none of the words would be counted as correct unless all of the visually presented digits were correctly recalled in the exact order in which they were presented.

The talker variable was manipulated in a between-subjects design. Subjects were randomly assigned to one of two talker conditions: single talker or multiple talker. Memory preload was also manipulated between subjects. Subjects were randomly assigned to one of three preload conditions: zero-digit, three-digit, or six-digit preload.

Results and Discussion

Word recall and digit recall were examined separately as dependent variables. The presentation and discussion of the data are divided into two parts for ease of exposition.

Preload digit recall—Because no digits were actually presented in the zero-digit preload condition, only the three-digit and six-digit preload conditions were analyzed. Digits were scored as correct if, and only if, they were recalled in the exact serial position in which they were presented.

For the three-digit preload groups, more digits were recalled in the single-talker condition (89.4%) than the multiple-talker condition (84.3%). For the six-digit preload groups, more digits were recalled in the single-talker condition (82%) than the multiple-talker condition (72.6%).

A two-way ANOVA was conducted on the digit recall data for talker and preload condition. The analysis revealed significant main effects of talker, $F(1, 44) = 4.91, p < .03$, and preload condition, $F(1, 44) = 8.49, p < .01$, on digit recall. The interaction between the two variables was not significant, although there was a strong trend for the differences between the two talker conditions to become larger at the higher preload.⁴

Spoken word recall—The percentage of words correctly recalled as a function of talker condition, preload condition, and serial position is shown in Table 1.

A three-way ANOVA was conducted on the word recall data for talker, preload condition, and serial position. A main effect of talker was not obtained. A significant main effect of serial position was obtained $F(9, 585) = 180.6, p < .001$. Post hoc tests revealed that fewer words were recalled as preload increased. The interaction of talker and preload condition was not significant. Significant interactions of talker and serial position, $F(9, 585) = 2.0, p < .04$, and preload condition and serial position, $F(18, 585) = 2.86, p < .01$, were obtained. The three-way interaction was not significant.

In order to investigate these interactions, separate three-way ANOVAS were conducted for the primacy (List Positions 1–3), middle (List Positions 4–7), and recency (List Positions 8–10) portions of the serial position curve.

⁴An arcsine transformation was performed on the digit recall data. Analysis of the transformed data did not change the pattern of results obtained.

In the primacy portion (Positions 1–3), a marginally significant main effect of talker was obtained, $F(1, 65) = 3.9, p < .06$. Better recall was observed for early list items in the single-talker condition compared with the multiple-talker condition. A significant main effect of preload condition was also obtained, $F(2, 65) = 4.37, p < .02$. Newman-Keuls tests revealed that the percentage of words recalled was reliably greater in the zero-digit preload condition compared with the other two conditions and that the percentage of words recalled in the three-digit condition was greater than in the six-digit condition. A significant main effect of serial position was also obtained, $F(2, 130) = 242.7, p < .001$. No significant interactions were obtained in this analysis.

In the middle portion (Positions 4–7), a significant main effect of talker was not observed. However, a significant main effect of preload condition was found, $F(2, 65) = 11.2, p < .001$. Newman-Keuls tests revealed that word recall was higher in the zero-digit preload condition than in the other two preload conditions. Word recall did not differ between the three-digit and six-digit preload conditions for items in the middle portion of the curve. A significant main effect of serial position was also obtained, $F(3, 195) = 4.96, p < .01$. No significant interactions were obtained.

In the recency portion (Positions 8–10), a main effect of talker was not observed. However, a significant main effect of preload condition was obtained, $F(2, 65) = 30.4, p < .001$. Newman-Keuls tests revealed that recall was higher in the zero-digit preload condition compared with the other two preload conditions. Recall was also higher in the three-digit condition than in the six-digit condition. A significant main effect of serial position was also obtained, $F(2, 130) = 376.9, p < .001$. No significant interactions were obtained.

Taken together, the digit recall and word recall data suggest that the processing of spoken items from multiple-talker word lists requires more resources in working memory, and the allocation of these additional resources interferes with the rehearsal of items in working memory. The digit recall data provide evidence that the rehearsal of preload items was impaired by the subsequent presentation of spoken word lists produced by different talkers. It is unlikely that the effect of talker variability on digit recall could be due to differences in encoding of visually presented digits because these items were presented to subjects *before* the spoken word lists were presented. In addition, it is also unlikely that the effect of talker variability on digit recall could be due to differences produced by the retrieval of items from single-talker and multiple-talker lists because all of the visually presented digits were recalled *before* the spoken words were recalled. Thus, any effects must be due to competition for common processing resources between the two sets of items.

The increased demands and competition for processing resources for multiple-talker lists clearly affected the rehearsal and subsequent transfer of digit items into long-term memory. Talker variability evidently affects the rehearsal of information in working memory.

Experiment 3

The results of the first two experiments suggest that the processing of spoken word lists produced by different talkers requires greater resources in working memory than word lists produced by only a single talker. Recall of items in the primacy portion of the serial position curve could be affected, however, by variables other than those related to the allocation of processing resources in working memory. It is possible that the differences obtained between single-talker and multiple-talker conditions might reflect differences in search and retrieval processes that occur independently of any capacity demands incurred during the encoding and rehearsal of specific items. As noted earlier, there is some evidence that a representation of a talker's voice is retained in long-term memory and can be used to

facilitate the retrieval of items in recognition memory tasks (Craik & Kirsner, 1974; Geiselman & Bellezza, 1976, 1977).

If cues to a talker's voice are transferred into long-term memory along with associated item and order information, they may differentially affect search and retrieval processes at the time of recall. In immediate recall paradigms, talker-specific acoustic information from terminal list items may be available in working memory at the time of recall. This information might be used to facilitate the retrieval of early list items from long-term memory. If talker-specific information can be used to facilitate search of long-term memory, memory search may be more efficient when the voice characteristics of only one talker are used. Facilitation of search and retrieval may not occur, or may not be as efficient, when list items are spoken by different talkers. In this case, voice characteristics associated with terminal items will not match the characteristics of the talkers that produced early items.

Experiment 3 assessed recall performance when voice cues in working memory were eliminated. If the differences are due to the use of talker-specific cues in working memory during search and retrieval, then differences in recall in the primacy portion between the two talker conditions should be reduced when talker-specific information in working memory is eliminated by an interference task. In this experiment we combined a retroactive interference task (Peterson & Peterson, 1959) with the recall paradigm employed in Experiment 1. Subjects were presented with a list of spoken words for serial recall and then were asked to perform an arithmetic task that was designed to eliminate the contents of working memory before recall. The interference task thus forced subjects to rely on long-term memory for the recall of list items.

Method

Subjects—Subjects were 108 undergraduates at Indiana University who volunteered to fulfill a course requirement. Each subject participated in one hour-long session. All subjects were native speakers of English who reported no history of a speech or hearing disorder at the time of testing.

Stimuli—The stimuli used in Experiment 2 were also used in Experiment 3. All aspects of the stimuli remained exactly the same.

Procedure—The experimental procedure was identical to Experiment 1, except that a retroactive interference task was inserted after each word list. Subjects saw a three-digit number presented visually on a CRT monitor located in front of them. The three digits in each number were randomly sampled without replacement from the digits 1 through 9 and were presented simultaneously on the CRT monitor. Subjects were required to silently count backwards by threes from this number, subtracting three every time they heard a signal tone over their headphones. The tones occurred at 2-s intervals after the presentation of the three-digit number. The end of the arithmetic task was signaled by the presentation of two sequential tones. After subjects heard the two tones, they were required to write down the number they currently had in memory from the subtraction task. When this was completed, subjects were required to recall the words in the serial order in which they were presented. Subjects were told that their recall of the words would be counted as correct if, and only if, the items were recalled in the correct serial position. In order to ensure that subjects paid full attention to the arithmetic task, they were also told that their responses could not be scored unless they also produced the correct number from the subtraction task at the beginning of each recall period.

Talker variability was manipulated in a between-subjects design. Subjects were randomly assigned to one of two conditions: single talker or multiple talker. The length of the

retroactive interference interval was also manipulated as a between-subjects variable to produce three retention conditions: 4-s, 8-s, and 12-s retention.

Results and Discussion

Table 2 shows the percentage of words correctly recalled as a function of talker, retention interval, and serial position. The data obtained in Experiment 1 are included here as the zero-second interference condition and were used in the statistical analysis as the immediate recall condition in the design.

Inspection of Table 2 reveals that recall of items in the primacy portion was consistently higher for single-talker word lists compared with multiple-talker word lists across all retention intervals, thus replicating again our earlier findings.

A three-way ANOVA was conducted for talker, retention interval, and serial position. A significant main effect of talker was obtained $F(1, 128) = 14.7, p < .001$. Recall of items from single-talker lists was superior to recall of items from multiple-talker lists. A significant main effect of retention interval was also obtained $F(3, 128) = 32.5, p < .001$. Recall performance decreased as the duration of the retention interval increased. A significant main effect of serial position was also obtained, $F(9, 1152) = 333.3, p < .001$. Talker and retention interval did not interact. Significant interactions of serial position with interference condition, $F(27, 1152) = 12.6, p < .001$, and talker and serial position were also obtained, $F(9, 1152) = 11.1, p < .001$. The three-way interaction was not significant.

To investigate these interactions, separate three-way ANOVAS were carried out for the primacy, middle, and recency portions of the serial position curve. In the primacy portion (Positions 1–3), a significant main effect of talker was obtained, $F(1, 128) = 52.9, p < .001$. Recall of items from early list positions was higher for single-talker lists than multiple-talker lists. A significant main effect of serial position was also obtained, $F(2, 256) = 474.4, p < .001$. A main effect of retention interval was not significant, nor were any of the interactions significant. Thus, the superior recall of items from single-talker word lists did not change reliably as a function of the retention interval in the interference task.

In the middle portion (Positions 4–7), the main effect of talker was not significant. However, a significant effect of retention interval was observed, $F(3, 128) = 8.26, p < .001$. Newman-Keuls tests revealed that recall was higher in the immediate recall condition than in any of the other conditions. Recall was higher in the 4-s retention condition than in the 8-s and 12-s retention conditions. Recall in the 8-s and 12-s conditions did not differ reliably from each other.

In the recency portion (Positions 8–10), the main effect for talker was not significant. As expected, however, a highly significant effect of the retention interval was obtained, $F(3, 128) = 70.8, p < .001$. Newman-Keuls tests revealed that recall in the immediate recall condition was superior to all other conditions and that recall in the 4-s retention interval was better than recall in the 8-s and 12-s retention intervals. The 8-s and 12-s conditions did not differ reliably from each other. A significant main effect of serial position was also obtained, $F(2, 256) = 330.5, p < .001$. No significant interactions were obtained.

The results of Experiment 3 demonstrate that the length of the retention interval did not reliably affect the difference in the recall of early list items between the two talker conditions. As expected, the interference task did reduce performance for terminal list items, confirming the prediction that the arithmetic task would reduce the amount of information available in working memory at the time of recall. These results suggest that the advantage

in recall of early list items for lists produced by a single talker is not due to differential use of talker-specific acoustic information in working memory at the time of recall.⁵

General Discussion

The results of the present set of experiments provide evidence to support the hypothesis that the processing of words produced by different talkers requires more resources in working memory than the processing of words produced by a single talker. In three separate experiments, we observed decreases in recall of early list items for multiple-talker word lists. In addition, subjects recalled more visually presented preload digits when the digits were followed by a word list spoken by a single talker than when the word list was spoken by different talkers. Finally, the use of an interference task showed that differences in recall for early list items were not due to the use of talker-specific cues in working memory at the time of recall.

The present findings are consistent with other results showing that stimulus variability reduces the recall of early list items in memory tasks (Tulving & Colotla, 1970; Watkins & Watkins, 1980). The present findings differ, however, from those obtained by Mattingly et al. (1983), who reported that talker variability did not reduce recall performance for early list items. In their study, the digit names were used as stimuli. Digits are a highly overlearned set of stimulus materials, and it is likely that subjects used these constraints to improve their performance in the more demanding conditions. Moreover, Mattingly et al. (1983) used only three talkers in their multiple-talker condition, which may have prevented them from observing any effects of talker variability on recall.

The processing capacity explanation suggested here is consistent with a number of recent perceptual findings. Earlier work has demonstrated that identification performance is adversely affected when the talker's voice changes from trial to trial (Mullennix et al., 1989). Another more recent study has provided evidence that the perceptual deficits produced by talker variability are related to failures in selective attention. Using the Garner (1974) speeded classification task, Mullennix and Pisoni (1987) found that a stimulus dimension related to the talker's voice and a stimulus dimension related to the phonetic feature of voicing were perceived as integral perceptual dimensions. Thus, when phonetic information was attended to, voice information could not be selectively ignored, and vice versa. Mullennix and Pisoni (1987) also found that when the number of talkers was increased, selective attention was disrupted to an even greater degree, as shown by consistent increases in response latencies. Thus, as talker variability increased, processing demands increased. These results suggest that processing information about a talker's voice is mandatory in Fodor's sense (Fodor, 1983).

The precise nature of the differences in capacity demands between single-talker and multiple-talker word lists is unclear at this time, although several accounts are worth considering. One possibility is that more processing resources are needed only for the initial perceptual encoding of words produced by different talkers, resulting in fewer available resources for subsequent rehearsal of the words. Differences in the amount, efficiency, or

⁵The interference task was silent and may not have removed auditory information from an acoustical store, such as Crowder and Morton's (1969) PAS, by the time of recall. Nevertheless, the data from Experiment 3 strongly suggest that acoustic information was not used to facilitate the search and retrieval of single-talker items in the present set of experiments. To the extent that the interference task eliminated the contents of short-term memory, recall performance reflected retrieval from long-term memory only. If acoustic information facilitated the retrieval of single-talker items from long-term memory, this would have produced a single-talker advantage across all list positions in Experiment 3. Because talker differences were confined to early list items, we conclude that information from a sensory-based acoustic store did not facilitate the retrieval of single-talker items.

speed of rehearsal could thus be due to initial differences in encoding these items at the time of input.

A second possibility is that talker variability does not affect the speed or efficiency of initial encoding processes, but rather affects only the efficiency of rehearsal processes that operate after the stimulus items have been encoded into working memory. In this case, more processing resources would be needed only for the rehearsal of multiple-talker items because of uncertainty about who the talker was.

It is useful to compare the present findings with those reported by Luce et al. (1983), who examined the recall of natural and synthetic word lists by using the same methodology. Luce et al. found that recall for synthetically produced words was lower than recall for naturally produced words across all serial positions. In addition to the main effect of voice, they also observed an interaction between voice and serial position. The differences in recall between natural and synthetic speech were largest in early list positions and reduced in the middle and terminal positions. Luce et al. explained the reduced recall across all serial positions for synthetic speech by assuming that the synthetic speech produced a much larger number of encoding errors at the time of input. In the present experiments, it appears that talker variability did not produce encoding errors at the time of input because recall of items from the recency portion of the curve was comparable for single- and multiple-talker lists. Talker variability therefore appears to increase the demands for processing resources required for the rehearsal of list items. The differences in recall in the primacy portion of the serial position curve obtained by Luce et al. are similar to the present results and reflect, in our view, greater capacity demands affecting rehearsal of synthetic speech compared with natural speech.

In summary, the present results demonstrate that talker variability has significant consequences for the allocation of resources in serial recall tasks. Compensation for stimulus variability between talkers requires additional processing capacity and appears to interfere with the ability of subjects to actively maintain information in working memory and subsequently to transfer items into long-term memory. Our results are consistent with the hypothesis that speech perception utilizes a resource-demanding talker normalization mechanism that compensates for the stimulus variability produced by different talkers. Compensation for talker variability is not capacity free and apparently has consequences not only for the perception of speech but also for memory processes that operate on these perceptual representations. The present findings raise important questions about our current understanding of rehearsal processes in working memory, the transfer of speech from working memory into long-term memory, and the subsequent retrieval of these representations at the time of recall. Although it may appear at first glance that listeners perceive speech from different talkers without cost, the present set of findings, taken together with earlier perceptual data, suggests that this is only an illusion, and like many other cognitive processes, a number of complex processing activities actually underlie what we observe informally on the surface.

Acknowledgments

This research was supported by National Institutes of Health (NIH) Research Grant NS-12179-11 and NIH Training Grant NS-07134-09 to Indiana University. The authors thank Paul A. Luce for his comments and editorial suggestions on an earlier version of this article.

References

- Allard F, Henderson L. Physical and name codes in auditory memory: The pursuit of an analogy. *Quarterly Journal of Experimental Psychology*. 1976; 28:475–482. [PubMed: 1005652]

- Assmann PF, Nearey TM, Hogan JT. Vowel identification: Orthographic, perceptual, and acoustic aspects. *Journal of the Acoustical Society of America*. 1982; 71:975–989. [PubMed: 7085986]
- Atkinson, RC.; Shiffrin, RM. Human memory: A proposed system and its control processes. In: Spence, KW.; Spence, JT., editors. *The psychology of learning and motivation*. Vol. 2. New York: Academic Press; 1968. p. 89–105.
- Baddeley, AD.; Hitch, GJ. Working memory. In: Bower, GH., editor. *The psychology of learning and memory*. Vol. 8. New York: Academic Press; 1974. p. 47–89.
- Cole RA, Coltheart M, Allard F. Memory of a speaker's voice: Reaction time to same- or different-voiced letters. *Quarterly Journal of Experimental Psychology*. 1974; 26:1–7. [PubMed: 4814860]
- 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.
- Crowder, RG. *Principles of learning and memory*. Hillsdale, NJ: Erlbaum; 1976.
- Crowder RG, Morton J. Precategorical acoustic storage (PAS). *Perception & Psychophysics*. 1969; 5:365–373.
- Fodor, JA. *The modularity of mind*. Cambridge, MA: MIT Press; 1983.
- Garner, WR. *The processing of information and structure*. Potomac, MD: Erlbaum; 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 and Verbal Behavior*. 1983; 22:15–23.
- Glanzer, M. Storage mechanisms in recall. In: Bower, GT.; Spence, JT., editors. *The psychology of learning and motivation*. Vol. 5. New York: Academic Press; 1972. p. 129–193.
- Joos MA. Acoustic phonetics. *Language*. 1948; 24(Suppl 2):1–136.
- Luce PA, Feustel TC, Pisoni DB. Capacity demands in short-term memory for synthetic and natural speech. *Human Factors*. 1983; 25:17–32. [PubMed: 6840769]
- Mattingly IG, Studdert-Kennedy M, Magen H. Phonological short-term memory preserves phonetic detail. *Journal of the Acoustical Society of America*. 1983; 73:S4.
- Mullennix, JW.; Pisoni, DB. Research on Speech Perception Progress Report 13. Bloomington, IN: Department of Psychology, Indiana University; 1987. Talker variability effects and processing dependencies between word and voice.
- 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]
- Peterson LJ, Peterson MJ. Short-term retention of individual verbal items. *Journal of Experimental Psychology*. 1959; 58:193–198. [PubMed: 14432252]
- Rundus D. Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*. 1971; 89:43–50.
- Shiffrin, RM. Memory search. In: Norman, D., editor. *Models of human memory*. New York: Academic Press; 1970. p. 375–447.
- Shiffrin, RM. Capacity limitations in information processing, attention, and memory. In: Estes, WK., editor. *Handbook of learning and cognitive processes*. Vol. 4. Hillsdale, NJ: Erlbaum; 1976. p. 177–236.
- Strange W, Verbrugge RR, Shankweiler DP, Edman TR. Consonant environment specifies vowel identity. *Journal of the Acoustical Society of America*. 1976; 60:213–224. [PubMed: 956528]
- Summerfield, Q. Report on Research in Progress in Speech Perception. Vol. 2. Belfast, Northern Ireland: The Queen's University of Belfast; 1975. Acoustic and phonetic components of the influence of voice changes and identification times for CVC syllables; p. 73–98.

- Summerfield, Q.; Haggard, MP. Report on Research in Progress in Speech Perception. Vol. 2. Belfast, Northern Ireland: The Queen's University of Belfast; 1973. Vocal tract normalisation as demonstrated by reaction times; p. 1-12.
- Tulving E, Colotla V. Free recall of trilingual lists. *Cognitive Psychology*. 1970; 1:86–98.
- Verbrugge RR, Strange W, Shankweiler DP, Edman TR. What information enables a listener to map a talker's vowel space? *Journal of the Acoustical Society of America*. 1976; 60:198–212. [PubMed: 956527]
- Watkins OC, Watkins MJ. Echoic memory and voice quality: Recency recall is not enhanced by varying presentation voice. *Memory & Cognition*. 1980; 8:26–30.
- Waugh NC, Norman DA. Primary memory. *Psychological Review*. 1965; 72:89–104. [PubMed: 14282677]
- Weenink, DJM. The identification of vowel stimuli from men, women, and children. Proceedings from the Institute of Phonetic Sciences of the University of Amsterdam; 1986. p. 41-54.

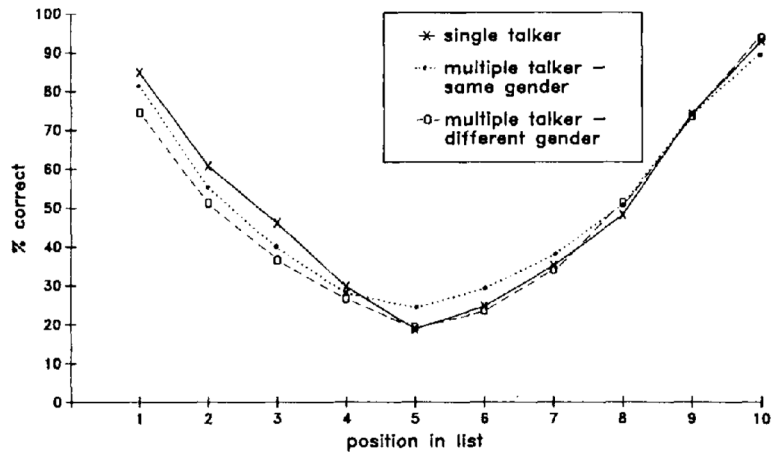


Figure 1. Mean percent correct serial recall collapsed over subjects as a function of serial position and talker condition for Experiment 1.

\$watermark-text

\$watermark-text

\$watermark-text

Table 1
 Percent Word Recall for Experiment 2 by Subject Condition and Serial Position

Condition	Serial position									
	1	2	3	4	5	6	7	8	9	10
Preload = 0										
Single	79.9	59.8	45.7	32.0	27.9	30.6	42.0	54.3	84.0	90.3
Multiple	70.8	50.0	35.0	27.9	25.4	27.5	37.9	51.3	75.8	95.4
Preload = 3										
Single	81.4	50.9	37.7	28.6	18.2	18.6	21.4	26.4	50.9	86.4
Multiple	63.8	48.8	30.0	22.9	19.6	16.7	21.3	32.1	57.9	85.4
Preload = 6										
Single	70.0	39.6	25.8	17.9	10.8	10.0	7.9	18.3	39.6	70.0
Multiple	61.2	37.7	22.3	17.3	13.8	15.0	19.6	24.2	43.5	72.7

Table 2
 Percent Word Recall for Experiment 3 by Subject Condition and Serial Position

Condition	Serial position									
	1	2	3	4	5	6	7	8	9	10
No retention interval										
Single	85.2	60.7	46.1	30.1	21.1	24.7	35.5	48.8	74.2	92.9
Multiple	74.6	51.4	36.4	26.6	19.3	23.6	34.1	51.4	73.4	93.4
4-s retention interval										
Single	83.3	60.0	42.9	30.8	16.7	11.7	17.5	30.8	47.9	73.3
Multiple	63.8	41.7	30.0	20.4	17.1	14.2	18.3	27.5	50.0	70.4
8-s retention interval										
Single	85.0	61.2	43.0	31.0	17.1	10.0	15.3	26.0	36.7	61.7
Multiple	61.7	40.8	27.9	18.8	14.2	10.8	18.3	26.2	36.7	61.7
12-s retention interval										
Single	84.0	61.1	45.2	32.5	18.0	12.2	17.0	25.0	29.9	50.1
Multiple	62.9	40.8	30.0	26.2	14.2	10.8	17.1	24.2	30.0	47.9