Effects of stimulus variability on perception and representation of spoken words in memory

LYNNE C. NYGAARD

Indiana University, Bloomington, Indiana

MITCHELL S. SOMMERS

Washington University, St. Louis, Missouri

and

DAVID B. PISONI

Indiana University, Bloomington, Indiana

A series of experiments was conducted to investigate the effects of stimulus variability on the memory representations for spoken words. A serial recall task was used to study the effects of changes in speaking rate, talker variability, and overall amplitude on the initial encoding, rehearsal, and recall of lists of spoken words. Interstimulus interval (ISI) was manipulated to determine the time course and nature of processing. The results indicated that at short ISIs, variations in both talker and speaking rate imposed a processing cost that was reflected in poorer serial recall for the primacy portion of word lists. At longer ISIs, however, variation in talker characteristics resulted in improved recall in initial list positions, whereas variation in speaking rate had no effect on recall performance. Amplitude variability had no effect on serial recall across all ISIs. Taken together, these results suggest that encoding of stimulus dimensions such as talker characteristics, speaking rate, and overall amplitude may be the result of distinct perceptual operations. The effects of these sources of stimulus variability in speech are discussed with regard to perceptual saliency, processing demands, and memory representation for spoken words.

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 (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967), linguistic stress (Klatt, 1976), utterance length (Klatt, 1976; Oller, 1973), vocal tract size and shape (Fant, 1973; Joos, 1948; Peterson & Barney, 1952), and speaking rate (Miller, 1981, 1987a), for example, can all have profound effects on the acoustic realization of linguistic units. The consequences of these numerous sources of variability for speech perception is that phonetic segments do not necessarily have any invariant acoustic form (but see Kewley-Port, 1983; Stevens & Blumstein, 1978). Rather, the stable linguistic units that listeners perceive are based on an acoustic signal that is highly variable. The purpose of the present series of experiments was to examine in detail three of these sources of variability—changes in talker characteristics,

changes in speaking rate, and changes in overall amplitude. The overall goal of these studies was to assess the consequences of changes along each of these stimulus dimensions for the perceptual processing and subsequent representation of spoken words in memory.

Traditionally, accounts of speech perception have characterized variability in the acoustic speech signal as a perceptual problem that listeners must solve (Shankweiler, Strange, & Verbrugge, 1976). Perceivers are thought to achieve consistent phonetic percepts given variation in talker and speaking rate through a perceptual normalization process in which linguistic units are evaluated relative to the prevailing rate of speech (Miller, 1987a; Miller & Liberman, 1979; Summerfield, 1981) or relative to specific characteristics of a talker's vocal tract (Joos, 1948; Ladefoged & Broadbent, 1957; Summerfield & Haggard, 1973). Implicit in this traditional view of perceptual normalization is the assumption that the end product of speech perception is a sequence of idealized canonical linguistic units. Variability is assumed to be stripped away to arrive at the abstract symbolic representations that are used for further linguistic analysis (see, e.g., Halle, 1985; Joos, 1948; Kuhl, 1991, 1992). Despite the important role that information such as talker identity and speaking rate may play in the communicative setting (Ladefoged & Broadbent, 1957; Laver, 1989; Laver & Trudgill, 1979), little acknowledgment has been made of how that information might be used in real-time speech perception. Recently, however, re-

This research was supported by NIH Research Grant DC-00111-16 and NIH Training Grant DC-00012-13 to Indiana University. Portions of this research were presented at the 123rd meeting of the Acoustical Society of America in Salt Lake City, Utah. We thank Terrance M. Nearey and two anonymous reviewers for useful comments and suggestions. We also wish to thank Luis Hernandez for his help with computer programming and systems analysis. Correspondence should be addressed to L. C. Nygaard, Speech Research Laboratory, Department of Psychology, Indiana University, Bloomington, IN 47405 (e-mail: lnygaard@indiana.edu).

searchers have begun to examine the relationship between perceptual mechanisms that are dedicated to resolving differences in talker characteristics and speaking rate and those mediating the apprehension of the linguistic content of the speech signal (e.g., Miller & Volaitis, 1989; Nygaard, Sommers, & Pisoni, 1994; Sommers, Nygaard, & Pisoni, 1992, 1994; Volaitis & Miller, 1992).

Talker Variability

Consider first the perceptual consequences of variation in talker characteristics. Several experiments have demonstrated that talker variability can affect both vowel perception (Assmann, Nearey, & Hogan, 1982; Summerfield, 1975; Summerfield & Haggard, 1973; Verbrugge, Strange, Shankweiler, & Edman, 1976; Weenink, 1986) and word recognition (Cole, Coltheart, & Allard, 1974; Creelman, 1957; Mullennix, Pisoni, & Martin, 1989). For example, Mullennix, Pisoni, and Martin (1989) found that word recognition performance was poorer when listeners were presented with words produced by multiple talkers than when the same items were produced by only a single talker. Using a speeded classification task (Garner, 1974), Mullennix and Pisoni (1990) also observed that subjects had difficulty in ignoring irrelevant variation in a talker's voice when asked to classify words by initial phoneme. Likewise, subjects took longer to classify talkers' voices when there was irrelevant variation in phonetic context. Taken together, these findings suggest that variability due to changes in talker characteristics is both time and resource demanding. As the talker's voice changes from trial to trial in these tasks, listeners appear to devote additional processing resources to recover the phonetic content of the utterance. These additional processing requirements are reflected in longer response latencies and poorer identification performance for words produced in multiple-talker contexts than for words produced in single-talker contexts.

Additional research has demonstrated that talker variability can affect memory processes as well as perceptual encoding. Martin, Mullennix, Pisoni, and Summers (1989) found that serial recall of spoken word lists produced by multiple talkers was poorer than recall of lists produced by a single talker, but only in the primacy portion of the serial recall curve (Positions 1–3 in 10-word lists). According to Martin et al., these findings suggest that recall of lists of words spoken by multiple talkers required greater processing resources for the encoding and subsequent rehearsal of words in working memory than did the recall of single-talker lists. When listeners memorized lists of words produced by multiple talkers, the variations introduced by changing talker characteristics required the use of limited processing resources that were also needed for encoding, rehearsal, and transfer of list items into long-term memory. In addition, Martin et al. showed that recall of visually presented digits presented prior to the serial recall task was poorer when the subsequent to-be-remembered lists were produced by multiple talkers rather than by a single talker. These results provide additional evidence that the encoding and rehearsal of words from multiple-talker lists required more processing resources than did the single-talker

lists, leaving fewer resources available for storage and retrieval of the visually presented digits. In addition, it appears that these effects are not modality specific, and that they reflect use of general cognitive processing resources.

In a series of follow-up experiments, Goldinger, Pisoni, and Logan (1991) introduced an additional experimental manipulation into the serial recall task to determine the exact nature of the talker variability effects. Presentation rate, a variable assumed to affect primarily rehearsal processes (Jahnke, 1968; Murdock, 1962; Rundis, 1971), was used to assess the relative effects of talker variability on perceptual encoding and rehearsal processes. If talker variability affects only perceptual encoding, the differences between recall of multiple-talker lists and single-talker lists should remain relatively constant across different presentation rates. However, if talker variability affects rehearsal processes as well, this variable would be expected to interact with presentation rate. Goldinger et al. found that at relatively fast presentation rates (like those used by Martin et al., 1989), serial recall of spoken words was better in initial list positions for single-talker lists than for multiple-talker lists. At slower presentation rates, however, Goldinger et al. found that recall of words in initial list positions from multiple-talker lists was *superior* to recall of words from single-talker lists. The pattern of results suggested that at fast presentation rates, variation due to changes in the talker affects *both* 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 the concomitant talker information along with each word and are able to use the additional redundant talker information as an aid in retrieval.

Additional evidence that detailed talker-specific information is encoded, and used, in long-term memory comes from a recent series of experiments conducted by Palmeri, Goldinger, and Pisoni (1993; see also Nygaard et al., 1994). Using a continuous recognition memory procedure, Palmeri et al. showed that instance-specific voice information was retained along with lexical information about the word, and these attributes aided the subjects later in recognition memory. The finding that subjects were able to use talker-specific information suggests that this source of variability may not be discarded or normalized out in the process of speech perception, as has been widely assumed in the literature. Rather, variations in a talker's voice may become part of a very rich and highly detailed representation of the speaker's utterance in long-term memory (see also Craik & Kirsner, 1974; Geiselman, 1979; Geiselman & Bellezza, 1976, 1977; Geiselman & Crawley, 1983). The effects of talker variability on performance would then be due to the additional attention and resources necessary to encode information conveyed by a talker's voice.

Given this view of the processing of talker information in speech perception, we conducted a series of experiments to investigate the effects of additional sources of variability on the processing and representation of speech. First, we sought to replicate the earlier findings of the effects of talker variability on perceptual and memory processes involved in serial recall. Second, we sought to extend the in-

vestigation to two additional sources of variability—variability due to speaking rate and variability due to changes in overall amplitude.

Variation in Speaking Rate

An extensive body of research suggests that speech perception is affected by changes in a talker's speaking rate. For example, Miller and Liberman (1979) presented listeners with a synthetic /ba/-/wa/ continuum in which the stop-glide distinction was cued primarily by the duration of the formant transitions. When they varied the duration of the steady-state portion of the syllable, they observed a shift in identification boundaries toward longer values of transition duration as the overall duration of the syllable became longer. Miller and Liberman interpreted these findings as support for the proposal that speech is perceived in a rate-dependent manner; as the steady-state portion of the syllable became longer, specifying a slower speaking rate, listeners adjusted their perceptual judgments accordingly (Miller, 1981, 1987a).

Additional research has shown that listeners use information about the speaking rate of a precursor phrase, as well as visual articulatory rate information, in making perceptual judgments about phonetic contrasts. Summerfield (1981) conducted a series of experiments to evaluate the effect of a precursor phrase varying in rate of articulation on the identification of voiced versus voiceless syllable-initial stop consonants. His results showed that phoneme identification boundaries shifted to shorter values of voice onset time (VOT) as the articulation rate of the precursor phrase increased. Thus, listeners were apparently basing their classification of the initial stop consonants on information about the prevailing rate of speech in the precursor phrase. Green and Miller (1985) have found that segmental judgments that depend on temporal information are susceptible to rate information provided by visually presented articulations as well as by auditory information. Thus, it appears that visual and auditory articulatory rate information both contribute to the categorization of phonetic segments.

More recently, Miller and her colleagues (Miller & Volaitis, 1989; Volaitis & Miller, 1992) have shown that both phonetic category boundaries and accompanying phonetic category structures that rely on temporal information are quite sensitive to relative changes in articulation rate. Their results indicate that changes in speaking rate affect the mapping of acoustic (and perhaps visual) information onto phonetic category structures. Thus, listeners seem to compensate for changes in speaking rate not only by shifting category boundaries, but also by restructuring their phonetic categories.

Although it appears clear that listeners are sensitive to changes in rate of articulation, less attention has been paid to the perceptual consequences, in terms of attention, processing resources, and memory, of this source of variability in the speech signal. At issue here is whether the observed changes in perceptual judgments are due to compensatory processes that require time and attention or whether adjustments to differences in speaking rate are

more automatic and less costly in terms of analysis than adjustments to differences in talker characteristics. Recently, Sommers, Nygaard, and Pisoni (1994) investigated this issue by presenting listeners with lists of words mixed in noise under two conditions—one in which all the words in the list were presented at a single speaking rate and one in which words in the list were presented at different speaking rates. The results were quite similar to those found earlier for talker variability (Mullennix et al., 1989). Although items in the single- and mixed-rate conditions were identical, listeners were better able to identify words mixed in noise if all the words were produced at a single speaking rate. Changes in speaking rate from word to word in a list apparently incurred a processing cost that made identifying words more difficult. These results suggest that if a compensatory process does exist, it must demand the attention and processing resources of the listener, leaving fewer resources available for the identification of words.

Given the effects of overall speaking rate on the perceptual processing of temporally based phonetic contrasts, the question arises whether rate variability would have effects similar to those found for talker variability in terms of perceptual encoding and memory representation. That is, given that changes in speaking rate appear to incur processing costs during perception, is speaking rate also encoded into long-term memory in the same manner as is talker variability? By comparing performance across different interstimulus intervals (ISIs) (e.g., Goldinger et al., 1991) in a serial recall task with high- and low-variability word lists, we hoped to assess how speaking rate would be encoded and represented during speech perception and spoken word recognition.

Amplitude Variability

In contrast to the demonstrated effects of variability in speaking rate and talker characteristics, whether or not variability in the overall amplitude of an utterance affects phonetic processing has received little attention. Although variations in relative amplitude across linguistic units in an utterance can contribute to the perception of linguistic stress, vowel height, and intonation, overall amplitude or the amplitude at which a stimulus is presented has not been found to affect phonetic or linguistic judgments of speech (see Klatt, 1985). As such, a compensatory or a normalization process has not been proposed for variation in presentation level. For example, in the study conducted by Sommers, Nygaard, and Pisoni (1994) discussed earlier, we compared word identification under several conditions of signal distortion for lists with items presented at multiple amplitudes versus lists with the same items presented at a single overall amplitude. No difference was observed in identification performance for lists with amplitude levels varying over a 30-dB range in comparison with lists with a constant amplitude level. These results suggest that variability in amplitude may not incur a significant processing cost in the analysis of spoken words.

In our second experiment, changes in overall amplitude were studied to assess the effects of variability along a dimension of the speech signal that is assumed to be irrele-

vant for segmental analysis. Is it the case that any source of variability in the acoustic speech signal has consequences for the perceptual encoding and rehearsal of spoken words, or is it the case that only sources of variation that are phonetically relevant in a given task affect perception and memory of spoken words? That is, would variation along a phonetically irrelevant dimension such as overall amplitude or presentation level affect memory for serial order in the same manner as talker variability? If so, this would suggest that variation in all aspects of spoken words is attention- and resource-demanding regardless of their effect on the perception of phonetic distinctions. If variation in amplitude does not affect perception and memory of spoken words, this would suggest either that only aspects of spoken words that are linguistically relevant affect processing, encoding, and rehearsal of spoken words or that the perceptual system adjusts for changes in overall amplitude level relatively quickly and easily in the analysis of speech, and consequently, that these changes have little effect on subsequent stages of processing.

EXPERIMENT 1

The first experiment was conducted for two reasons. First, we wanted to replicate and confirm Goldinger et al.'s (1991) talker variability results. Second, we wanted to investigate the effects of changes in speaking rate on serial recall. Although our talker variability replication was similar to Goldinger et al.'s previous experiments, two differences should be noted. First, our stimulus set consisted of a set of monosyllabic words with a wider phonetic inventory and more varied syllable structure than the materials used by Goldinger et al. Second, our presentation rates were slightly different from those used by Goldinger et al. (1991). Although both studies varied presentation rate by changing the ISI of words in the lists, our ISIs were 100, 1,000, and 4,000 msec, whereas Goldinger et al. used 250-, 500-, 1,000-, 2,000-, and 4,000-msec ISIs.¹ Despite these minor differences, we predicted that at the fastest presentation rate (100-msec ISI), listeners would recall more words in initial list positions from single-talker lists than from multiple-talker lists. Serial recall performance in initial list positions is assumed to reflect the efficiency of rehearsal and storage of items in long-term memory (Glanzer & Cunitz, 1966; Murdock, 1962). Thus, effects of variability on perception, rehearsal, and storage of items in long-term memory should be seen in serial recall performance in initial list positions. At the medium presentation rate (1,000-msec ISI), we predicted that there would be no difference between multiple- and single-talker lists. Finally, at the slowest presentation rate (4,000-msec ISI), we predicted that there would be superior recall for items in initial list positions from the multiple-talker lists as opposed to such recall from the single-talker lists.

For the speaking rate manipulation, our method was identical to that used in the talker variability experiment, except that the effects of variations in articulation rate rather than talker characteristics were measured. Consequently,

serial recall of lists of words produced at a single speaking rate was compared to serial recall of lists produced at multiple speaking rates. Our prediction was that because speaking rate has been shown to affect phonetic judgments and thus requires processing resources and attention, variation in speaking rate should also produce a pattern of results similar to that found for talker variability in the serial recall task. Again, three ISIs (100, 1,000, 4,000 msec) were used to determine the role of processing time on the perceptual encoding and representation of rate information. At short ISIs, we expected that recall for words in initial list positions would be poorer for multiple-rate lists than for single-rate lists. In the 1,000-msec ISI condition, recall performance was expected to be comparable for multiple and single speaking rate lists. Finally, at the longest ISIs, we expected that as variations in speaking rate were fully processed and encoded, recall performance for multiple speaking rate lists in initial list positions would be superior to single speaking rate lists.

Method

Subjects

One hundred eighty undergraduate students enrolled in introductory psychology courses at Indiana University 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 any speech or hearing disorder at the time of testing.

Stimuli

The stimuli consisted of a set of 100 monosyllabic words drawn from phonetically balanced (PB) word lists (ANSI, 1971). To obtain the database of words produced by different talkers at different speaking rates, words were embedded in the carrier phrase "Please say the word _____" for presentation to speakers. Ten speakers (6 female and 4 male) were asked to pronounce each sentence at three different speaking rates—slow, medium, and fast—for a total of 3,000 words (100 words $\times$ 10 speakers $\times$ 3 rates). Order of productions for the different speaking rates was determined randomly. The words were digitized on line at a 10-kHz sampling rate, low-pass filtered at 4.8 kHz, and subsequently edited from the carrier phrase for presentation. The average root mean square amplitude of all stimulus tokens was equated by using a signal processing software package (Luce & Carrell, 1981).

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 utterances, 5 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 the 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 3,000 words, eight 10-word lists were constructed. 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 successive word in a list was produced at a different speaking rate. In the single-talker and single-rate condition, all words were produced by one of two talkers (one male or one female) at a normal or medium speaking rate.

Procedure

Subjects were tested in groups of 5 or fewer in a quiet testing room. Stimuli were presented over matched and calibrated TDH-39 headphones at approximately 80 dB (SPL). A PDP-11/34 computer was used to present stimuli and to control the experimental procedure in real time. The digitized stimuli were presented using a 12-bit digital-to-analog converter and were low-pass filtered at 4.8 kHz.

During the experiment, subjects first heard a 500-msec, 1,000-Hz warning tone to alert them that a list was about to be presented. After a 500-msec silent interval, a list of 10 words was presented, with one of three ISIs—100, 1,000, or 4,000 msec—between words. After each list was presented, another warning tone sounded, indicating the beginning of the recall period. Subjects had 60 sec 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 and to write down their responses on answer sheets provided by the experimenter.

Talker, speaking rate, and ISI were all between-subjects variables. Of the 180 subjects, 60 were tested at each ISI. Of those 60, 20 subjects were tested in the multiple-talker condition, 20 were tested in the multiple speaking rate condition, and 20 were tested 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 interword intervals. In addition, for the single-talker/single speaking rate condition, half of the subjects (10) received lists produced by a male speaker and half received lists produced by a female speaker to ensure that differences between conditions were not due to the idiosyncratic characteristics of a particular talker. Because this factor was never found to influence recall performance, it will not be included in subsequent discussions or analyses.

Results and Discussion

Subjects' responses were scored as correct only if the target word or a phonetically equivalent spelling of the target word was recalled in the same serial position as that of the word presented in the list. Figure 1 shows the effects of talker variability on serial recall at the three ISIs. Percentage 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-msec ISI, the middle panel shows recall performance from the 1,000-msec ISI condition, and the bottom panel shows the results from the 4,000-msec ISI condition. A three-way analysis of variance (ANOVA) with talker condition (multiple vs. single), serial position (1–10), and ISI (100, 1,000, 4,000 msec) as factors 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 [$F(9,1026) = 153.20, p < .001$]. In addition, a significant main effect of ISI was found [$F(2,114) = 39.21, p < .001$]. Total recall performance increased overall as ISIs increased. The interaction of serial position and ISI was also significant [$F(18,1026) = 3.72, p < .001$], indicating that the shape of the serial position curve changed as recall performance improved at the longer ISIs. No other main effects or interactions were found to be significant in this analysis.

In order to assess the effects of talker variability on serial recall performance more directly, separate ANOVAs were conducted just for the early (List Positions 1–3), middle (List Positions 4–7), and late (List Positions 8–10)

Figure 1. Mean percentage of correctly recalled words for both the single- and multiple-talker lists as a function of serial position and interstimulus interval (ISI).

portions of the serial position curve. In the analysis of the early portion of the curve, with list position (1–3), ISI, and talker as factors, significant main effects of position [$F(2,228) = 118.50, p < .001$] and ISI [$F(2,114) = 42.52, p < .001$] were found, reflecting differences in recall performance as a function of position in the lists and improved recall performance as ISIs became longer. In addition, a significant interaction was found between serial position and ISI [$F(4,228) = 2.45, p < .05$], indicating that the serial position curve over the first three list positions changed as recall performance improved with longer ISIs. Finally, a significant interaction was found between ISI and talker variability [$F(2,114) = 2.92, p < .05$]. To assess the locus of this interaction, separate additional two-way ANOVAs for the shortest ISI condition and for the longest ISI condition were conducted with responses only to items in early list positions (1–2). List position and talker variability conditions were factors. At the shortest ISI, a significant main effect of talker variability was found, indicating that recall of words from single-talker lists was superior to recall of

words from multiple-talker lists [$F(1,38) = 6.72, p < .02$]. At the longest ISI as well, a significant main effect of talker variability was found [$F(1,38) = 3.98, p < .05$]. However, at the longest ISI, recall of words from multiple-talker lists was superior to recall of words from single-talker lists.

In the analysis of the middle portion of the serial position curve, with list position (4–7), ISI, and talker as factors, significant main effects of position [$F(3,342) = 17.53, p < .001$] and ISI [$F(2,114) = 27.98, p < .001$] were found, reflecting differences in recall performance as a function of position in the lists and improved recall performance as ISIs became longer. In addition, a significant interaction was found between serial position and ISI [$F(6,342) = 2.84, p < .01$], indicating that the middle portion of the curve changed in shape as recall performance improved with longer ISIs. No other main effects or interactions were significant.

In the analysis of the late portion of the serial position curve, with list position (8–10), ISI, and talker as factors, significant main effects of serial position [$F(2,228) = 130.56, p < .001$] and ISI [$F(2,114) = 6.11, p < .001$] were found. Recall performance varied as a function of position in the lists, and more items were recalled overall with longer ISIs. In addition, a significant interaction was found between serial position and ISI [$F(4,228) = 3.58, p < .05$], indicating that the late portion of the serial position curve changed as recall performance improved with longer ISIs. No other main effects or interactions were significant. It should be noted that the effect of talker variability does not result in a significant main effect for items in the recency portion of the serial recall curve even though there are apparent differences in Figure 1. Because these recency effects are not reliable for our data set, and because our predictions concern the recall of items in early list positions, we will not consider them further in this paper.

Figure 2 shows the effects of variations in speaking rate on serial recall performance at the three ISIs. 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-msec ISI. The middle panel shows recall performance from the 1,000-msec ISI condition, and the bottom panel shows the results from the 4,000-msec ISI condition. A three-way ANOVA with speaking rate (multiple vs. single), serial position, and ISI as factors revealed significant main effects for ISI [$F(2,114) = 21.00, p < .001$], serial position [$F(9,1026) = 147.86, p < .001$], and speaking rate [$F(1,114) = 3.99, p < .05$]. Recall performance was better overall at long ISIs than at short ISIs; recall varied as a function of list position; and finally, recall performance was affected by variability in speaking rate. In addition, a significant three-way interaction [$F(18,1026) = 2.17, p < .005$] was found, indicating that differences between multiple- and single-rate lists changed as a function of list position and ISI.

To further assess the effects of speaking rate on recall performance, separate ANOVAs were conducted for early (1–3), middle (4–7), and late (8–10) list positions. In the

Figure 2. Mean percentage of correctly recalled words for both the single and multiple speaking rate lists as a function of serial position and interstimulus interval (ISI).

analysis of the early portion of the serial position curve, with list position (1–3), ISI, and speaking rate as factors, significant main effects of position [$F(2,228) = 95.05, p < .001$] and ISI [$F(2,114) = 32.93, p < .001$] were found. Recall performance changed as a function of list position, and more items were recalled with longer ISIs. In addition, a significant interaction was found between serial position and ISI [$F(4,228) = 2.64, p < .04$], indicating that the serial position curve over the first three list positions changed as recall performance improved with longer ISIs. A significant two-way interaction was found between ISI and speaking rate [$F(2,114) = 3.08, p < .05$] and is considered below. Finally, a significant three-way interaction among speaking rate, serial position, and ISI was also found [$F(4,228) = 2.40, p < .05$], indicating that recall of items from early list positions changed both as a function of ISI and speaking rate.

To assess the locus of the two-way interaction between ISI and speaking rate, separate additional two-way ANOVAs for the shortest ISI condition and for the longest ISI condition were conducted with responses only to items in early list positions (1–2). List position and rate variability were factors. At the shortest ISI, a significant main effect of rate variability was found, indicating that recall of words from single speaking rate lists was superior to recall of words from multiple speaking rate lists [$F(1,38) = 10.71$, $p < .005$]. At the longest ISI, however, no comparable significant main effect of rate variability was found, indicating that recall of words from multiple-rate lists was similar to recall of words from single-rate lists.

In the analysis of the middle portion of the serial position curve, with list position (4–7), ISI, and speaking rate as factors, significant main effects of position [$F(3,342) = 16.76$, $p < .001$] and ISI [$F(2,114) = 23.30$, $p < .001$] were found. Recall performance changed as a function of list position, and more items were recalled overall with longer ISIs. In addition, a significant interaction was found between position and ISI [$F(6,342) = 5.12$, $p < .001$], indicating that the middle portion of the serial position curve changed in shape as recall performance improved with longer ISIs. No other main effects or interactions were significant.

In the analysis of the late portion of the serial position curve, with list position (8–10), ISI, and speaking rate as factors, significant main effects of position [$F(2,228) = 166.63$, $p < .001$] and ISI [$F(2,114) = 3.06$, $p < .05$] were found, reflecting differences in recall performance as a function of position in the lists and improved recall performance as ISIs became longer. In addition, a significant interaction was found between serial position and ISI [$F(4,228) = 5.07$, $p < .001$], indicating that the late portion of the serial position curve changed as recall performance improved with longer ISIs. No other main effects or interactions were significant. Again, it should be noted that the effect of variability does not result in a significant main effect for items in the recency portion of the serial recall curve even though there are apparent differences in Figure 2. Because these recency effects for rate variability are not reliable for our data set, and because our predictions concern the recall of items in early list positions, we will not consider them further in this paper.

It should be noted that the critical difference between the talker variability and the rate variability conditions can be seen in the first few serial positions in the longest ISI condition (i.e., 4 sec). The separate two-way ANOVAs reported earlier for the talker and rate conditions conducted with responses only to items in early list positions (1–2) and only at the longer ISI revealed a significant main effect of talker [$F(1,38) = 3.98$, $p < .05$], but no significant main effect of rate. In contrast, main effects of variability were found for both talker and rate condition at the shortest ISI. That is, the analysis of the talker data at the longest ISI revealed a difference between multiple-talker and single-talker conditions, but the analysis of the rate condition at the longest ISI revealed no difference between multiple-rate and single-rate conditions for initial list positions. Recall

performance in the initial list positions was better for multiple-talker lists than for single-talker lists at the long ISI. However, a comparable benefit of variation in speaking rate was not found in the primacy portion of the curve. Thus, the two sources of variability did not have comparable effects on recall performance at the longest ISI.

First, in regard to the effects of talker variability, our findings replicate the major results reported by Martin et al. (1989) and Goldinger et al. (1991). Further, given that the previous studies used monosyllabic words selected from the Modified Rhyme Test (MRT), the present results demonstrate reliable effects of talker variability with a new set of stimulus materials (PB words). At short ISIs, recall of items in early list positions is poorer for multiple-talker lists than for single-talker lists. Assuming that recall performance in the primacy portion of the serial recall function reflects the amount of processing and rehearsal needed to transfer items into long-term memory (Baddeley & Hitch, 1974), poorer recall of multiple-talker lists suggests that listeners incur a processing cost due to increased talker variability that affects the efficient transfer of items into long-term memory. In the 1,000-msec 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 longest ISI, multiple-talker lists actually displayed an advantage in recall performance, replicating a result reported earlier by Goldinger et al. (1991). In this condition, words in early list positions were recalled better in multiple-talker lists than in single-talker lists. As Goldinger et al. (1991) have argued, this advantage in recall for multiple-talker lists at long ISIs suggests that talker information is retained in long-term memory and appears to be used by subjects to aid in subsequent recall.

In contrast to the effects observed for talker variability, the results from the multiple and single speaking rate conditions suggest a somewhat different picture. As observed in the talker variability conditions, at the short ISI, recall of words in initial list positions was better for single-rate lists than for multiple-rate lists. Again, this finding suggests that variation in speaking rate incurs a processing cost that affects the successful encoding and rehearsal of early list items, especially at short ISIs. At the 1,000-msec ISI, just as for talker variability, no difference was found in recall performance between multiple speaking rate and single speaking rate lists. With variability in speaking rate, given sufficient time, subjects are able to process and encode multiple-rate lists as well as single-rate lists. At the 4,000-msec ISI, however, a difference does emerge between the effects of talker variability and the effects of variation in speaking rate. For variations in speaking rate, no benefit was found for multiple speaking rate versus single speaking rate lists at the longest ISI. This finding contrasts with recall performance at the longest ISI with talker variability. In the latter case, talker information appeared to aid in recall because subjects presumably had sufficient time to make use of the distinctive attributes provided by each different voice. Information about speaking rate in the long-

est ISI condition, in contrast, does not appear to be used by subjects to aid in their serial recall. In short, speaking rate does not appear to be encoded into a long-term memory representation of items used by listeners in this task.

Taken together, these findings suggest that the perceptual analysis of both talker and rate information requires time and resources. However, even though both sources of stimulus variability reduce performance at short ISIs, presumably requiring more time for processing and rehearsal, only talker information appears to be retained in memory for later use as an additional redundant cue to encoding temporal order in the serial recall task. The question remains whether the processing time and computational resources needed to accommodate variability in speaking rate and talker characteristics have a common origin or whether they are due to separate processing operations. We consider this issue again in the General Discussion.

EXPERIMENT 2

The second experiment was designed to study the effects of variation in the overall amplitude level of spoken words. As mentioned previously, variability in the overall level of presentation has not been shown to affect phonetic processing per se, and, as such, is not assumed to require additional time and processing resources for perceptual analysis and encoding. By comparing the effects of amplitude variability with the effects of talker and rate variability, we sought to evaluate the role of phonetic relevance on the encoding, rehearsal, and transfer of spoken words into memory. If variations in overall amplitude do produce effects on recall performance in the serial recall task, this finding would suggest that the processing of any stimulus variation in the speech signal draws upon a common pool of processing resources that are also used for the evaluation of the phonetic content of an utterance. If variability in overall amplitude does not affect recall performance, it can be argued that only factors that affect the perceptual analysis of spoken words will compete for time and resources in the serial recall task. In this case, amplitude differences either could be analyzed and compensated for by peripheral auditory mechanisms or simply might not be attended to in the serial recall task.

Method

Subjects

One hundred twenty undergraduate students enrolled in introductory psychology courses at Indiana University 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 at the time of testing.

Stimuli

The stimuli consisted of the 100 medium speaking rate words used in Experiment 1. To create different presentation levels, words were digitally manipulated to create three different overall amplitudes via a signal processing software package. The maximum level for a waveform was set to a specified value, and the remaining amplitude values in the digital file were then rescaled relative to that maximum. The three root mean square overall amplitudes—40, 50, and 60 dB—used in the present investigation were chosen to be both perceptually

salient and free of distortion. Thus, from the original 100 medium rate words used in Experiment 1, 300 new stimuli were created—100 words at each of the three amplitude levels. From this set of 300 words, eight 10-word lists were constructed. In the multiple-amplitude condition, successive words in each list were chosen from a different overall amplitude level. In the single-amplitude condition, each word in the list had been created at the medium amplitude of 50 dB. The presentation level for the experiment was calibrated so that the medium amplitude stimuli were presented at approximately 80 dB (SPL), and the high- and low-amplitude stimuli were either 10 dB higher or 10 dB lower, depending on their relative amplitude differences.

Procedure

The design and procedures were the same as those in Experiment 1. Subjects were given a warning tone to signal the presentation of a list. Following this, a list of 10 words at one of three ISIs (100, 1,000, or 4,000 msec) was presented and subjects were asked to recall and write down the words from the list in the order in which they were presented.

Amplitude (single vs. multiple) and ISI (100, 1,000, 4,000 msec) were the between-subjects variables. Of the 120 subjects, 40 were tested at each ISI. Of those 40, 20 subjects were tested in the multiple-amplitude condition and 20 subjects were tested in the single-amplitude condition. The same words were heard by all subjects, but the overall presentation level of items varied, depending on condition. In addition, for the single-amplitude condition, half of the subjects received lists produced by one female speaker and half of the subjects received lists produced by another female speaker. Talker was varied in this manner, as it was in Experiment 1, to control for idiosyncratic characteristics of individual talkers. Again, no significant main effects or interactions with voice were observed; therefore, this factor will not be included in any subsequent discussions and analyses.

Results and Discussion

As in the first experiment, subjects' responses were scored as correct only if the target word (or a phonetically equivalent spelling) was recalled in the same serial position as that of the word presented in the list. Figure 3 shows the effects of variability in overall amplitude on serial recall at the three ISIs. The top panel shows serial recall performance for multiple-amplitude versus single-amplitude lists at the 100-msec ISI. The middle panel shows recall performance from the 1,000-msec ISI, and the bottom panel shows the results from the 4,000-msec ISI. A three-way ANOVA with amplitude condition (multiple vs. single), serial position (1–10), and ISI (100, 1,000, 4,000 msec) as factors was conducted on the number of correct responses. As expected, significant main effects of serial position [$F(9,1026) = 143.21, p < .001$] and ISI [$F(2,114) = 52.81, p < .001$] were found, showing changes in recall performance as a function of list position and better recall performance overall as ISI increased. Finally, no significant main effect of amplitude condition was found and none of the interactions were significant. Likewise, additional separate three-way ANOVAs conducted just for items in early (1–3), middle (4–7), or late (8–10) list positions uncovered no significant main effects or interactions involving amplitude. As with the effects of talker and rate, the effect of amplitude did not result in a significant main effect in any analysis, even though for later list positions, the single-amplitude word list appeared to be recalled better

Figure 3. Mean percentage of correctly recalled words for both the single- and multiple-amplitude lists as a function of serial position and interstimulus interval (ISI).

than the multiple-amplitude word lists in Figure 3. Again, because these effects are not reliable for our data set, they will not be considered further.

These results suggest that variation in overall presentation level has little effect on the perceptual encoding and rehearsal processes involved in the serial recall task. Although null results such as these must be interpreted with caution, it appears that amplitude variability does not demand the processing time and resources that would have resulted in impaired recall of words from initial list positions. These results replicate and extend recent findings on the effects of amplitude variability on perceptual identification found by Sommers et al. (1994) and on the effects of overall amplitude variations on auditory repetition priming reported by Church and Schacter (1994). In the Sommers et al. study, we found that variability in overall level did not impair subjects' ability to identify digitally distorted spoken words. Similarly, Church and Schacter observed no effect of overall amplitude on repetition prim-

ing. Thus, the present results provide additional evidence that the introduction of amplitude variability into a set of spoken words requires significantly less processing time and resources than does variability from either talker or speaking rate.

The absence of an effect of amplitude variability on serial recall performance has two implications for understanding the nature of perceptual encoding and memory representation for spoken words. The first is that not all sources of variability present in the speech signal affect perceptual processing equally. Variability introduced by changes in overall amplitude level does not require the same type of perceptual accommodation as that involved in the resolution of variation due to talker and speaking rate.

The second implication of these results is that even though the changes in amplitude from trial to trial in this serial recall task were perceptually salient to listeners, the subjects did not use this source of information to aid recall. That is, amplitude variability, like rate variability, does not appear to be encoded into long-term memory as talker information was in our first experiment. Although the reason for the differences in recall performance at the long ISIs between the talker variability condition and the rate- and amplitude-variability conditions is unclear, it does seem that variations in both speaking rate and amplitude may be the result of separate, but similar, mandatory processes occurring at an early peripheral level (Fodor, 1983; Miller, 1987b). Variability due to changes in speaking rate and overall amplitude may be processed and discarded at an early stage of auditory analysis, and hence these sources of information may not be available to benefit listeners' recall performance at long ISIs.

GENERAL DISCUSSION

The present set of experiments was designed to assess the effects of talker, speaking rate, and amplitude variability on the perceptual encoding, rehearsal, and memory processes employed in the recall of lists of spoken words. To accomplish this, serial recall performance for lists with high stimulus variability (i.e., multiple talker/rate/amplitude conditions) was compared with serial recall of lists with low stimulus variability (i.e., single talker/rate/amplitude conditions). In addition, ISI was manipulated to assess the time course of processing and encoding of these different sources of information.

Taken together, the results of our experiments showed that each of these three types of stimulus variability in speech appears to be analyzed and encoded in a distinct manner. The pattern of results across ISIs suggests that differences may exist in the way listeners analyze, rehearse, and encode information about a talker's voice, speaking rate, and the overall presentation level of an utterance. At relatively short ISIs, variations in a talker's voice from trial to trial impaired the perceptual encoding and/or transfer of spoken words into long-term memory. Words from multiple-talker lists were not recalled as well as were words from single-talker lists, replicating earlier results reported by Martin et al. (1989) and Goldinger et al. (1991). The

same results were found for variability introduced by changes in speaking rate in the short ISI condition. Words from multiple-speaking rate lists were not recalled as well as were words from single-speaking rate lists, suggesting that variations in speaking rate also require processing, which competes for a limited pool of resources. When there is little time for rehearsal and elaboration at short ISIs, variability in talker characteristics and speaking rate interfered with the successful perceptual analysis and transfer of words into memory. These results may be contrasted with amplitude variability, which did not impair serial recall performance in the short ISI condition.

One explanation for the differential effects of talker and rate versus amplitude variability is that differences in the first two dimensions can have profound ramifications for the resolution of the spectral and temporal properties of segmental phonetic contrasts in speech (see, e.g., Lade-foged & Broadbent, 1957; Miller & Liberman, 1979; Miller & Volaitis, 1989; Mullenix & Pisoni, 1990; Volaitis & Miller, 1992), whereas variation in amplitude does not. If this explanation is correct, listeners may be sensitive only to the variations in the speech signal that are phonetically relevant. That is, listeners may devote a significant amount of processing resources only to perceiving and encoding changes in the speech signal that affect the nature and structure of their phonetic categorization and subsequent word recognition (Miller & Volaitis, 1989; Volaitis & Miller, 1992). If that is the case, then serial recall of lists with resource- and attention-demanding variability (e.g., talker characteristics and speaking rate) should produce poorer recall performance than should lists with variation that is not relevant to phonetic analysis.

An alternative account of the pattern of results is that the resolution of variations in overall amplitude level may occur very early in processing and may be carried out automatically without making any demands on processing resources. Accordingly, all sources of variability, including talker, rate, and amplitude, are attended to and processed by the listener. The differences in recall performance among the three sources of variability investigated here would simply be due to differences in relative ease of processing. Consequently, although listeners must compensate for changes in overall amplitude, this analysis may require very little time and very few processing resources. The absence of an effect of amplitude variability therefore may simply reflect the ease of processing speech varying along this stimulus dimension.

In regard to the effects of stimulus variability on serial recall in the long ISI condition, an entirely different pattern of results emerges. In this case, for talker variability, our study replicates the findings of Goldinger et al. (1991), in which recall of words in early list positions in multiple-talker lists was superior to recall of words in single-talker lists when listeners were given sufficient processing time to fully encode talker information. Apparently, the distinctive information provided by each of the different voices associated with the words in the list allowed listeners to remember both the word and its temporal position in the list. Thus, it appears that listeners in the present

study and in the Goldinger et al. (1991) experiment were encoding distinctive talker information along with the linguistic content of the spoken words into a memory representation that could be used for serial recall. These findings are consistent with recent findings from our laboratory showing that talker-specific information is retained in long-term memory and can be used not only to aid recognition memory (Palmeri et al., 1993), but also to facilitate the subsequent perceptual analysis of the phonetic content of a talker's novel utterance (see Nygaard et al., 1994). In the latter study, we found that learning to explicitly identify a talker's voice facilitated subsequent perceptual processing of novel words produced by that talker. These findings provide evidence that both a phonetic description of the utterance and a structural description of the characteristics of voice are retained in long-term memory (see also Remez, Fellowes, Pardo, & Rubin, 1993).

In examining the effects of variability in speaking rate at long ISIs, however, we did not obtain the same pattern of results that we found for talker variability. No advantage was observed for multiple speaking rate lists over single speaking rate lists at long ISIs. One reason that speaking rate may not be beneficial at long ISIs is simply that the two different sources of variability may be processed and encoded differently—at least in memory tasks of this type, using isolated words. For example, changes in speaking rate and talker characteristics have very different effects on the acoustic realization of spoken words (e.g., Gay, 1978; Miller, 1987a; Miller & Baer, 1983; Nearey, 1989; Peterson & Barney, 1952). Further, such changes also have very different roles in terms of their properties and functions in speech perception. Talker characteristics are relatively permanent and can provide potentially important information about a talker's gender, dialect, age, physical size, and emotional and physical state (Laver & Trudgill, 1979). Changes in speaking rate, in contrast, are often thought to vary more quickly and may convey prosodic and emotional information, but generally at the level of the phrase or sentence, and not at the level of the word. For this reason, speaking rate may not necessarily be encoded in long-term memory representations in the same manner as talker-specific attributes for a memory task using isolated words. Instead, speaking rate may be used only to assess the phonetic value of linguistic segments and then be discarded or ignored by the listener as irrelevant for this task (see Biederman & Cooper, 1991; Srinivas, 1993). This interpretation implies that rate information in speech may at times be treated in a fundamentally different manner than talker information. Likewise, the lack of any effect of amplitude variability may simply be a consequence of the low information value of this dimension in the encoding and rehearsal of spoken words in the serial recall task. Not only does overall amplitude provide little distinctive information to the listener, but also, it may not even be relevant for the perception of the phonetic content of a talker's utterance.

A similar account has been suggested by Fisher and Cuervo (1983) to explain memory for the physical features of spoken discourse. In a series of experiments, listeners

were presented with passages in which the gender or language of the talker was either relevant or irrelevant to the meaning of the passage. Memory for the physical features of the passages was evaluated by having subjects indicate the language or gender of a series of statements selected from the earlier passage. The results showed that memory for the gender or language of the talker who produced the original passage was superior when those physical features were relevant to the content of the passage. Although these results address the relationship between physical features of speech and discourse comprehension, they also provide an appropriate account for the serial recall task. At the longer ISIs, listeners encode and retain information about a talker's voice because, for isolated words, talker characteristics are highly salient and may provide additional distinctive temporal order cues for the serial recall task. Variations in speaking rate and amplitude, on the other hand, may not be encoded and retained in long-term memory in this task because these stimulus dimensions may not provide additional unique or salient information to the listener at the level of the isolated word.

It should be noted, however, that the lack of an effect at the longer ISIs for speaking rate and amplitude variability should be interpreted with caution. Given the task demands of the serial recall task, we might expect that any perceptible cue to the temporal order of words in the list should be encoded and retained. Thus, it remains unclear why any perceptually salient variation across words would not be used. It is possible that the serial recall task is not sensitive enough to reflect the retention of dimensions such as speaking rate and amplitude in long-term memory. However, speaking rate and amplitude may indeed be retained in a representation of talkers' utterances; consequently, a different task, such as a continuous recognition memory task (Palmeri et al., 1993) or a test of implicit memory (Jacoby, 1983; Jacoby & Brooks, 1984; Jacoby & Hayman, 1987; Schacter, 1990), might reveal this type of representation. In these tasks, the listener might rely more strongly on the physical or surface features of the stimulus. For example, Schacter (1990) has proposed that a perceptual representation system (PRS) that represents highly detailed aspects of a stimulus event underlies tasks that involve an implicit memory component. Recognition or recall tasks are assumed to be subserved primarily by a conceptually driven episodic memory system. Perhaps the serial recall task that was used relied primarily on an episodic memory store that preserves few, if any, surface features of the spoken words. According to this view, information about the surface form of spoken words such as speaking rate and amplitude would not necessarily be encoded into an episodic memory trace and therefore would not be available in this memory task.

Although the sensitivity and nature of the serial recall task may have been an important factor in these experiments, the large effects of variability in speaking rate observed at the shortest ISIs argues against an alternative interpretation. If the serial recall task was not sensitive to changes in speaking rate, there should be no effect of that

variable at the shortest ISIs as well. Likewise, given the lack of effects of amplitude variability in other experimental situations (Church & Schacter, 1994; Sommers et al., 1992, 1994), these results point to differences in the encoding and rehearsal of words varying along each of these dimensions rather than to task-specific factors.

In summary, the present set of findings suggests that all sources of stimulus variability in speech may not be created equally. Variation in the speech signal due to speaking rate, overall amplitude, and talker characteristics appears to be processed and encoded in memory in a fundamentally different manner. Variation in speaking rate appears to affect the initial encoding and categorization of linguistic units but may not be transferred into long-term memory, at least in this serial recall task. Amplitude variability, however, does not appear to be encoded into memory at all; or, alternatively, it may be processed automatically, therefore making few if any demands on limited attentional resources and time. These distinct patterns of results for each stimulus dimension investigated here suggests that each source of variability encountered in the speech signal may engage different adaptive mechanisms and may affect different levels of processing. The perceptual mechanisms used in speech perception and spoken word recognition not only appear to process many kinds of stimulus variation effectively, but also appear to be quite flexible, and, specifically, to process different kinds of variation in different ways. Depending on the relevance of each source of variability and the mandatory nature of the perceptual mechanisms devoted to resolving the variation, long-term memory for spoken words incorporates only certain linguistic and extralinguistic information into detailed representations of a talker's utterance.

REFERENCES

- AMERICAN NATIONAL STANDARDS INSTITUTE. (1971). *Method for measurement of monosyllabic word intelligibility* (American National Standard S3.2-1960 [R1971]). New York: Author.
- ASSMANN, P. F., NEAREY, T. M., & HOGAN, J. T. (1982). Vowel identification: Orthographic, perceptual, and acoustic aspects. *Journal of the Acoustical Society of America*, **71**, 975-989.
- BADDELEY, A. D., & HITCH, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-90). New York: Academic Press.
- BIEDERMAN, I., & COOPER, E. E. (1991). Evidence for complete translational and reflectional invariance in visual object priming. *Perception*, **20**, 585-593.
- CHURCH, B. A., & SCHACTER, D. L. (1994). Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 521-533.
- 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.
- CRAIK, F. I. M., & KIRSNER, K. (1974). The effect of speaker's voice on word recognition. *Quarterly Journal of Experimental Psychology*, **26**, 274-284.
- CREELMAN, C. D. (1957). Case of the unknown talker. *Journal of the Acoustical Society of America*, **29**, 655.
- FANT, G. (1973). *Speech sounds and features*. Cambridge, MA: MIT Press.
- FISHER, R. P., & CUERVO, A. (1983). Memory for physical features of

- discourse as a function of their relevance. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **9**, 130-138.
- FODOR, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.
- GARNER, W. R. (1974). *The processing of information and structure*. Potomac, MD: Erlbaum.
- GAY, T. (1978). Effect of speaking rate on vowel formant movements. *Journal of the Acoustical Society of America*, **63**, 223-230.
- 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 & Verbal Behavior*, **22**, 15-23.
- GLANZER, M., & CUNITZ, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning & Verbal Behavior*, **5**, 351-360.
- GOLDINGER, S. D., PISONI, D. B., & LOGAN, J. S. (1991). On the nature of talker variability effects on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 152-162.
- GREEN, K. P., & MILLER, J. L. (1985). On the role of visual rate information in phonetic perception. *Perception & Psychophysics*, **38**, 269-276.
- HALLE, M. (1985). Speculations about the representation of words in memory. In V. A. Fromkin (Ed.), *Phonetic linguistics* (pp. 101-114). New York: Academic Press.
- JACOBY, L. L. (1983). Perceptual enhancement: Persistent effects of an experience. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **9**, 21-38.
- JACOBY, L. L., & BROOKS, L. R. (1984). Nonanalytic cognition: Memory, perception, and concept learning. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 18, pp. 1-47). New York: Academic Press.
- JACOBY, L. L., & HAYMAN, C. A. G. (1987). Specific visual transfer in word identification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **13**, 456-463.
- JAHNKE, J. C. (1968). Presentation rate and the serial-position effect of immediate serial recall. *Journal of Verbal Learning & Verbal Behavior*, **7**, 608-612.
- JOOS, M. A. (1948). Acoustic phonetics. *Language*, **24**, 1-136.
- KEWLEY-PORT, D. (1983). Time-varying features as correlates of place of articulation in stop consonants. *Journal of the Acoustical Society of America*, **73**, 322-335.
- KLATT, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, **59**, 1208-1221.
- KLATT, D. H. (1985). A shift in formant frequencies is not the same as a shift in the center of gravity of a multiformant energy concentration [Summary]. *Journal of the Acoustical Society of America*, **77**, S7.
- KUHL, P. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, **50**, 93-107.
- KUHL, P. (1992). Psychoacoustics and speech perception: Internal standards, perceptual anchors, and prototypes. In L. A. Werner & E. W. Rubel (Eds.), *Developmental psychoacoustics* (pp. 293-332). Washington, DC: American Psychological Association.
- LADEFOGED, P., & BROADBENT, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America*, **29**, 98-104.
- LAVER, J. (1989). Cognitive science and speech: A framework for research. In H. Schnelle & N. O. Bersen (Eds.), *Logic and linguistics: Research directions in cognitive science. European perspectives* (Vol. 2, pp. 37-70). Hillsdale, NJ: Erlbaum.
- LAVER, J., & TRUDGILL, P. (1979). Phonetic and linguistic markers in speech. In K. R. Scherer & H. Giles (Eds.), *Social markers in speech* (pp. 1-32). Cambridge: Cambridge University Press.
- LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. P., & STUDDERT-KENNEDY, M. (1967). Perception of the speech code. *Psychological Review*, **74**, 431-461.
- LUCE, P. A., & CARRELL, T. D. (1981). *Creating and editing waveforms using WAVES* (Research in Speech Perception, Progress Report No. 7). Bloomington: Indiana University Speech Research Laboratory.
- 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: Human Perception & Performance*, **15**, 676-684.
- MILLER, J. L. (1981). Effects of speaking rate on segmental distinctions. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech* (pp. 39-74). Hillsdale, NJ: Erlbaum.
- MILLER, J. L. (1987a). Rate-dependent processing in speech perception. In A. Ellis (Ed.), *Progress in the psychology of language* (pp. 119-157). Hillsdale, NJ: Erlbaum.
- MILLER, J. L. (1987b). Mandatory processing in speech perception. In J. L. Garfield (Ed.), *Modularity in knowledge representation and natural-language understanding* (pp. 309-322). Cambridge, MA: MIT Press.
- MILLER, J. L., & BAER, T. (1983). Some effects of speaking rate on the production of /b/ and /w/. *Journal of the Acoustical Society of America*, **73**, 1751-1755.
- MILLER, J. L., & LIBERMAN, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics*, **25**, 457-465.
- MILLER, J. L., & VOLAITIS, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*, **46**, 505-512.
- 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. (1989). 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.
- NEAREY, T. (1989). Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America*, **85**, 2088-2113.
- NYGAARD, L. C., SOMMERS, M. S., & PISONI, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, **5**, 42-46.
- OLLER, D. K. (1973). The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America*, **54**, 1235-1247.
- PALMERI, T. J., GOLDINGER, S. D., & PISONI, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 309-328.
- PETERSON, G. E., & BARNEY, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, **24**, 175-184.
- REMEZ, R. E., FELLOWES, J. M., PARDO, J. S., & RUBIN, P. E. (1993). *Voice recognition based on phonetic information*. Paper presented at the meeting of the Psychonomic Society, Washington, DC.
- RUNDIS, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, **89**, 63-77.
- SCHACTER, D. L. (1990). Perceptual representation systems and implicit memory: Toward a resolution of the multiple memory systems debate. In A. Diamond (Ed.), *Development and neural bases of higher cortical functions* (Annals of the New York Academy of Sciences, Vol. 608, pp. 543-571). New York: New York Academy of Sciences.
- SHANKWEILER, D. P., STRANGE, W., & VERBRUGGE, R. R. (1976). Speech and the problem of perceptual constancy. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, knowing: Toward an ecological psychology* (pp. 315-346). Hillsdale, NJ: Erlbaum.
- SOMMERS, M. S., NYGAARD, L. C., & PISONI, D. B. (1992). Stimulus variability and the perception of spoken words: Effects of variations in speaking rate and overall amplitude. In J. J. Ohala, T. M. Nearey, B. L. Derwing, M. M. Hodge, & G. E. Wiebe (Eds.), *ICSLP 92 Proceedings: 1992 International Conference on Spoken Language Processing* (Vol. 1, pp. 217-220). Edmonton, AB: Priority Printing.
- SOMMERS, M. S., NYGAARD, L. C., & PISONI, D. B. (1994). Stimulus variability and spoken word recognition: I. Effects of variability in speaking rate and overall amplitude. *Journal of the Acoustical Society of America*, **96**, 1314-1324.
- SRINIVAS, K. (1993). Perceptual specificity in nonverbal priming. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 582-602.

- STEVENS, K. N., & BLUMSTEIN, S. E. (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America*, **64**, 1358-1368.
- SUMMERFIELD, Q. (1975). *Acoustic and phonetic components of the influence of voice changes and identification times for CVC syllables* (Report on Research in Progress in Speech Perception, Vol. 2, pp. 73-98). Belfast, Northern Ireland: Queen's University.
- SUMMERFIELD, Q. (1981). On articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 1074-1095.
- SUMMERFIELD, Q., & HAGGARD, M. P. (1973). *Vocal tract normalisation as demonstrated by reaction times* (Report on Research in Progress in Speech Perception, Vol. 2, pp. 12-23). 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.
- VOLAITIS, L. E., & MILLER, J. L. (1992). Phonetic prototypes: Influences of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America*, **92**, 723-735.
- WEENINK, D. J. M. (1986). The identification of vowel stimuli from men, women, and children. *Proceedings from the Institute of Phonetic Sciences of the University of Amsterdam*, **10**, 41-54.

NOTE

1. To avoid confusion between the presentation rate and speaking rate manipulations, we refer to presentation rate as interstimulus interval (ISI) throughout the paper. Because presentation rate was varied by changing the ISI, short ISIs between items in a list resulted in fast presentation rates and long ISIs resulted in slow presentation rates. Thus, ISI or presentation rate was manipulated independently from the articulation rate of the words within the list.

(Manuscript received May 9, 1994;
revision accepted for publication March 6, 1995.)