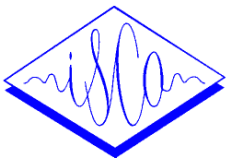

2nd International Conference on
Spoken Language Processing (ICSLP 92)
Banff, Alberta, Canada
October 12-16, 1992

EFFECTS OF SPEAKING RATE AND TALKER VARIABILITY ON THE REPRESENTATION OF SPOKEN WORDS IN MEMORY

Lynne C. Nygaard, Mitchell S. Sommers, and David B. Pisoni

Speech Research Laboratory, Indiana University
Bloomington, Indiana 47405

ABSTRACT

The present paper reports a series of experiments designed to investigate the nature of perceptual compensation and the memory representations for spoken words produced at different speaking rates. The aim was to determine if variability in speaking rate has consequences for the encoding and processing of spoken words and if these consequences are comparable to those found for talker variability. A serial recall task was used to study the effects of changes in speaking rate and talker variability on the initial encoding, rehearsal, and recall of lists of spoken words. Presentation rate was manipulated to determine the time course and nature of processing. The results indicate that at fast presentation rates, variations in both speaking rate and talker characteristics incur a processing cost which influences the initial encoding and subsequent rehearsal of spoken words. At slower presentation rates, however, variation in talker results in improved recall in initial list positions while variation in speaking rate has no effect on recall performance. These results suggest that the processing of variability due to changes in speaking rate and talker differences may be the result of distinct operations. Talker information appears to be integrated into long-term representations of spoken words while rate information may be discarded or lost after initial stages of processing.

10.21437/ICSLP.1992-46

I. INTRODUCTION

One of the fundamental problems confronting theories of speech perception is how to characterize the listener's ability to extract consistent phonetic percepts from a highly variable acoustic signal. Factors such as phonetic context [1], linguistic stress, and utterance length [2], for example, can all have profound effects on the acoustic realization of linguistic units. The consequences of these many sources of variability for speech perception is that phonetic segments do not necessarily have any invariant acoustic form [3]. Rather, stable linguistic units or equivalence classes that listeners report stem from an acoustic signal that may be anything but stable. The purpose of the present series of experiments was to examine in detail two of these sources of variability--changes in talker characteristics and changes in speaking rate. Our aim was to assess the consequences of changes along each of these dimensions for perceptual processing and the subsequent representation of spoken words in human memory.

Traditionally, accounts of speech perception have characterized variation in the acoustic speech signal as a perceptual problem that perceivers must solve [4]. Listeners are thought to achieve consistent phonetic percepts given variation in talker and rate through some kind of normalization process in which linguistic units are evaluated relative to the prevailing rate of speech [5, 6] or relative to specific talker characteristics [7, 8]. Indeed, several recent studies have shown that listeners are sensitive to changes in talker characteristics and in speaking rate.

Consider the perceptual consequences of variation in talker characteristics. Summerfield and Haggard [9] as well as Mullennix et al. [10] have shown that phoneme and word recognition performance is poorer when listeners are presented with words produced by multiple talkers compared to words produced by a single talker. Likewise, using a Garner [11] speeded classification task, Mullennix and Pisoni [12] reported that subjects had difficulty ignoring irrelevant variation in a talker's voice when asked to classify words by initial phoneme. Taken together, these findings suggest that variations due to changes in talker characteristics are time and resource demanding. Further, the processing of talker information appears to be inseparable from the processing of the phonetic content of the message.

Additional research has suggested that talker variability can affect memory processes as well. At relatively fast presentation rates, Martin et al. [13] and Goldinger et al. [14] found that serial recall of spoken words is better in initial list positions when all the words in the list are produced by a single speaker compared to a condition in which each word is produced by a different speaker. Interestingly, at longer presentation rates, Goldinger et al. [14] found that recall of multiple-talker lists is superior to single-talker lists. These results suggest that at fast presentation rates, variation due to changes in the talker affects the initial encoding and subsequent rehearsal of items in the to-be-remembered lists. At slower presentation rates, on the other hand, listeners are able to fully process and encode each word along with the concomitant talker information. Consequently, listeners are able to use the additional talker information to aid in their recall task.

Further evidence that talker information is encoded and retained in a long-term episodic memory store comes from recent experiments conducted by Palmeri et al. [15]. Using a continuous recognition memory procedure, specific voice information was shown to be retained along with word information and these attributes were found to aid later recognition. The finding that subjects are able to use talker-specific information suggests that this source of variation may not be discarded or normalized in the process of speech perception, as widely assumed in the literature. Rather, variation in a talker's voice may become part of a rich and highly detailed representation of the speaker's utterance. The decrements in performance due to talker variability would then be due to the additional attention and resources necessary to encode the indexical information conveyed by a talker's voice.

The purpose of the present set of experiments was two-fold. First, we wanted to replicate the earlier findings on the effects of talker variability on perceptual and memory processes involved in serial recall. Second, we wanted to extend the investigation to another source of variability -- variability due to speaking rate. An extensive body of research suggests that listeners are susceptible to changes in speaking rate as well. Speaking rate affects performance in phoneme identification [5, 6] as well as in perceptual identification tasks [16]. In addition, Miller and Volaitis [17] have shown recently that phonetic category boundaries and accompanying phonetic category structures that rely on temporal information are quite sensitive to relative changes in rate of articulation.

Given the effects of overall speaking rate on the perceptual processing of temporal contrasts, the question arises whether rate variability would have effects similar to those found for talker variability in terms of processing time and resources. That is, do changes in speaking rate incur processing costs that draw upon a limited pool of resources and further, is speaking rate encoded into long-term memory representations in the same manner as talker variability? To answer these questions, we sought to assess the consequences for perceptual processing and memory representation of changes in rate of articulation.

Our first step was to replicate and confirm the effects of talker variability on the serial recall of spoken words at three different presentation rates. Recall of lists of words produced by a single

talker was compared to recall of lists produced by multiple talkers. Words were presented at 100, 1000, or 4000 ms inter-word intervals. Our second step was to extend the investigation to speaking rate. Recall of lists produced at multiple speaking rates was compared to lists produced at a single speaking rate at the same three presentation rates used for the voice manipulation. This experimental design allowed us to compare the effects of variability due to talker with the effects of variability in speaking rate on the serial recall of spoken words. In addition, using the presentation rate manipulation, we were interested in determining if variability due to speaking rate, like talker information, aids serial recall at slower presentation rates.

II. METHOD

2.1 Subjects

One hundred and eighty undergraduate students enrolled in introductory psychology courses served as subjects. They were given partial course credit for their participation. All subjects were native speakers of American English and reported no history of speech or hearing disorders.

2.2 Stimuli

The stimuli consisted of a set of 100 monosyllabic words drawn from phonetically balanced (PB) word lists. To obtain a database of words produced by different talkers at different speaking rates, each word was embedded in the carrier phrase "The next word is _____" for presentation to speakers. Ten speakers were asked to pronounce each phrase at three different speaking rates -- slow, medium, and fast -- for a total of 3000 words (100 words x 10 speakers x 3 rates). The words were digitized on-line and subsequently edited from the carrier phrase for presentation. To ensure that variations in articulation rate were perceptually salient, rate judgments were collected for the complete set of words from a separate group of listeners. For each speaker's utterance, subjects were asked to judge whether the words were produced at a fast, medium, or slow rate. Percent correct as defined by the percentage of times subjects chose a rate that corresponded to the intended rate of the talker was 83, 81, and 75 for slow, medium, and fast words respectively. In addition, durations of words produced at each rate by each speaker were measured. The durations for slow, medium, and fast words averaged across speakers were 903, 564, and 383 msec, respectively. Thus, the rate judgments and measured durations confirm that the stimulus materials included a wide range of articulation rates and that this variation was perceptually salient to listeners.

From this original set of 3000 words, 8 ten-word lists were constructed for each condition. In the multiple-talker condition, each word in a list was produced by a different talker. Likewise, in the multiple-speaking rate condition, words were selected from slow, medium, and fast items such that each word in a list was produced at a different speaking rate. In the single-talker and single-rate condition, all words were produced by a single talker at a normal speaking rate.

2.3 Procedure

Subjects were tested in groups of 6 or fewer in a quiet testing room. Stimuli were presented over matched and calibrated TDH-39 headphones at approximately 80 dB (SPL) and were low-pass filtered at 4.8 kHz. A PDP-11/34 computer was used to control the experiment in real time.

During the experiment, subjects first heard a 500 ms, 1000 Hz warning tone to alert them that a list was about to be presented. Then, a list of ten words was presented at one of three rates--words were presented with either 100, 1000, or 4000 msec between items in the list. After the list had been presented, another warning tone sounded indicating the beginning of the recall period. Subjects had 60 seconds in which to recall the words in the list. A third tone signaled the end of the recall period. Subjects were instructed to recall the words in the exact order in which they were presented.

Talker, speaking rate, and presentation rate were all between-subjects variables. Of the 180 subjects, sixty were tested at each presentation rate. Of those sixty, twenty were tested in the multiple-talker, twenty in the multiple-speaking rate, and twenty in the single-talker/single-speaking rate conditions. The same words were heard by all subjects. The variables that changed across conditions were the number of talkers, the number of speaking rates, and the presentation rate.

III. RESULTS

Figure 1 shows the effects of talker variability on serial recall at the three presentation rates. Percent correct recall is plotted as a function of serial position for single- versus multiple-talker conditions. The top panel shows serial recall performance for multiple-talker and single-talker lists presented at the 100 ms ISI. The middle panel shows recall performance from the 1000 ms ISI and the bottom panel shows the results from the 4000 ms ISI. An overall ANOVA using talker, serial position, and presentation rate as main effects was conducted on the number of correct responses. As expected, a significant main effect of serial position was found showing reliable primacy and recency effects in recall ($p<.001$). In addition, a significant main effect of presentation rate was found ($p<.005$). Recall performance was better overall at slower presentation rates. Additional analyses were conducted separately for early, middle, and late list positions. These analyses revealed a significant interaction between presentation rate and talker variability at early list positions ($p<.05$), but not for middle and late list positions. Recall of words from the early portions of multiple-talker lists was affected more by changes in presentation rate than recall of words from the single-talker lists.

Figure 2 shows the effects of variations in speaking rate on serial recall at the three presentation rates. Again, percent correct is plotted as a function of serial position. The top panel shows recall performance for multiple-speaking rate and single-speaking rate lists presented at the 100 ms ISI. The middle panel shows recall performance from the 1000 ms ISI condition and the bottom panel shows the results from the 4000 ms ISI. An overall ANOVA (speaking rate x serial position x presentation rate revealed main effects for presentation rate ($p<.005$), serial position ($p<.001$) and speaking rate ($p<.05$). Recall performance was better overall at slow presentation rates than at fast presentation rates; recall was better at early and late list positions; and finally, recall was better overall for the single-rate lists than the multiple-rate lists. In addition, a significant three-way interaction ($p<.05$) was found indicating that a larger difference was obtained between multiple- and single-rate lists at the fast presentation rates than at the slow presentation rate. Additional analyses were conducted separately for early, middle, and late list positions. The results revealed a significant interaction ($p<.05$) between speaking rate and presentation rate, but only for early list positions. This finding indicates that the difference between single- and multiple-rate lists was smaller at the slow presentation rate than at the fast presentation rates.

The critical difference between the talker variability conditions and the rate variability conditions can be found at the slowest presentation rate. Posthoc analyses revealed a difference between multiple-talker and single-talker conditions, but no difference between multiple-rate and single-rate conditions in initial list positions. Recall performance was better for multiple-talker lists than for single-talker lists at the slow presentation rate. However, a comparable benefit of variation in speaking rate was not found in the primacy portion of the curve.

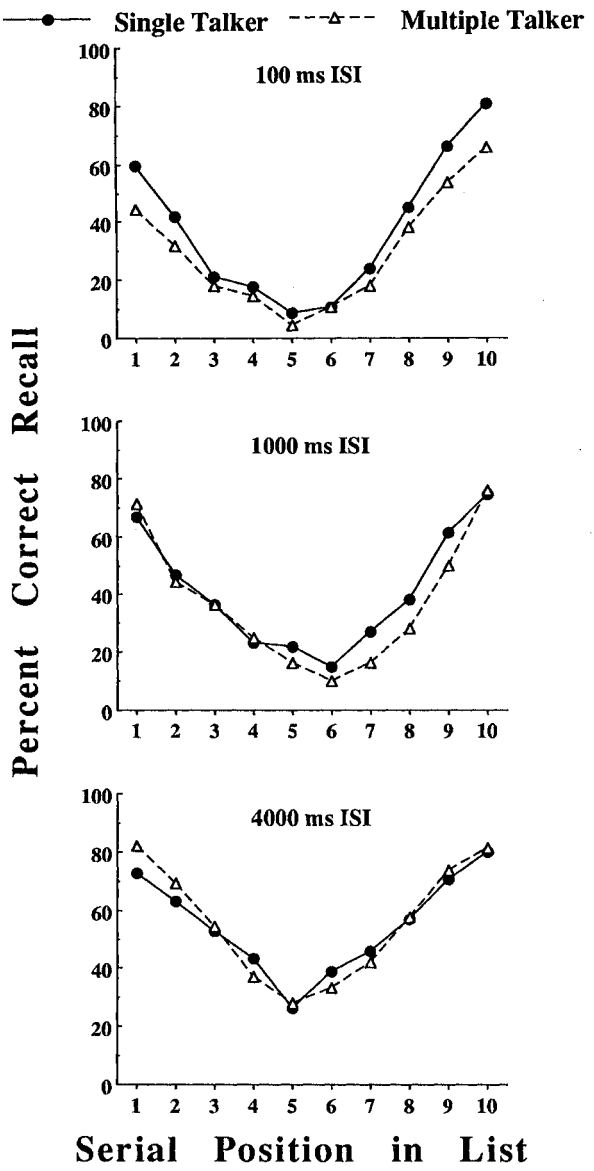

Figure 1 - Mean percentages of correctly recalled words for single- and multiple-talker lists as a function of serial position and presentation rate.

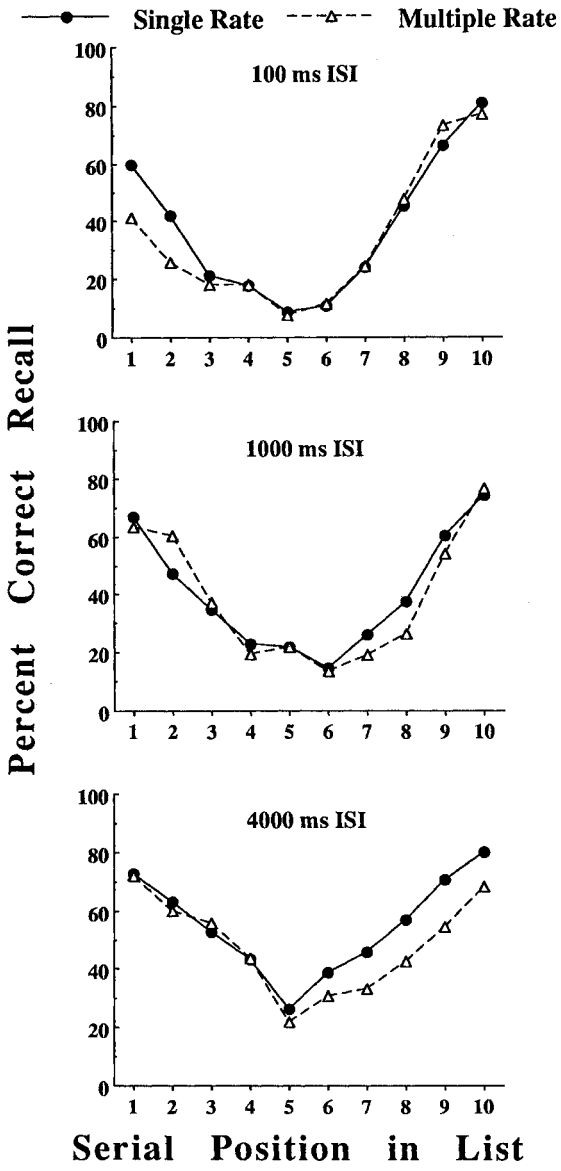

Figure 2 - Mean percentages of correctly recalled words for single- and multiple-speaking rate lists as a function of serial position and presentation rate.

IV. DISCUSSION

The present set of experiments was designed to assess the effects of talker variability and speaking rate on the perceptual encoding, rehearsal, and memory processes involved in recall of lists of spoken words. To that end, serial recall performance for multiple-talker and multiple-speaking rate lists was compared to serial recall of single-talker/single-speaking rate lists. In addition, presentation rate was manipulated to determine the time course of processing and encoding of both talker and rate information.

Turning first to the effects of talker variability, our findings replicate the results reported by Martin et al. [13] and Goldinger et al. [14]. Further, given that the previous findings used MRT words, the present results demonstrate reliable effects of talker variability with a new set of stimulus materials (PB words). At fast presentation rates, recall of items in early list positions is poorer for multiple-talker lists than single-talker lists. Assuming that recall performance in the primacy portion of the serial recall function reflects the amount of processing and the efficiency of rehearsal processes needed to transfer items into long-term memory [18], poorer recall of multiple-talker lists suggests that talker variability incurs a processing cost which somehow affects the successful transfer of items into long-term memory. In the 1000 ms ISI condition, we found no difference between multiple-talker and single-talker lists. This result suggests that given additional time subjects are able to process and encode multiple-talker lists at least as well as single-talker lists. Interestingly, at the slowest presentation rate, multiple-talker lists displayed an advantage in recall performance. In this condition, words in early list positions were recalled better in multiple-talker lists than in single-talker lists. As Goldinger et al. [14] have argued, this advantage in recall for multiple-talker lists at slow presentation rates suggests that talker information is retained in the representation of items and appears to be used by subjects to aid in subsequent recall. This change from talker variability impairing recall performance at fast list presentation

rates to aiding recall performance at slow presentation rates suggests that talker information may be integrated into subjects' representation of spoken words.

The results from the multiple- and single-speaking rate conditions suggest a somewhat different picture. As for the talker variability conditions, at the fast presentation rate, serial recall of words in initial list positions was better for single-speaking rate lists than for multiple-speaking rate lists. Again, this finding suggests that variation in speaking rate incurs some kind of processing cost which affects the successful encoding and rehearsal of early list items, especially at fast presentation rates. At 1000 ms ISI, just as for talker variability, no difference was found in recall performance between multiple-speaking rate and single-speaking rate lists. With speaking rate as well, given sufficient time, subjects are able to process and encode multiple-rate lists as well as single-rate lists. At the 4000 ms ISI, however, a difference does emerge between the talker variability conditions and the speaking rate variability conditions. For variations in speaking rate, no benefit was found for multiple-speaking rate versus single-speaking rate lists at the slow presentation rate. This finding contrasts with recall performance at the slowest presentation rate with talker variability. In this case, talker information appeared to aid in serial recall because subjects presumably had sufficient time to make use of the distinctive attributes provided by each different voice. Information about speaking rate, in contrast, does not appear to provide any additional information at the slow presentation rate that subjects can use to aid in their serial recall.

One reason that speaking rate may not be beneficial at slow presentation rates is simply that the two different sources of variability are processed and encoded differently. Changes in speaking rate and talker variability are realized acoustically quite differently. Further, they also have very different roles in terms of indexical properties and function in speech. For these reasons, speaking rate may not necessarily be encoded in long-term memory representations in the same manner as talker-specific attributes. Rather, speaking rate may be used only to assess the phonetic value of linguistic segments. This interpretation implies that rate information may be treated in a fundamentally different manner than talker information.

ACKNOWLEDGMENTS

This research was supported by NIH Training Grant #DC-00012-12 and #DC-00012-13 and NIDCD Research Grant DC-00111-16 to Indiana University in Bloomington.

REFERENCES

[1] A. M. Liberman, F. S. Cooper, D. P. Shankweiler, and M. Studdert-Kennedy. "Perception of the Speech Code," *Psychological Review*, vol. 74, pp. 431-461, 1967.

[2] D. H. Klatt. "Linguistic Uses of Segmental Duration in English: Acoustic and Perceptual Evidence," *Journal of the Acoustical Society of America*, vol. 59, pp. 1208-1221, 1976.

[3] K. N. Stevens and S. E. Blumstein. "Invariant Cues for Place of Articulation in Stop Consonants," *Journal of the Acoustical Society of America*, vol. 64, pp. 1358-1368, 1978.

[4] D. P. Shankweiler, W. Strange, and R. R. Verbrugge. "Speech and the Problem of Perceptual Constancy," in R. Shaw & J. Bransford (Eds.), *Perceiving, Acting, Knowing: Toward an Ecological Psychology*. New Jersey, Erlbaum, pp. 315-346, 1976.

[5] J. L. Miller and A. M. Liberman. "Some Effects of Later Occurring Information on the Perception of Stop Consonant and Semivowel," *Perception & Psychophysics*, vol. 25, pp. 457-465, 1979.

[6] Q. Summerfield. "On Articulatory Rate and Perceptual Constancy in Phonetic Perception," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 7, pp. 1074-1095, 1981.

[7] M. A. Joos. "Acoustic Phonetics," *Language*, vol. 24, pp. 1-136, 1948.

[8] P. Ladefoged and D. E. Broadbent. "Information Conveyed by Vowels," *Journal of the Acoustical Society of America*, vol. 29, pp. 98-104, 1957.

[9] Q. Summerfield and M. P. Haggard. "Vocal Tract Normalisation as Demonstrated by Reaction Times," *Report on Research in Progress in Speech Perception*, vol. 2, Northern Ireland: Queen's University, 1973.

[10] J. W. Mullennix, D. B. Pisoni, and C. S. Martin. "Some Effects of Talker Variability on Spoken Word Recognition," *Journal of the Acoustical Society of America*, vol. 85, pp. 365-378, 1988.

[11] W. R. Garner. *The Processing of Information and Structure*. Potomac, MD: Erlbaum, 1973.

[12] J. W. Mullennix and D. B. Pisoni. "Stimulus Variability and Processing Dependencies in Speech Perception," *Perception & Psychophysics*, vol. 47, pp. 379-390, 1990.

[13] C. S. Martin, J. W. Mullennix, D. B. Pisoni, and W. V. Summers. "Effects of Talker Variability on Recall of Spoken Word Lists," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 15, pp. 676-684, 1989.

[14] S. D. Goldinger, D. B. Pisoni, and J. S. Logan. "On the Nature of Talker Variability Effects on Recall of Spoken Word Lists," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 17, pp. 152-162, 1991.

[15] T. J. Palmeri, S. D. Goldinger, and D. B. Pisoni. "Episodic Encoding of Voice Attributes and Recognition Memory for Spoken Words," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Manuscript under review, 1992.

[16] M. S. Sommers, L. C. Nygaard, and D. B. Pisoni. "The Effects of Speaking Rate and Amplitude Variability on Perceptual Identification," *Journal of the Acoustical Society of America*, vol. 91, p. 2340, 1992.

[17] J. L. Miller and L. E. Volaitis. "Effect of Speaking Rate on the Perceptual Structure of a Phonetic Category," *Perception & Psychophysics*, vol. 46, pp. 505-512, 1989.

[18] A. D. Baddeley and G. J. Hitch. "Working Memory," in G. H. Bower (Ed.), *The Psychology of Learning and Memory, vol. 8*, New York, Academic, pp. 47-90, 1974.