Published in final edited form as:
J Exp Psychol Learn Mem Cogn. 1993 March ; 19(2): 309–328.

Episodic Encoding of Voice Attributes and Recognition Memory for Spoken Words

Thomas J. Palmeri,
Department of Psychology, Indiana University

Stephen D. Goldinger, and
Department of Psychology, Arizona State University

David B. Pisoni
Department of Psychology, Indiana University

Abstract

Recognition memory for spoken words was investigated with a continuous recognition memory task. Independent variables were number of intervening words (lag) between initial and subsequent presentations of a word, total number of talkers in the stimulus set, and whether words were repeated in the same voice or a different voice. In Experiment 1, recognition judgments were based on word identity alone. Same-voice repetitions were recognized more quickly and accurately than different-voice repetitions at all values of lag and at all levels of talker variability. In Experiment 2, recognition judgments were based on both word identity and voice identity. Subjects recognized repeated voices quite accurately. Gender of the talker affected voice recognition but not item recognition. These results suggest that detailed information about a talker's voice is retained in long-term episodic memory representations of spoken words.

The speech signal varies substantially across individual talkers as a result of differences in the shape and length of the vocal tract (Carrell, 1984; Fant, 1973; Summerfield & Haggard, 1973), glottal source function (Carrell, 1984), positioning and control of articulators (Ladefoged, 1980), and dialect. According to most contemporary theories of speech perception, acoustic differences between talkers constitute noise that must be somehow filtered out or transformed so that the symbolic information encoded in the speech signal may be recovered (e.g., Blandon, Henton, & Pickering, 1984; Disner, 1980; Gerstman, 1968; Green, Kuhl, Meltzoff, & Stevens, 1991; Summerfield & Haggard, 1973). In these theories, some type of "talker-normalization" mechanism, either implicit or explicit, is assumed to compensate for the inherent talker variability[1] in the speech signal (e.g., Joos, 1948). Although many theories attempt to describe how idealized or abstract phonetic representations are recovered from the speech signal (see Johnson, 1990, and Nearey, 1989, for reviews), little mention is made of the fate of voice information after lexical access is complete. The talker-normalization hypothesis is consistent with current views of speech perception wherein acoustic-phonetic invariances are sought, redundant surface forms are quickly forgotten, and only semantic information is retained in long-term memory (see Pisoni, Lively, & Logan, 1992).

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

[1] *Talker variability* refers to differences between talkers. All references to talker variability and voice differences throughout this article refer to such between-talker differences. Differences between words produced by the same talker are not implied by this term.

According to the traditional view of speech perception, detailed information about a talker's voice is absent from the representations of spoken utterances in memory. In fact, evidence from a variety of tasks suggests that the surface forms of both auditory and visual stimuli are retained in memory. Using a continuous recognition memory task (Shepard & Teghtsoonian, 1961), Craik and Kirsner (1974) found that recognition memory for spoken words was better when words were repeated in the same voice as that in which they were originally presented. The enhanced recognition of same-voice repetitions did not deteriorate over increasing delays between repetitions. Moreover, subjects were able to recognize whether a word was repeated in the same voice as in its original presentation. When words were presented visually, Kirsner (1973) found that recognition memory was better for words that were presented and repeated in the same typeface. In a parallel to the auditory data, subjects were also able to recognize whether a word was repeated in the same typeface as in its original presentation. Kirsner and Smith (1974) found similar results when the presentation modalities of words, either visual or auditory, were repeated.

Long-term memory for surface features of text has also been demonstrated in several studies by Kolers and his colleagues. Kolers and Ostry (1974) observed greater savings in reading times when subjects reread passages of inverted text that were presented in the same inverted form as an earlier presentation than when the same text was presented in a different inverted form. This savings in reading time was found even 1 year after the original presentation of the inverted text, although recognition memory for the semantic content of the passages was reduced to chance (Kolers, 1976). Together with the data from Kirsner and colleagues, these findings suggest that physical forms of auditory and visual stimuli are not filtered out during encoding but instead remain part of long-term memory representations. In the domain of spoken language processing, these findings suggest a voice-encoding hypothesis, consistent with nonanalytic views of cognition, wherein episodic traces of spoken words retain their surface forms (e.g., Jacoby & Brooks, 1984).

In several recent experiments from our laboratory, investigators have examined the effects of varying the voice of the talker from trial to trial on memory for spoken words. Martin, Mullennix, Pisoni, and Summers (1989) examined serial recall of word lists spoken either by a single talker or by multiple talkers. Recall of items in the primacy portion of the lists was reduced by introducing talker variability; recall of items from the middle or end of the list was not affected. These results were replicated in later studies (Goldinger, Pisoni, & Logan, 1991; Lightfoot, 1989; Logan & Pisoni, 1987).

Reduced recall in the primacy portion of a list is typically assumed to indicate a decrease in the amount or efficiency of short-term memory rehearsal, which affects the success of transferring items into long-term memory (Atkinson & Shiffrin, 1968; Waugh & Norman, 1965). Martin et al. (1989) argued that this reduction in efficiency of rehearsal may arise from two processes. First, perceptual processing of words in multiple-talker lists may require access to a normalization mechanism that may not be used in the processing of single-talker lists. Changing the talker from trial to trial forces the normalization mechanism to recalibrate, which demands greater processing resources. Thus multiple-talker word lists may leave fewer resources available for short-term memory rehearsal, thereby decreasing the success of long-term memory encoding. Second, unique voice-specific information may be encoded in memory along with each of the items. The added variability of the voice information in multiple-talker lists may attract attention and usurp resources from rehearsal, without a normalization process involved (Martin et al., 1989).

Recently, Goldinger et al. (1991) conducted a serial recall experiment to test these alternative hypotheses. They varied the interstimulus interval (ISI) between words in the lists to assess the interaction of talker variability and rehearsal time. For lists with short ISIs, Martin et al.’s (1989) results were replicated. For lists with longer ISIs, however, the effect of talker variability on serial recall was reversed: Words in the primacy portion of multiple-talker lists were recalled as accurately as words in the primacy portion of single-talker lists (see also Lightfoot, 1989). Goldinger et al. suggested that shorter ISIs did not provide sufficient time for encoding both words and voices in long-term memory. Longer ISIs enabled complete encoding and rehearsal of both sources of information. The different voices in multiple-talker lists provided additional retrieval cues, thus improving recall.

The results of Goldinger et al. (1991) suggest that voice information is encoded along with lexical information in the representations of spoken words. In our study, we were interested in measuring how long voice information is retained in memory and in learning more about the nature of the representation of voices in memory. Following Craik and Kirsner’s (1974) procedure, we used a continuous recognition memory task (Shepard & Teghtsoonian, 1961). In this task, subjects were presented with a continuous list of spoken words. Words were presented and later repeated after a variable number of intervening items. The subject judged whether each word was “old” or “new.” Half of the words were presented and later repeated in the same voice, and the others were presented in one voice but later repeated in a different voice. If voice information were encoded along with abstract lexical information, same-voice repetitions would be expected to be recognized faster and more accurately than different-voice repetitions.

By manipulating the number of intervening items between repetitions, we could measure how long voice information is retained in memory. At short lags, recognition judgments could be based on short-term memory; if recognition judgments were enhanced by a match of attributes of the physical stimulus with information stored in short-term memory, subjects would recognize same-voice repetitions faster and more accurately than different-voice repetitions. At longer lags, if a same-voice advantage were still observed, it could be assumed that some voice information must have been encoded into long-term memory. Thus by comparing accuracy and response times for same-voice and different-voice repetitions across a range of lags, we could assess the encoding and retention of voice information.

Beyond the question of voice encoding is a more general issue concerning the nature of the representation of voices in memory. Is the memory of a spoken word an analog representation, true to the physical form of the speech percept, or are more symbolic attributes the primary aspects of the signal that are retained? Geiselman and Bellezza (1976, 1977) proposed a voice-connotation hypothesis to account for the retention of voice characteristics in memory. In a series of experiments, Geiselman and his colleagues found that subjects were able to judge whether repeated sentences were originally presented in a male or a female voice. They argued that the talker’s gender modified the semantic interpretation or connotation of the message (Geiselman & Bellezza, 1976, 1977; Geiselman & Crawley, 1983). In accordance with symbolic views of cognition, Geiselman argued that voice information is encoded through semantic interpretation, rather than as an independent perceptual attribute.

In view of Geiselman’s claim, it is difficult to determine which aspects of voice were retained in memory to improve performance on same-voice trials in the experiments reported by Craik and Kirsner (1974). As Craik and Kirsner noted, only two voices were used (a male and female), and thus either detailed voice information or some sort of more abstract gender code could have been encoded in memory. In our study, we increased the number of talkers in the stimulus set. This enabled us to assess whether the recognition advantage observed for same-voice repetitions was attributable to the retention of gender information or to the retention of more detailed voice characteristics. With more than two talkers, different-voice repetitions can be produced by talkers of either gender. Thus it was possible to determine whether same- and different-gender repetitions produced equivalent recognition deficits. If only gender information were retained in memory, we would expect no differences in recognition between same-voice repetitions and different-voice/same-gender repetitions. However, if more detailed information were retained, we would expect recognition deficits for words repeated in a different voice, regardless of gender.

EXPERIMENT 1

In Experiment 1, we examined continuous recognition memory for spoken words as a function of the number of talkers in the stimulus set, the lag between the initial presentation and repetition of words, and the voices of repetitions. Subjects were required to attend only to word identity; they were told to classify repeated words as “old,” regardless of whether the voice was the same or different.

In accordance with previous results (Craik & Kirsner, 1974; Hockley, 1982; Kirsner, 1973; Kirsner & Smith, 1974; Shepard & Teghtsoonian, 1961), decreased recognition accuracy and increased response times were expected with increases in lag. Of more importance, increased recognition accuracy and decreased response times were expected for same-voice repetitions, in relation to different-voice repetitions. These results would replicate the two-talker results of Craik and Kirsner (1974).

Increasing the number of talkers in the stimulus set also enabled us to assess the separate effects of voice and gender information. Thus we could evaluate the voice-connotation hypothesis by comparing the effects of gender matches and exact voice matches on recognition memory performance. In addition, increasing the number of talkers enabled us to measure perceptual processing deficits caused by changing the talker’s voice from trial to trial. Increasing the amount of stimulus variability could increase the degree of recalibration necessary to normalize for changes in voice, leaving fewer resources available for encoding items into long-term memory. We also included, as a control, a single-talker condition in which each word was spoken by the same talker, which enabled us to assess the effects of talker variability on recognition performance.

Method

Subjects—Subjects were 200 undergraduate students from introductory psychology courses at Indiana University. Forty subjects served in each of five talker variability conditions. Each subject participated in one half-hour session and received partial course credit. All subjects were native speakers of English who reported no history of a speech or hearing disorder at the time of testing.

Stimulus Materials—The stimulus materials were lists of words spoken either by a single talker or by multiple talkers. All items were monosyllabic words selected from the vocabulary of the Modified Rhyme Test (MRT; House, Williams, Hecker, & Kryter, 1965). Each word was recorded in isolation on audiotape and digitized by a 12-bit analog-to-digital converter. The root mean squared amplitude levels for all words were digitally equated.

The lists were constructed from a digital database of 300 MRT words spoken by 20 talkers (10 male and 10 female). Each word was presented and repeated once. The repetition of any given word occurred after a lag of 1, 2, 4, 8, 16, 32, or 64 intervening words; the repetition itself counted as the last of the intervening words. Each lag value was used an equal number of times in each list. At every position, except at the beginning of the list, a new word or a repeated word occurred with equal probability. One-hundred forty pairs of presented and repeated items were used in each list. Each subject was given an initial practice list of 15 words with which to become familiarized with the task; none of these words were repeated in the experiment. The subject was not told that the next 30 words of each list were used to establish a memory load and were not considered in the data analyses. None of these initial 30 words were included in the test portion of the list. Twenty filler words were also randomly distributed throughout the test portion of each list to simplify the algorithm used to generate the proper lag distributions; these words were never repeated and were not considered in the final data analyses. The 15 practice words, 30 load words, 140 test pairs, and 20 filler words constituted a total of 345 spoken words used in each session

We manipulated talker variability by selecting a subset of stimuli from the database of 20 talkers. Single-talker lists were generated by randomly selecting 1 of the 20 talkers as the source of all of the words. We produced multiple-talker lists of 2, 6, 12, and 20 talkers by randomly selecting an equal number of men and women from the pool of 20 talkers. Each talker's voice was used an equal number of times. On the initial presentation of a word, one of the available talkers in this set was selected at random. The probabilities of a same-voice or different-voice repetition of a given word were equal. For the different-voice repetitions, the numbers of male-male, female-female, female-male, and male-female pairings were equated.

Design—The lag between the initial presentation and repetition of a word (1, 2, 4, 8, 16, 32, or 64) and voice of the repetitions (same voice or different voice) were manipulated as within-subject variables. The total number of talkers in the list (1, 2, 6, 12, or 20) was manipulated as a between-subjects variable. Each group of subjects was presented a different list of stimulus words.

Procedure—Subjects were tested in groups of 5 or fewer in a room equipped with sound-attenuated booths used for speech perception experiments. Stimulus presentation and data collection were controlled on-line by a PDP-11/34 minicomputer. Each stimulus word was presented by a 12-bit digital-to-analog converter, low-pass filtered at 4.8 kHz, and presented binaurally over matched and calibrated TDH-39 headphones at 80 dB. After hearing each word, the subject was allowed a maximum of 5 s to respond by pressing either a button labeled *new* if the word was judged new or a button labeled *old* if the word was judged old. A 1-s delay separated response to one word and presentation of the next word. Subjects rested one finger from each hand on the two response buttons and were asked to respond as quickly and as accurately as possible. Subjects were given a maximum of 5 s to respond in each trial. If no response was made, that trial was not recorded. A failure to respond was rarely observed during the experimental trials. Button selections and response times were recorded on-line. Response times were measured from the offset of each stimulus word. No feedback was provided.

Results

In all of the following analyses, a hit was defined as responding “old” to a repeated word and a false alarm as responding “old” to a new word. Unless otherwise stated, all reported response time data are for hits. *Item recognition* refers to judgments of words as old or new.

Item-Recognition Accuracy—We examine first overall performance from the multiple-talker conditions and then an analysis of the single-talker condition and an analysis of the effects of the gender of the talkers for different-voice repetitions.

Overall analysis: Figure 1 displays item-recognition accuracy from all of the multiple-talker conditions for same- and different-voice repetitions as a function of talker variability and lag. Both panels show that recognition performance was better for same-voice repetitions than for different-voice repetitions. As shown in the upper panel, we found no effects of talker variability. However, as shown in the lower panel, recognition performance decreased as lag increased.

We conducted a 4 × 7 × 2 analysis of variance (ANOVA) for the between-subjects variable of talker variability and the within-subject variables of lag and same- and different-voice repetitions (voice). The main effect of talker variability was not significant. Increasing the number of talkers in the set had no effect on recognition hit rates (or on false alarms, as described later), as reflected by the horizontal lines in the upper panel of Figure 1. A significant main effect of lag was observed, $F(6, 936) = 196.16$, $MS_e = 0.014$, $p < .0001$. Increasing lag caused a decrease in item recognition, as shown by the negative slopes of both lines in the lower panel of Figure 1. In all subsequent subanalyses, we failed to find any effect of talker variability. We did consistently find significant effects of lag. The discussion and statistical analyses of these two results are not repeated.

A significant main effect of voice was also observed $F(l, 156) = 186.94$, $MS_e = 0.012$, $p < .0001$. Same-voice repetitions yielded higher hit rates than did different-voice repetitions at all levels of talker variability and at all values of lag, as shown by the separation of the two functions in both panels of Figure 1. The two-way Talker Variability × Voice and Voice × Lag interactions were not significant.

In addition to the main effects, a significant two-way Talker Variability × Lag interaction, $F(18, 936) = 1.86$, $MS_e = 0.014$, $p < .05$, and a significant three-way Talker Variability × Lag × Voice interaction, $F(18, 936) = 2.14$, $MS_e = 0.010$, $p < .01$, were observed. Although statistically significant, the magnitude of these interactions was very small in comparison with that of the main effects of lag and voice. Each interaction accounted for less than 1 % of the variance, in contrast to 55.8% and 7.8% accounted for by lag and voice, respectively.

Single-talker analysis: To assess whether introducing any amount of talker variability would decrease recognition performance, we compared item recognition from the single-talker condition with item recognition from the same-voice repetitions in each of the multiple-talker conditions. We subjected the data to a 5 × 7 (Talker Variability × Lag) ANOVA. As in the analysis of the multiple-talker conditions alone, we found a significant effect of lag, although the main effect of talker variability was not significant. Recognition accuracy in the single-talker condition did not significantly differ from the accuracy of same-voice trials in the multiple-talker conditions.

Gender analysis: To assess the specific effects of gender matches and mismatches on recognition of different-voice repetitions, we conducted an additional analysis on a subset of the data from the multiple-talker conditions. Only the six-, twelve-, and twenty-talker conditions were included in this analysis because only in these conditions were different-voice repetitions spoken by talkers of both genders. Figure 2 displays item-recognition accuracy for same-voice repetitions compared with different-voice/ same-gender and different-voice/different-gender repetitions. As shown in both panels, performance was consistently better for same-voice repetitions than different-voice repetitions. Moreover, we found no difference in performance for different-voice repetitions by talkers of the same gender, in comparison with different-voice repetitions by talkers of a different gender.

We conducted a 3 × 7 × 3 (Talker Variability × Lag × Voice) ANOVA with the recognition data. The analyses included three voice conditions: (a) same-voice, (b) different-voice/same-gender, and (c) different-voice/different-gender. A significant main effect of voice was observed $F(2, 234) = 53.64$, $MS_e = 0.022$, $p < .0001$.

Tukey's Honestly Significant Difference (HSD) analyses revealed that recognition accuracy rates for different-voice/same-gender and different-voice/different-gender repetitions did not differ significantly, although both were significantly lower than recognition accuracy rates for same-voice repetitions (throughout this article, all reported post hoc comparisons were significant at $p < .05$ or less). None of the two-way interactions were significant. No gender effect was observed at any value of lag or any level of talker variability.

Item-Recognition Response Times—As with the accuracy data, we examine first response times from the multiple-talker conditions and then an analysis of the single-talker condition and an analysis of the effects of the gender of the talkers for different-voice repetitions.

Overall analysis: Figure 3 displays the item-recognition response times for hits from all multiple-talker conditions. As shown in both panels, recognition was consistently faster for same-voice repetitions than for different-voice repetitions. The upper panel shows that talker variability had no effect on response times; the lower panel shows that recognition responses, were slower at longer lags.

We conducted a 4 × 7 × 2 (Talker Variability × Lag × Voice) ANOVA with the response times.[2] The main effect of talker variability was not significant. As shown by the horizontal lines in the upper panel of Figure 3, increasing the number of talkers had no significant effect on response times. However, a significant main effect of lag was observed, $F(6, 936) = 161.20$, $MS_e = 0.021$, $p < .0001$.[3] Increasing the lag produced an increase in response times, as shown by the rising curves in the lower panel of Figure 3. In all subsequent analyses, we failed to find any effects of talker variability. We did consistently find significant effects of lag. The discussion and statistical analyses of these two results are not repeated.

A significant main effect of voice was observed, $F(l, 156) = 30.00$, $MS_e = 0.017$, $p < .0001$. Same-voice repetitions yielded shorter response times than did different-voice repetitions at all levels of talker variability, as shown by the separation of lines in both panels of the figure. A significant Lag × Voice interaction was also observed, $F(6, 936) = 7.08$, MS_e 0.016, $p < .0001$. Tukey's HSD analyses revealed that responses were significantly faster for same-voice repetitions than for different-voice repetitions but only at lags of one and two intervening items. However, as shown in the lower panel of Figure 3, responses to same-voice repetitions tended to be faster than those to different-voice repetitions at all values of lag, except at a lag of 64 items. A significant three-way Variability × Lag × Voice interaction was also observed, $F(18, 936) = 2.09$, $MS_e = 0.016$, $p < .01$. The magnitude of this interaction was small and difficult to interpret; it accounted for less than 1% of the variance, in contrast to 53.5% accounted for by lag. The main effect of voice and the Voice × Lag interaction were also quite small in the response time data, accounting for only 1.3% and 1.5% of the variance, respectively.

Single-talker analysis: As with item-recognition accuracy, the response times from the single-talker condition were compared with the response times from the same-voice repetitions of the multiple-talker conditions. Figure 4 displays response times for the single-talker condition and the average response times for the same-voice repetitions of the

[2] It is possible to estimate the amount of time between the initial presentation of a word and a repetition after 64 intervening items. The average length of each stimulus word was 550 ms, the average response time was 1,133 ms, and there was a 1-s delay between trials. This implies 2,683 ms per item. Hence approximately 172 s elapsed between the initial presentation of a word and a repetition 64 items later.

[3] Throughout this article all mean squared error (MS_e) terms are reported in seconds squared, whereas all data in figures are reported in milliseconds.

multiple-talker conditions. As shown in the upper panel, recognition was faster in the single-talker condition (i.e., 1 talker) than in any of the multiple-talker conditions (i.e., 2, 6, 12, or 20 talkers). As shown in the lower panel, recognition was consistently faster in the single-talker condition across all values of lag.

We conducted a 5 × 7 (Talker Variability × Lag) ANOVA with the response times. Although the response times for the single-talker condition tended to be shorter than the response times for the multiple-talker conditions, the main effect of talker variability was not significant. In addition to the main effect of lag, a significant two-way Talker Variability × Lag interaction was observed, $F(24, 1170) = 1.63$, $MS_e = 0.015$, $p < .05$. The differences in response times between the single-talker condition and the same-voice repetitions of the multiple-talker conditions were larger at longer lags.

Gender analysis: As with item-recognition accuracy, an additional analysis with the different-voice repetitions in terms of gender was conducted. In both panels of Figure 5, the same-voice repetitions are compared with different-voice/same-gender and different-voice/different-gender repetitions. As shown in both panels, response times were somewhat shorter for same-voice repetitions than for different-voice repetitions. Gender did not affect response times appreciably except in the six-talker condition. In that condition, responses to different-voice/different-gender repetitions were slightly faster than those to different-voice/same gender repetitions.

We conducted a 3 × 7 × 3 (Talker Variability × Lag × Voice) ANOVA with the response time data. A significant main effect of voice was observed, $F(2, 234) = 4.07$, $MS_e = 0.038$, $p < .05$. Tukey's HSD analyses revealed that the response times for different-voice/same-gender and different-voice/different-gender repetitions did not significantly differ, although both were significantly longer than the response times for same-voice repetitions. There was also a significant two-way Voice × Lag interaction, $F(12, 1404) = 2.39$, $MS_e = 0.037$, $p < .01$. The advantage for same-voice repetitions was clearly present at lags of one or two items. At longer lags, however, there were no clear differences among the three conditions.

Comparison of Hits and False Alarms—Comparison of hits and false alarms provides an assessment of overall item-recognition performance. False alarms were examined to determine whether overall discriminability was affected by increases in talker variability. As shown in Table 1, there were few differences in the false alarm rates and in the false alarm response times across all talker variability conditions. We conducted separate one-way ANOVAs with the false alarm rates and response times over all five levels of talker variability. Because false alarms were responses to new items, they could not be analyzed in terms of lag or voice. The main effect of talker variability was not significant in either analysis. However, the data revealed that false alarms, like hits, were committed some-what faster in the single-talker condition than in the multiple-talker conditions.

Discussion

Several important findings were obtained in Experiment 1: First, increasing lag between the initial and subsequent presentations of a spoken word decreased recognition accuracy and increased response time. This finding replicates those of a number of earlier studies (e.g., Craik & Kirsner, 1974; Hockley, 1982; Kirsner, 1973; Kirsner & Smith, 1974; Shepard & Teghtsoonian, 1961). Decreased recognition performance would be expected with increases in lag because items cannot be completely encoded into long-term memory and because intervening items produce interference. This lag effect has been explained by various memory models in terms of decay of activation over time (Anderson, 1983), increased noise in a common memory substrate that combines all items (Eich, 1985; Murdock, 1982), or changesin context over the course of the experiment (Gillund & Shiffrin, 1984).

The second major finding was that increasing the stimulus variability from two to twenty talkers had no effect on overall recognition accuracy and had little effect on response times. We also found no reliable difference in recognition accuracy between the single-talker condition and the same-voice repetitions of the multiple-talker conditions, although response times were shorter in the single-talker condition, especially at long lags. There appears to have been a constant effect of introducing any amount of talker variability that slows responses but does not affect accuracy. This response time deficit with multiple-talker stimuli is consistent with earlier findings (e.g., Mullennix, Pisoni, & Martin, 1989; Nusbaum & Morin, 1992).

According to several accounts of talker normalization in speech perception, a unique vocal-tract coordinate system is constructed for each new voice (Johnson, 1990; Nearey, 1989). The construction of this coordinate system usurps processing resources from short-term memory rehearsal and other high-level processes (Martin et al., 1989). With more talkers, the voices change more often and more radically, hypothetically creating a need for additional recalibration and decreasing recognition memory performance. We found no such decrease in recognition performance in Experiment 1. Furthermore, if voice information were encoded strategically, increasing the number of talkers from two to twenty should have impaired subjects' ability to process, encode, and retain the voice characteristics of all the talkers. The equivalent performances despite increases in talker variability provide some evidence for the proposal that voice encoding is largely automatic, not strategic.

The absence of talker variability effects in the accuracy data is not inconsistent with a voice-encoding hypothesis. If detailed voice information is encoded into long-term memory representations of spoken words, voice attributes can be used as retrieval cues in recognition. In a familiarity-based model of recognition (e.g., Gillund & Shiffrin, 1984; Hintzman, 1988), words in a list with low talker variability are similar to many previous words because of numerous voice matches, which thus increase their overall level of familiarity. Conversely, words in a list with high talker variability are similar to fewer previous words because of few voice matches, which thus decreases their overall level of familiarity. Because any increase or decrease in familiarity is equivalent for both targets and distractors, no net change in overall recognition performance is predicted when talker variability increases in the recognition task. This hypothesis, however, predicts concomitant increases in both hit rates and false alarm rates, which were not found. Subjects knew that equal numbers of "new" and "old" responses would be required. We assume that subjects adjusted an internal criterion of familiarity to equate the number of "new" and "old" responses (cf. Gillund & Shiffrin, 1984), thereby producing no net change in hit rates and false alarm rates with increases in talker variability.

The third and most important finding from this experiment was that words presented and later repeated in the same voice were recognized faster and more accurately than words presented in one voice but repeated in another voice. The magnitude of the same-voice advantage was independent of the total number of talkers in the stimulus set. Moreover, the same-voice advantage in accuracy was found at all values of lag; it was observed with immediate repetitions and with repetitions after 64 intervening items. The same-voice advantage in response time was large at short lags but was no longer present at a lag of 64 items. This same-voice advantage was obtained without any explicit instructions to the subjects to attend to voice. Furthermore, recognition of different-voice repetitions was not affected by a talker's gender. These findings suggest that some form of detailed voice information, beyond an abstract gender code, is retained in memory over fairly long periods of time.

EXPERIMENT 2

Craik and Kirsner (1974) reported that listeners not only recognized same-voice repetitions more reliably but could also explicitly judge whether repetitions were in the same voice as the original items. Like Craik and Kirsner, we were interested in our subjects' ability to explicitly judge such voice repetitions. As in Experiment 1, the number of talkers in the stimulus set was varied as a between-subjects factor, and the lag and the voices of the repetitions were varied as within-subject factors. In addition to judging whether each word was "old" or "new," subjects also were to determine whether old items were repeated in the same voice or in a different voice. After hearing each word, subjects responded by pressing a button labeled *new* if the word had not been heard before, one labeled *same* if the word had been heard before in the same voice, or one labeled *different* if the word had been heard before in a different voice.

By combining "same" and "different" responses together to produce an "old" response, we could compare the results of Experiments 1 and 2. This provided another way of assessing whether strategic or automatic processes are used to encode voice information. Evidence for voice encoding was found in Experiment 1, even though no explicit instructions to remember voices had been given. We were interested in examining whether explicit instructions to remember voices would increase the amount of voice encoding. Results of previous experiments have suggested that voice information is encoded automatically without any conscious awareness (Geiselman, 1979; Geiselman & Bellezza, 1976, 1977; Mullennix & Pisoni, 1990), and so we anticipated little effect on recognition performance as a result of changing task demands.

Craik and Kirsner (1974) found that subjects could correctly judge voice repetition in a two-voice set with lags of up to 32 intervening items. In Experiment 2, we extended the maximal lag to 64 intervening items. We also increased the number of talkers in the stimulus set, thereby increasing the number of voices to be held in memory. As in Experiment 1, increasing the number of talkers also enabled us to analyze the role of gender in recognition memory for spoken words and voices.

Method

Subjects—One hundred sixty subjects were recruited from the Indiana University community. Each subject participated for one half-hour session and received $5.00 for services. Forty subjects served in each of four conditions of talker variability. All subjects were native speakers of English and reported no history of a speech or hearing disorder at the time of testing.

Stimulus Materials—To prevent fatigue as a result of the additional voice judgment, the experimental lists were shorter than those used in Experiment 1. Only 84 test pairs were used, in contrast to the 140 test pairs used in Experiment 1. Fifteen practice words, 30 load words, 84 test pairs, and 12 filler words constituted a total of 225 spoken words in each session. Only multiple-talker lists were used. Otherwise, the stimulus materials and list generation procedures in Experiment 2 were identical to those used in Experiment 1.

Procedure—All aspects of Experiment 2 concerning test conditions and stimulus presentation were identical to those in Experiment 1. The only procedural difference concerned a change in the response categories: After hearing each word, the subject was allowed a maximum of 5 s to respond by pressing a button labeled *new* if the word had not been heard before, one labeled *same* if the word had been heard before in the same voice, or one labeled *different* if the word had been heard before in a different voice. The index finger from one hand was used to press the *new* button, and the index and middle fingers of the other hand were used to press the *same* and *different* buttons.

Results

The results are discussed first for item recognition and then for voice recognition. *Item recognition* refers to judgments of words as “old” or “new.” *Voice recognition* refers to judgments of voices as “same” or “different.”

Old/New Item Recognition—To assess item recognition performance, we combined “same” and “different” responses into an “old” response category. Thus a hit was defined as responding “same” or “different” to an old item.

Item-Recognition Accuracy: We first discuss an analysis of overall item-recognition accuracy and then compare the results of Experiments 1 and 2. Then, as with Experiment 1, we examine the gender of the talkers for different-voice repetitions.

Overall analysis: Figure 6 displays item-recognition accuracy for same-voice and different-voice repetitions as a function of talker variability and lag. As shown in both panels, recognition performance was better for same-voice repetitions than for different-voice repetitions. The upper panel shows that recognition performance was not affected by increases in talker variability; the lower panel shows that recognition performance decreased as lag increased.

A 4 × 7 × 2 (Talker Variability × Lag × Voice) ANOVA was applied to the data. The main effect of talker variability was not significant. As shown by the horizontal lines in the upper panel of Figure 6, increasing the number of talkers had no effect on recognition performance. However, a significant main effect of lag was observed, $F(6, 936) = 169.56$, $MS_e = 0.025$, $p < .0001$. Increasing the lag decreased recognition, as shown by the negative slopes of the lines in the lower panel of Figure 6.

A significant main effect of voice was also observed, $F(l, 156) = 165.15$, $MS_e = 0.021$, $p < .0001$. Same-voice repetitions produced higher recognition accuracy at all levels of talker variability and at all values of lag, as shown by the separation of the two functions in both panels of Figure 6. A normal approximation of the binomial distribution was used to establish a liberal confidence interval for determining chance performance (0.5 ± 0.03). All rates of recognition displayed in Figure 6 were significantly greater than chance. The main effects of lag and voice accounted for 52.2% and 3.5% of the variance, respectively.

Combined analysis of Experiments 1 and 2: We compared the results of Experiments 1 and 2. To assess the effects of the different task demands on recognition performance, comparison of Figures 1 and 6 reveals that the overall pattern of results was similar across the experiments. Item-recognition accuracy decreased more across lag in Experiment 2 than in Experiment 1. A 2 × 4 × 7 × 2 (Experiment × Talker Variability × Lag × Voice) ANOVA was applied to the combined data from both experiments. Only the main effect and interactions resulting from the differences between experiments are reported. A significant main effect of experiment was observed, $F(1, 312) = 5.30$, $MS_e = 0.070$, $p < .05$. This reflected the generally higher recognition accuracy in Experiment 1 than in Experiment 2. A significant two-way Experiment × Lag interaction was also observed, $F(6, 1872) = 4.36$, $MS_e = 0.020$, $p < .001$. The differences in recognition accuracy between Experiments 1 and 2 were observed primarily at longer lags; Tukey’s HSD analyses revealed a significant difference between experiments only at a lag of 64 items. Although statistically significant, each of the experiment-related effects accounted for less than 1% of the variance.

Gender analysis: As in Experiment 1, we compared the effects of gender matches and mismatches on item-recognition performance. Figure 7 displays item-recognition accuracy for same-voice repetitions compared with different-voice/same-gender and different-voice/different-gender repetitions. As shown in both panels, same-voice repetitions were recognized more accurately than different-voice repetitions, regardless of gender. In addition, different-gender repetitions were recognized more accurately than same-gender repetitions.

We conducted a 3 × 7 × 3 (Talker Variability × Lag × Voice) ANOVA with the accuracy data. As in the overall analyses, a significant main effect of lag was observed; the main effect of talker variability was not significant. More important, a significant main effect of voice was observed, $F(2, 234) = 66.16$, $MS_e = 0.050$, $p < .0001$. Tukey's HSD analyses revealed that same-voice repetitions produced significantly higher item-recognition accuracy than either different-voice/same-gender or different-voice/different-gender repetitions. Different-voice/different-gender repetitions, in turn, produced significantly higher item-recognition accuracy than different-voice/same-gender repetitions. A significant Lag × Voice interaction was also observed, $F(12, 1404) = 3.02$, $MS_e = 0.051$, $p < .001$. As shown in the lower panel of Figure 7, this interaction was attributed largely to the crossover observed at a lag of 32 items. A significant two-way Talker Variability × Lag interaction was also observed, $F(12, 702) = 2.25$, $MS_e = 0.063$, $p < .01$, although Tukey's HSD analyses did not reveal any significant trends.

Item-Recognition Response Times: As with the accuracy data, we first examine overall performance and then compare the results of Experiments 1 and 2 and assess the effects of gender on response times.

Overall analysis: Figure 8 displays item-recognition response times for same- and different-voice repetitions as a function of talker variability and lag. As shown in both panels, same-voice repetitions were recognized faster than different-voice repetitions. The upper panel shows that responses in the six- and twelve-talker conditions were somewhat faster than in the two- and twenty-talker conditions; the lower panel shows that responses were generally slower as lag increased.

We conducted a 4 × 7 × 2 (Talker Variability × Lag × Voice) ANOVA with the response time data.[4] The main effect of talker variability was not significant. Increasing the number of talkers had no significant effect on the speed of correctly recognizing old words. A significant main effect of lag was observed $F(6, 936) = 135.32$, $MS_e = 0.16$, $p < .0001$. Increasing the lag between items slowed responses, as shown by the increasing curves in the lower panel of Figure 8.

A significant main effect of voice was also observed, $F(l, 156) = 43.41$, $MS_e = 0.13$, $p < .0001$. As shown by the separation of the functions in both panels of Figure 8, same-voice repetitions produced faster responses at all levels of talker variability. A significant two-way Voice × Lag interaction was observed, $F(6, 936) = 6.01$, $MS_e = 0.13$, $p < .0001$. This interaction was caused by the crossover of same- and different-voice repetitions at a lag of

[4] It is possible to estimate the amount of time between the initial presentation of a word and a repetition after 64 intervening items. The average length of each stimulus word was 550 ms, the average response time was 1,895 ms, and there was a 1-s delay between trials. This implies 3,445 ms per item. Hence approximately 220 s elapsed between the initial presentation of a word and a repetition 64 items later.

64 items. The Talker Variability × Voice interaction approached significance, $F(18, 936) = 2.67$, $MS_e = 0.13$, $p = .05$. The differences between responses to same- and different-voice repetitions were larger in the two- and twenty-talker conditions. The main effects of lag and voice accounted for 45.1% and 2.0% of the variance, respectively. The two-way Voice × Lag interaction accounted for 1.4% of the variance. The two-way Talker Variability × Voice interaction accounted for less than 1% of the variance.

Combined analysis of Experiments 1 and 2: As with item-recognition accuracy, we compared the response times from Experiments 1 and 2 to assess the effects of the different task demands. Figures 3 and 8 indicate that although the patterns of results were very similar, responses were consistently much slower in Experiment 2 than in Experiment 1. This is not surprising inasmuch as subjects in Experiment 2 were required to explicitly recognize both items and voices. We conducted a 2 × 4 × 7 × 2 (Experiment × Talker Variability × Lag × Voice) ANOVA with the response times. Only the main effect and interactions resulting from differences between experiments are reported: The main effect of experiment was highly significant, $F(l, 312) = 588.83$, $MS_e = 1.16$, $p < .0001$. Overall, response times were several hundred milliseconds faster in Experiment 1 than in Experiment 2. A significant two-way Experiment × Lag interaction was observed, $F(6, 1872) = 48.04$, $MS_e = 0.090$, $p < .0001$. The differences in response times between the two experiments were smaller at short lags than at long lags. A significant two-way Experiment × Voice interaction was also observed, $F(1, 312) = 18.53$, $MS_e = 0.074$, $p < .0001$. The difference in response times between same- and different-voice repetitions was larger in Experiment 2 than in Experiment 1.

Gender analysis: Figure 9 displays response times for same-voice repetitions in comparison with different-voice/same-gender and different-voice/different-gender repetitions. As shown in both panels, same-voice repetitions were recognized faster than all different-voice repetitions. Different-gender repetitions were generally recognized faster than same-gender repetitions.

We conducted a 3 × 7 × 3 (Talker Variability × Lag × Voice) ANOVA with the response time data. A portion of the response time data (4.8%) in this analysis was missing; these values were replaced with the mean response time across subjects for that cell.[5] As in the overall analysis, the main effect of talker variability was not significant, but the main effect of lag was. A significant main effect of voice was also observed, $F(2, 234) = 16.27$, $MS_e = 0.18$, $p < .0001$, although Tukey's HSD analyses did not reveal any significant differences. However, as shown in Figure 9, the same-voice repetitions tended to be recognized faster than both types of different-voice repetitions. We also observed significant interactions of Talker Variability × Voice, $F(4, 234) = 6.45$, $MS_e = 0.18$, $p < .001$; Lag × Voice, $F(12, 1404) = 4.18$, $MS_e = 0.20$, $p < .0001$; and Talker Variability × Lag, $F(12, 702) = 3.80$, $MS_e = 0.20$, $p < .0001$. Tukey's HSD analyses did not reveal any meaningful trends.

Comparison of Hits and False Alarms: False alarm data were examined to determine whether overall performance was affected by increases in talker variability. As shown in

[5] In Experiment 2, only 12 observations were collected from each subject for every value of lag. Of these 12 observations, 6 were same-voice and 6 were different-voice repetitions. Of the 6 different-voice repetitions in conditions with more than 2 talkers, 3 were in voices of the same gender and 3 were in voices of the other gender. With only three observations, several subjects had hit rates of 0.0 for repetitions of a given gender at a given lag. Recall that we report only response times for hits. Including this 0.0 value in the calculation of a mean hit rate is valid, but there is no corresponding response time to report. To conduct our analyses, we calculated mean response times for each condition with all existing values and inserted those mean response times for the missing values. This method decreases the validity of the analyses as more and more missing values are replaced, because each replacement decreases the overall variance. We include these analyses to maintain our approach of reporting parallel analyses of hit rates and response times. The results of such an analysis, however, should be considered carefully as suggestive rather than conclusive evidence.

Table 2, there was little difference between the false alarm rates and false alarm response times across all the talker variability conditions. We conducted separate one-way ANOVAs with the false alarm rates and response times over all four levels of talker variability. The main effect of talker variability was not significant in either analysis.

Same/Different Voice Recognition: We calculated voice recognition accuracy for each subject by dividing the number of correct voice recognitions (i.e., responding "same" to a repetition in the same voice or responding "different" to a repetition in a different voice) by the total number of correct item recognitions (i.e., responding "same" or "different" to a repetition). Thus voice recognition accuracy was defined conditionally as the proportion of correctly recognized voices from the set of correctly recognized items.

Voice Recognition Accuracy

Overall analysis: Figure 10 displays voice recognition accuracy for same- and different-voice repetitions as a function of talker variability and lag. As shown in both panels, voice recognition was more accurate for same-voice repetitions than for different-voice repetitions; that is, subjects more accurately recognized a same-voice repetition as "same" than a different-voice repetition as "different." As shown in the upper panel, increases in talker variability had a small effect on voice recognition. The overall accuracy and the difference between recognizing same- and different-voice repetitions were somewhat greater in the two-talker condition. In addition, there was only a small difference between recognizing the same- and different-voice repetitions in the six-talker condition, in relation to the other conditions. As shown in the lower panel, accuracy dropped off quickly for repetitions after one or two items but then leveled off to above-chance performance at longer lags.

We conducted a 4 × 7 × 2 (Talker Variability × Lag × Voice) ANOVA with the voice recognition accuracy data. A significant main effect of talker variability was observed, $F(3, 156) = 8.50$, $MS_e = 0.064$, $p < .0001$. Tukey's HSD analyses revealed that voice recognition was highest in the two-talker condition; no differences were observed among the six-, twelve-, and twenty-talker conditions. A significant main effect of lag was also observed, $F(6, 936) = 140.89$, $MS_e = 0.042$, $p < .0001$. However, the two-way Talker Variability × Lag interaction was not significant.

A significant main effect of voice was observed, $F(l, 156) = 35.89$, $MS_e = 0.17$, $p < .0001$. Voice recognition for same-voice repetitions was more accurate than for different-voice repetitions, as shown by the separation of the functions in both panels of Figure 10. Although the same-voice and different-voice recognition rates were similar in the six-talker condition, the Talker Variability × Voice interaction was not significant. A significant two-way Lag × Voice interaction was observed, $F(6, 936) = 2.40$, $MS_e = 0.050$, $p < .05$. Tukey's HSD analyses revealed that the differences between same- and different-voice repetitions were not significant at lags of 4 or 64 items. By the confidence interval established earlier (0.5 ± 0.03), all rates of recognition for same-voice repetitions were significantly greater than chance. Of the 28 rates of recognition for different-voice repetitions, 25 were significantly greater than chance.

A significant three-way Talker Variability × Lag × Voice interaction was observed as well, $F(18, 936) = 2.86$, $MS_e = 0.050$, $p < .0001$. Unlike the other three-way interactions reported, this interaction had a clear interpretation: Same-voice repetitions were correctly recognized as "same" more often in the two-talker condition than in the other talker conditions but only at longer lags. In addition, different-voice repetitions were correctly recognized as "different" more often in the two-talker condition than in the other conditions but only at shorter lags. The main effects of talker variability, lag, and voice accounted for 1.7%, 36.9%, and 6.5% of the variance, respectively. The two-way Lag × Voice interaction accounted for less than 1% of the variance. The three-way Talker Variability × Lag × Voice interaction accounted for 1.8% of the variance.

Gender analysis: In both panels of Figure 11, responses to same-voice repetitions are compared with those to different-voice/same-gender and different-voice/different-gender repetitions. As shown in both panels, voice recognition was equivalent for same-voice repetitions and different-voice/different-gender repetitions; that is, subjects recognized same-voice repetitions as "same" as accurately as they recognized different-voice/different-gender repetition as "different." In addition, voice recognition accuracy for different-voice/ same-gender repetitions was less than chance; that is, subjects tended to judge different-voice/ same-gender repetitions as "same" rather than as "different."

The upper panel reveals that there were no effects of talker variability on voice recognition for same-voice repetitions and different-voice/different-gender repetitions. For different -voice /same -gender repetitions, however, "same" judgments were made more often in the six-talker condition than in the twelve- and twenty-talker conditions. The lower panel reveals that voice-recognition accuracy decreased as lag increased for same-voice repetitions and different-voice/different-gender repetitions. For different-voice/same-gender repetitions, however, "same" judgments were made more often at short lags; voice-recognition accuracy was nearly at chance at longer lags.

We conducted a 3 × 7 × 3 (Talker Variability × Lag × Voice) ANOVA for the data. In contrast to the overall analysis, no main effect of talker variability was observed. This was not surprising, because the main effect in the overall analysis was caused primarily by the two-talker condition, which was not included in this ANOVA. A significant main effect of lag was again observed, $F(6, 702) = 24.64$, $MS_e = 0.091$, $p < .0001$. A significant main effect of voice was also observed, $F(2, 234) = 139.65$, $MS_e = 0.18$, $p < .0001$. Tukey's HSD analyses revealed that rates of voice recognition accuracy for same-voice and different-voice/different-gender repetitions did not significantly differ, and both were significantly higher than the rate of accuracy for different-voice/same-gender repetitions. A significant two-way Talker Variability × Lag interaction was also observed, $F(12, 702) = 2.24$, $MS_e = 0.091$, $p < .01$, although Tukey's HSD analyses did not reveal any significant trends. A significant Lag × Voice interaction was also observed, $F(12, 1404) = 20.87$, $MS_e = 0.090$, $p < .0001$. Voice-recognition accuracy for same-voice and different-voice/different-gender repetitions decreased as lag increased. Unexpectedly, voice recognition for different-voice/ same-gender repetitions increased over short lags and then leveled off to approach chance at longer lags. Voice-recognition accuracy for different-voice/same-gender repetitions was less than 0.5 at every value of lag and was especially low at lags of one or two items. Hence subjects tended to recognize these repetitions as "same" even though they were actually repetitions in a different voice.

Voice Recognition Response Times

Overall analysis: Figure 12 displays the response times for correct voice recognitions. As shown both panels, voice recognition was faster for same-voice repetitions than different-voice repetitions. The upper panel shows that response times were not affected by increases in talker variability; the lower panel shows that response times increased as lag increased.

We conducted a 4 × 7 × 2 (Talker Variability × Lag × Voice) ANOVA with the response times for correct voice recognitions. A portion of the response time data (2.6%) in this analysis was missing; these values were replaced with mean response times (see Footnote [5]). The main effect of talker variability was not significant. A significant effect of lag was observed, $F(6, 936) = 111.04$, $MS_e = 0.19$, $p < .0001$, as shown by the general increase of response times in the lower panel of Figure 12.

A significant effect of voice was observed $F(1, 156) = 29.59$, $MS_e = 0.25$, $p < .0001$: "Same" responses to same-voice repetitions were generated faster than "different" responses to different-voice repetitions. A significant two-way Voice × Lag interaction was observed $F(6, 936) = 4.01$, $MS_e = 0.17$, $p < .001$, reflecting the crossover at a lag of 64 items. The main effects of lag and voice accounted for 40.4% and 2.3% of the variance, respectively. The two-way Voice × Lag interaction accounted for less than 1% of the variance.

Gender analysis: In both panels of Figure 13, the response times for voice recognition of same-voice repetitions are compared with the response times for voice recognition of different-voice/same-gender and different-voice/different-gender repetitions. As shown in both panels, voice recognition was faster for same-voice repetitions than for any different-voice repetition. No consistent pattern of results between same-gender and different-gender repetitions was observed.

We conducted a 3 × 7 × 3 (Talker Variability × Lag × Voice) ANOVA with the response time data. A fairly large portion of the response time data (16.8%) in this analysis was missing; these values were replaced with mean response times (see Footnote [5]). A significant main effect of talker variability was observed, $F(2, 117) = 3.55$, $MS_e = 1.91$, $p < .05$, reflecting the slower responses at higher levels of talker variability. As in the overall analysis, a significant main effect of lag was observed. A significant main effect of voice was also observed, $F(2, 234) = 9.62$, $MS_e = 0.28$, $p < .001$. Same-voice repetitions were recognized faster than any different-voice repetitions. A significant two-way Lag × Voice interaction was observed, $F(12, 1404) = 5.95$, $MS_e = 0.20$, $p < .0001$. As shown in Figure 13, most of the difference in response times occurred at short lags. We also observed a significant two-way Talker Variability × Lag interaction $F(12, 702) = 3.47$, $MS_e = 0.21$, $p < .001$; a significant Talker Variability × Voice interaction $F(4, 234) = 2.97$, $MS_e = 0.28$, $p < .05$; and a significant three-way Talker Variability × Lag × Voice interaction, $F(24, 1404) = 1.74$, $MS_e = 0.20$, $p < .05$.

Discussion

In Experiment 2, we replicated the major findings of Experiment 1: First, increasing the number of talkers in the stimulus set had little effect on item recognition. Second, increasing the lag between the initial and subsequent presentations of a word decreased accuracy and increased response time. Third, same-voice repetitions increased recognition accuracy and decreased response time, in relation to different-voice repetitions. Surprisingly, in contrast to Experiment 1, in which we found no differences in item recognition as a result of gender, different-gender repetitions were actually recognized more accurately than same-gender repetitions.

Overall, the pattern of item recognition results was not appreciably affected by the additional explicit voice-recognition task. Decreases in accuracy were found at longer lags in Experiment 2 than in Experiment 1, which indicates some loss as a result of the added voice decision. Inasmuch as the absolute time delays associated with each value of lag were longer in Experiment 2, it should not be surprising that item-recognition accuracy was somewhat lower at long lags. It is remarkable, however, that introducing the voice-recognition task had such a small effect on overall item-recognition accuracy. We did observe a large increase in response times in Experiment 2, in relation to Experiment 1, most likely as a result of the additional decision required in Experiment 2.[6]

We found that subjects were able to accurately recognize whether a word was presented and repeated in the same voice or in a different voice. This result replicates the findings of Craik and Kirsner (1974). The first few intervening items produced the largest decrease in voice-recognition accuracy and the largest increase in response time; the change was much more gradual after about four intervening items. These results suggest that subjects would be able to explicitly recognize voices after more than 64 intervening words. Moreover, the results suggest that surface information of spoken words, specifically source information, is retained and is accessible in memory for relatively long periods of time.

With only two talkers (a male and a female), voice recognition was more accurate for same-voice repetitions than for different-voice repetitions. Same-voice repetitions were recognized as “same” more quickly and accurately than different-voice repetitions were recognized as “different.” Surprisingly, these results differ from those reported by Craik and Kirsner (1974), who found no such difference in voice judgments. However, we used a larger set of lag values and a larger number of trials, and we tested a larger number of subjects per condition than did Craik and Kirsner (1974). As a result, we believe that our results are reliable and reflect meaningful differences in voice judgment.

With more than two talkers, when different-voice repetitions were analyzed in terms of gender, an interesting pattern of results emerged: Subjects were able to recognize same-voice repetitions as “same” as well as they could recognize different-voice/different-gender repetitions as “different.” However, there was a strong tendency to classify different-voice/same-gender repetitions as “same,” especially at short lags. Thus, when making explicit voice-recognition judgments, listeners relied on overall similarity of the voices, gender being the most salient dimension of similarity (Goldinger, 1992; McGehee, 1937). Even when the repetition was the very next word in the list, subjects relied on gender, classifying a word spoken by a different talker as “same” whenever the talker was of the same gender.

GENERAL DISCUSSION

These experiments extend previous results and provide several new insights into how the human voice is encoded and accessed in memory. First, the retention of voice attributes is largely automatic and does not rely on conscious or strategic processes. Second, voice information is retained in long-term memory for fairly long periods of time and is available for explicit use. Third, the encoding of voice information accounts for many details, not just gender. Fourth, the presence of detailed voice attributes affects recognition memory for spoken words.

The results from both experiments provide evidence that voice information is encoded automatically. First, if voice information were encoded strategically, increasing the number of talkers from two to twenty should have impaired subjects’ ability to process and encode voice information; however, we found little or no effect of talker variability on item recognition in either experiment. Second, the increased task demands in Experiment 2 should have affected performance. If voice information were encoded strategically, explicit instructions to attend to and remember voices should have increased the accuracy for same-voice repetitions and increased the difference in performance between same-voice and different-voice repetitions; however, we found a consistent same-voice advantage in recognition accuracy across all levels of talker variability and all values of lag in both experiments.[7] In addition, the proposal that voices are encoded in an automatic and

[6] One surprising result found in both experiments was our failure to find a same-voice advantage in response time at a lag of 64 items, even though there was an advantage in accuracy. It is not clear whether this is a true effect or an artifact of measurement. One complicating factor is that the response time data were for hits. Because accuracy decreased at longer lags, there were fewer observations per cell for response times, which possibly contributed to the surprising result.

obligatory manner is also consistent with previous research (Geiselman, 1979; Geiselman & Bellezza, 1976, 1977; Geiselman & Crawley, 1983; Mullennix & Pisoni, 1990).

Subjects recognized same-voice repetitions more accurately than they did different-voice repetitions, demonstrating that voice information is encoded into memory and affects later performance. Improved recognition as a result of repetition of surface forms has been shown in a number of previous studies: Subjects exhibit increased recognition accuracy for words presented and repeated in the same modality, either auditory or visual (Kirsner & Craik, 1971; Kirsner & Smith, 1974); for words presented and repeated in the same typeface or case (Hintzman, Block, & Inskeep, 1972; Kirsner, 1973); and for sentences presented and repeated in the same inverted direction (Kolers & Ostry, 1974; Masson, 1984). Apparently, so-called redundant surface information, such as sensory modality, typeface, direction of text, or voice, is encoded and retained in memory and affects later performance, regardless of whether the task is implicit or explicit (see also Goldinger, 1992).

In Experiment 2, subjects were able to recognize whether words were repeated in the same voice or in a different voice. This finding demonstrated that voice information is available for explicit recognition. Explicit judgments about the attributes of repeated stimuli have been shown in previous studies: Subjects are able to recognize whether repeated letter strings were originally written in the same case or typeface (Hintzman et al., 1972; Kirsner, 1973); whether repeated words were originally presented in the same modality (Hintzman et al., 1972; Kirsner & Smith, 1974); whether verbal concepts were originally presented as pictures or visual words (Light, Stansbury, Rubin, & Linde, 1973); and whether utterances were originally spoken by a male or female talker (Craik & Kirsner, 1974; Hintzman et al., 1972; Light et al., 1973).

Results from both experiments suggest that detailed voice information, not merely gender information, constitutes part of the long-term memory representations of spoken words. If only gender codes were prominent in memory representations of spoken words, as suggested by the voice-connotation hypothesis (Geiselman, 1979; Geiselman & Bellezza, 1976, 1977; Geiselman & Crawley, 1983), item recognition should have been better for words repeated by talkers of the same gender than for words repeated by talkers of a different gender. However, in both experiments, we found a clear disadvantage in recognition of different-voice repetitions, regardless of gender. It appears that something much more detailed than a gender code or connotative information about a talker's voice is retained in memory. Perhaps an analog representation of the spoken word or perhaps some record of the perceptual operations used to recognize speech signals would better characterize the episodic trace of a spoken word (e.g., Jacoby & Brooks, 1984; Klatt, 1979; Kolers, 1976; Schacter, 1990). Further research is necessary for determining the level of detail of voice encoding in long-term memory.

Over the past several years, Jacoby and his colleagues have argued that perception can rely on memory for prior episodes (Jacoby, 1983a, 1983b; Jacoby & Brooks, 1984; Jacoby & Hayman, 1987; Johnston, Dark, & Jacoby, 1985; see also Schacter, 1990). For example, Jacoby and Hayman (1987) found that prior presentation of a word improved later perceptual identification of that word when specific physical details were retained. The ease with which a stimulus is perceived is often called *perceptual fluency* and depends on the degree of physical overlap between representations stored at study and stimuli presented at test. Jacoby and Brooks (1984) argued that perceptual fluency can also play an important

[7] The differences in item recognition accuracy between Experiments 1 and 2 occurred primarily at longer lags. It is difficult to determine whether this difference was caused by the demands of the voice-recognition task or by the increase in the absolute delay associated with each lag in Experiment 2.

role in recognition memory judgments (see also Mandler, 1980). Stimuli that are easily perceived seem more familiar and are thus more likely to be judged as having previously occurred.

Because repetitions of visual details play an important role in visual word recognition (Jacoby & Hayman, 1987), it seems reasonable that repetitions of auditory details, such as attributes of a talker's voice, should also contribute to recognition of and memory for spoken words. In our experiments, same-voice repetitions physically matched previously stored episodes. These repetitions presumably resulted in greater perceptual fluency and were, in turn, recognized with greater speed and accuracy than different-voice repetitions. Increases in perceptual fluency apparently depend on repetition of very specific auditory details, such as exact voice matches, and not on categorical similarity, such as simple gender matches.

Using words spoken by different talkers, Goldinger (1992) recently conducted a series of explicit and implicit memory experiments. The similarity of the voices was measured directly with multidimensional scaling techniques. As in our experiments, same-voice advantages were consistently obtained in both explicit recognition memory and implicit perceptual identification tasks. For different-voice repetitions, however, similarity of the repeated voice to the original voice produced different effects in the two tasks. In explicit recognition, repetitions by similar voices produced only small increases in accuracy in relation to repetitions by dissimilar voices, which is consistent with our results. In implicit perceptual identification, in contrast repetitions by similar voices produced substantial increases in accuracy in relation to repetitions by dissimilar voices.

Geiselman and Bjork (1980) argued that voice information is encoded as a form of intraitem context. Just as preservation of extraitem context, such as the experimental room, can affect memory (Smith, Glenberg, & Bjork, 1978), intraitem context, such as voice, modality, or typeface, can also affect memory. If recognition depends on the degree to which intraitem aspects of context, such as voice, are reinstated at the time of testing, similarity of voices should result in similarity of context. Hence repetitions by a talker of the same gender as the original talker should reinstate intraitem context to a greater degree than should repetitions by a talker of the other gender, because gender is the most salient dimension on which voices differ (Carterette & Barnebey, 1975; Goldinger, 1992; McGehee, 1937). However, in both of our experiments, item recognition did not improve for repetitions produced by a similar voice; only exact repetitions provided an improvement in performance. Thus voice is not a contextual aspect of a word; rather, we argue that it is an integral component of the stored memory representation itself (see Glenberg & Adams, 1978; Goldinger, 1992; Mullennix & Pisoni, 1990).

When the task was explicit voice recognition, subjects judged whether the voice was the same or different by comparing the most salient aspect of the original talker, gender, with the most salient aspect of the current talker. Similarity of voices was an important factor determining voice-recognition accuracy. When the repeated voice was of the other gender, subjects recognized the voice as different quite easily. However, when the repeated voice was of the same gender, subjects were unable to discriminate it from the original voice. Perceptual fluency, which is dependent on repetition of specific auditory details, appears to be the basis for item recognition. Similarity judgment, which is dependent on salient perceptual attributes such as gender or dialect, appears to be the basis for explicit voice recognition (see Hecker, 1971).

Current models of spoken word recognition, such as the Fuzzy Logical Model of Perception (Massaro, 1987), LAFS (Klatt, 1979), Cohort Theory (Marslen-Wilson, 1990), the Neighborhood Activation Model (Luce, 1986; Luce, Pisoni, & Goldinger, 1990), and

TRACE (McClelland & Elman, 1986), do not account for stimulus variability, such as differences across talkers. In most of these theories, it is assumed, either explicitly or implicitly, that an early talker normalization process removes or reduces variability from the speech signal. Word recognition is assumed to operate on clean, idealized canonical representations of the spoken utterance that are devoid of surface variability. Our results and other recent findings (e.g., Goldinger, 1992; Goldinger et al., 1991; Martin et al., 1989) demonstrate that detailed voice information is encoded into long-term memory and may later facilitate recognition for spoken words in a variety of tasks.

Our findings also have implications for theoretical accounts of talker normalization in speech perception. A distinction between extrinsic and intrinsic normalization has been proposed in the literature (Johnson, 1990; Nearey, 1989; Nusbaum & Morin, 1992). With extrinsic normalization, vowels are rescaled with reference to a coordinate system constructed from previous vowels spoken by a specific talker (Disner, 1980; Gertsman, 1968; Joos, 1948; Ladefoged & Broadbent, 1957). Increasing the number of talkers should have caused a decrease in recognition performance because the processing resources used for recalibration of the normalization mechanism are not available for memory processes of encoding and retrieval (Martin et al., 1989). We found no evidence of such a decrease. Furthermore, implicit in these accounts of normalization is the loss of stimulus variability from memory representations. We found no evidence for such a loss either.

With intrinsic normalization, in contrast, it is assumed that phonetic identity can be recovered from various static and dynamic properties of the stimulus, without reference to other examples of the talker's speech (Assmann, Nearey, & Hogan, 1982; Fowler, 1986; Fowler, in press; Nearey, 1989; Shankweiler, Strange, & Verbrugge, 1977; Strange, Verbrugge, Shankweiler, & Edman, 1976; Verbrugge, Strange, Shankweiler, & Edman, 1976). Hence rescaling of the speech signal is not necessary because recoverable, invariant cues are present across all talkers. This account of normalization is more consistent with our results because it does not involve the loss of voice-specific attributes from the speech signal. It is possible, however, that both types of normalization may operate together (Johnson, 1990; Nearey, 1989; Nusbaum & Morin, 1992). It is apparent from our data that if either extrinsic or intrinsic normalization occurs, source information remains an integral part of the long-term memory representation of spoken words.

In summary, our results demonstrate that attributes of a talker's voice play an important role in recognition and are part of the representation of spoken words in memory. Same-voice repetitions are recognized faster and more accurately than are different-voice repetitions because voice information is retained in long-term episodic representations. Our findings constrain the types of normalization process that are plausible in theories of speech perception. Normalization does not discard information about a talker's voice; rather, attributes of a talker's voice appear to be represented in memory along with the phonetic (i.e., symbolic) form of spoken words. Voice encoding appears to be an integral part of speech perception: Words and voices are encoded into robust, multidimensional representations in long-term memory that subserve a variety of functions in spoken language processing.

Acknowledgments

This research was supported by National Institutes of Health Research Grant DC-00111-16 to Indiana University, Bloomington.

We thank Fergus Craik, Edward Geiselman, Leah Light, Scott Lively, Lynne Nygaard, Mitch Sommers, and Richard Shiffrin for their valuable comments and criticisms. We also thank Kristin Lively for collecting data in Experiment 2.

References

Anderson, JR. The Architecture of Cognition. Cambridge, MA: Harvard University Press; 1983.

Assmann PR, Nearey TM, Hogan JT. Vowel identification: Orthographic, perceptual, and acoustic aspects. Journal of the Acoustical Society of America. 1982; 71:975–989. [PubMed: 7085986]

Atkinson, RC.; Shiffrin, RM. Human memory: A proposed system and its control processes. In: Spence, KW.; Spence, JT., editors. The psychology of learning and motivation. Vol. Vol. 2. New York: Academic Press; 1968. p. 89-105.

Blandon RAW, Henton CG, Pickering JB. Towards an auditory theory of speaker normalization. Language and Communication. 1984; 4:59–69.

Carrell, TD. Contributions of fundamental frequency, formant spacing, and glottal waveform to talker identification. Bloomington: Indiana University, Department of Psychology; 1984. (Research on Speech Perception Tech. Rep. No. 5)

Carterette, EC.; Barnebey, A. Recognition memory for voices. In: Cohen, A.; Nooteboom, SG., editors. Structure and processes in speech perception. New York: Springer-Verlag; 1975. p. 246-265.

Craik FIM, Kirsner K. The effect of speaker's voice on word recognition. Quarterly Journal of Experimental Psychology. 1974; 26:274–284.

Disner SF. Evaluation of vowel normalization procedures. Journal of the Acoustical Society of America. 1980; 67:253–261. [PubMed: 7354193]

Eich JM. Levels of processing, encoding specificity, elaboration, and CHARM. Psychological Review. 1985; 92:1–38. [PubMed: 3983302]

Fant, G. Speech sounds and features. Cambridge, MA: MIT Press; 1973.

Fowler CA. An event approach to the study of speech perception from a direct-realist perspective. Journal of Phonetics. 1986; 14:3–28.

Fowler, CA. Listener-talker attunements in speech. In: Tighe, T.; Moore, B.; Santroch, J., editors. Human development and communication sciences. Hillsdale, NJ: Erlbaum; (in press)

Geiselman RE. Inhibition of the automatic storage of speaker's voice. Memory & Cognition. 1979; 7:201–204.

Geiselman RE, Bellezza FS. Long-term memory for speaker's voice and source location. Memory & Cognition. 1976; 4:483–489.

Geiselman RE, Bellezza FS. Incidental retention of speaker's voice. Memory & Cognition. 1977; 5:658–665.

Geiselman RE, Bjork RA. Primary versus secondary rehearsal in imagined voices: Differential effects on recognition. Cognitive Psychology. 1980; 12:188–205. [PubMed: 7371376]

Geiselman RE, Crawley JM. Incidental processing of speaker characteristics: Voice as connotative information. Journal of Verbal Learning and Verbal Behavior. 1983; 22:15–23.

Gerstman LJ. Classification of self-normalized vowels. IEEE Transactions on Audio and Electroacoustics, AU-16. 1968:78–80.

Gillund G, Shiffrin RM. A retrieval model for both recognition and recall. Psychological Review. 1984; 91:1–67. [PubMed: 6571421]

Glenberg A, Adams F. Type I rehearsal and recognition. Journal of Verbal Learning and Verbal Behavior. 1978; 17:455–463.

Goldinger, SD. Words and voices: Implicit and explicit memory for spoken words. Bloomington: Indiana University, Department of Psychology; 1992. (Research on Speech Perception Tech. Rep. No. 7)

Goldinger SD, Pisoni DB, Logan JS. On the nature of talker variability effects on recall of spoken word lists. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1991; 17:152–162.

Green KP, Kuhl PK, Meltzoff AN, Stevens EB. Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the McGurk effect. Perception & Psychophysics. 1991; 50:524–536. [PubMed: 1780200]

Hecker, MHL. Speaker recognition: An interpretive survey of the literature. Washington, DC: American Speech and Hearing Association; 1971. (ASHA Monographs, No. 16)

Hintzman DL. Judgments of frequency and recognition memory in a multiple-trace memory model. Psychological Review. 1988; 95:528–551.

Hintzman DL, Block RA, Inskeep NR. Memory for mode of input. Journal of Verbal Learning and Verbal Behavior. 1972; 11:741–749.

Hockley WE. Retrieval processes in continuous recognition. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1982; 8:497–512.

House AS, Williams CE, Hecker MHL, Kryter KD. Articulation-testing methods: Consonantal differentiation with a closed-response set. Journal of the Acoustical Society of America. 1965; 37:158–166. [PubMed: 14265103]

Jacoby LL. Perceptual enhancement: Persistent effects of an experience. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1983a; 9:21–38.

Jacoby LL. Remembering the data: Analyzing interactive processes in reading. Journal of Verbal Learning and Verbal Behavior. 1983b; 22:485–508.

Jacoby, LL.; Brooks, LR. Nonanalytic cognition: Memory, perception, and concept learning. In: Bower, G., editor. The psychology of learning and motivation. Vol. Vol. 18. New York: Academic Press; 1984. p. 1-47.

Jacoby LL, Hayman CAG. Specific visual transfer in word identification. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1987; 13:456–463.

Johnson K. The role of perceived speaker identify in F0 normalization of vowels. Journal of the Acoustical Society of America. 1990; 88:642–654. [PubMed: 2212287]

Johnston WA, Dark VJ, Jacoby LL. Perceptual fluency and recognition judgments. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1985; 11:3–11.

Joos MA. Acoustic phonetics. Language. 1948; 24(Suppl. 2):1–136.

Kirsner K. An analysis of the visual component in recognition memory for verbal stimuli. Memory & Cognition. 1973; 1:449–453.

Kirsner K, Craik FIM. Naming and decision processes in short-term recognition memory. Journal of Experimental Psychology. 1971; 88:149–157. [PubMed: 5577170]

Kirsner K, Smith MC. Modality effects in word identification. Memory & Cognition. 1974; 2:637–640.

Klatt DH. Speech perception: A model of acousticphonetic analysis and lexical access. Journal of Phonetics. 1979; 7:279–312.

Kolers PA. Reading a year later. Journal of Experimental Psychology: Human Learning and Memory. 1976; 2:554–565.

Kolers PA, Ostry DJ. Time course of loss of information regarding pattern analyzing operations. Journal of Verbal Learning and Verbal Behavior. 1974; 13:599–612.

Ladefoged P. What are linguistic sounds made of? Language. 1980; 56:485–502.

Ladefoged P, Broadbent DE. Information conveyed by vowels. Journal of the Acoustical Society of America. 1957; 29:98–104.

Light LL, Stansbury C, Rubin C, Linde S. Memory for modality of presentation: Within-modality discrimination. Memory & Cognition. 1973; 3:395–400.

Lightfoot, N. Effects of talker familiarity on serial recall of spoken word lists. Bloomington: Indiana University, Department of Psychology; 1989. (Research on Speech Perception Progress Rep. No. 15)

Logan, JS.; Pisoni, DB. Talker variability and the recall of spoken word lists: A replication and extension. Bloomington: Indiana University, Department of Psychology; 1987. (Research on Speech Perception Progress Rep. No. 13)

Luce, PA. Neighborhoods of words in the mental lexicon. Bloomington: Indiana University, Department of Psychology; 1986. (Research on Speech Perception Tech. Rep. No. 6)

Luce, PA.; Pisoni, DB.; Goldinger, SD. Similarity neighborhoods of spoken words. In: Altmann, D., editor. Cognitive representation of speech. Cambridge, MA: MIT Press; 1990. p. 122-147.

Mandler G. Recognizing: The judgment of previous occurrence. Psychological Review. 1980; 87:252–271.

Marslen-Wilson, W. Activation, competition, and frequency in lexical access. In: Altmann, GTM., editor. Cognitive models of speech processing: Psycholinguistic and computational perspectives. Cambridge, MA: MIT Press; 1990. p. 148-172.

Martin CS, Mullennix JW, Pisoni DB, Summers WV. Effects of talker variability on recall of spoken word lists. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1989; 15:676–684.

Massaro, DW. Speech perception by ear and eye: A paradigm for psychological inquiry. Hillsdale, NJ: Erlbaum; 1987.

Masson MEJ. Memory for the surface structure of sentences: Remembering with and without awareness. Journal of Verbal Learning and Verbal Behavior. 1984; 23:579–592.

McClelland JL, Elman JL. The TRACE model of speech perception. Cognitive Psychology. 1986; 18:1–86. [PubMed: 3753912]

McGehee F. The reliability of the identification of the human voice. Journal of General Psychology. 1937; 17:249–271.

Mullennix JW, Pisoni DB. Stimulus variability and processing dependencies in speech perception. Perception & Psychophysics. 1990; 47:379–390. [PubMed: 2345691]

Mullennix JW, Pisoni DB, Martin CS. Some effects of talker variability on spoken word recognition. Journal of the Acoustical Society of America. 1989; 85:365–378. [PubMed: 2921419]

Murdock BB. A theory for the storage and retrieval of item and associative information. Psychological Review. 1982; 89:609–626.

Nearey TM. Static, dynamic, and relational properties in vowel perception. Journal of the Acoustical Society of America. 1989; 85:2088–2113. [PubMed: 2659638]

Nusbaum, HC.; Morin, TM. Paying attention to differences among talkers. In: Tohkura, Y.; Vatikiotis-Bateson, E.; Sagisaka, Y., editors. Speech perception, production and linguistic structure. Tokyo: IOS Press; 1992. p. 113-134.

Pisoni, DB.; Lively, SE.; Logan, JS. Variability and invariance in speech perception and its role in perceptual learning: A new look at an old problem; Paper presented at the Workshop on Cross-language Speech Perception; University of South Florida, Tampa. 1992 May.

Schacter DL. Perceptual representation systems and implicit memory: Toward a resolution of the multiple memory systems debate. Annals of the New York Academy of Sciences. 1990; 608:543–571. [PubMed: 2075961]

Shankweiler, D.; Strange, W.; Verbrugge, R. Speech and the problem of perceptual constancy. In: Shaw, R.; Bransford, J., editors. Perceiving, acting, and knowing: Toward an ecological psychology. Hillsdale, NJ: Erlbaum; 1977. p. 315-345.

Shepard RN, Teghtsoonian M. Retention of information under conditions approaching a steady state. Journal of Experimental Psychology. 1961; 62:302–309. [PubMed: 13911664]

Smith SM, Glenberg AM, Bjork RA. Environmental context and human memory. Memory & Cognition. 1978; 6:342–353.

Strange W, Verbrugge RR, Shankweiler DP, Edman TR. Consonant environment specifies vowel identity. Journal of the Acoustical Society of America. 1976; 60:213–224. [PubMed: 956528]

Summerfield, Q.; Haggard, MP. Vocal tract normalisation as demonstrated by reaction times. Vol. Vol. 2. Belfast, Ireland: The Queen's University of Belfast, Department of Psychology; 1973. p. 1-12.(Report on Research in Progress in Speech Perception

Verbrugge RR, Strange W, Shankweiler DP, Edman TR. What information enables a listener to map a talker's vowel space? Journal of the Acoustical Society of America. 1976; 60:198–212. [PubMed: 956527]

Waugh NC, Norman DA. Primary memory. Psychological Review. 1965; 72:89–104. [PubMed: 14282677]

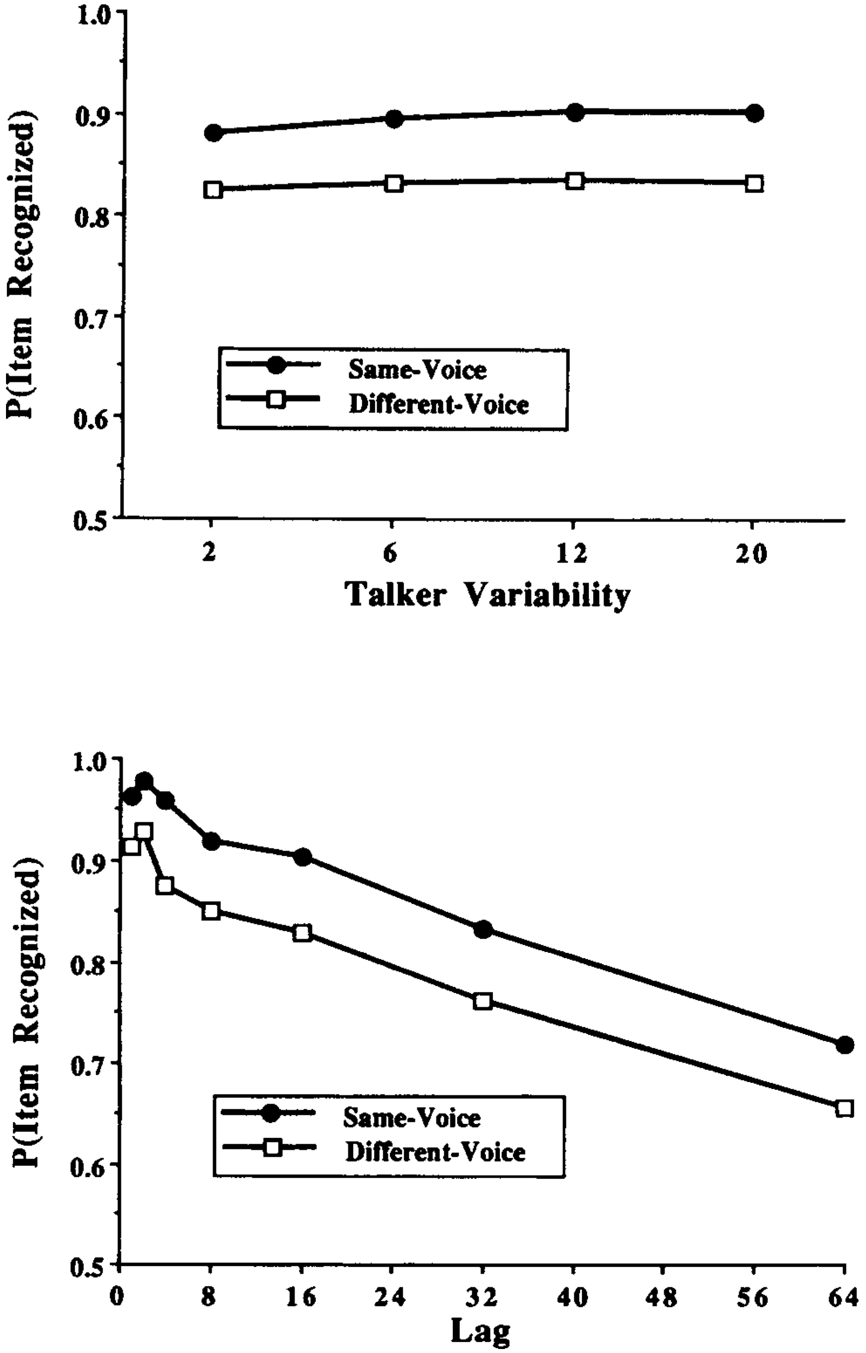

Figure 1.

Probability of correctly recognizing old items from all multiple-talker conditions in Experiment 1. (The upper panel displays item recognition for same-voice repetitions and different-voice repetitions as a function of talker variability, collapsed across values of lag; the lower panel displays item recognition for same- and different-voice repetitions as a function of lag, collapsed across levels of talker variability.)

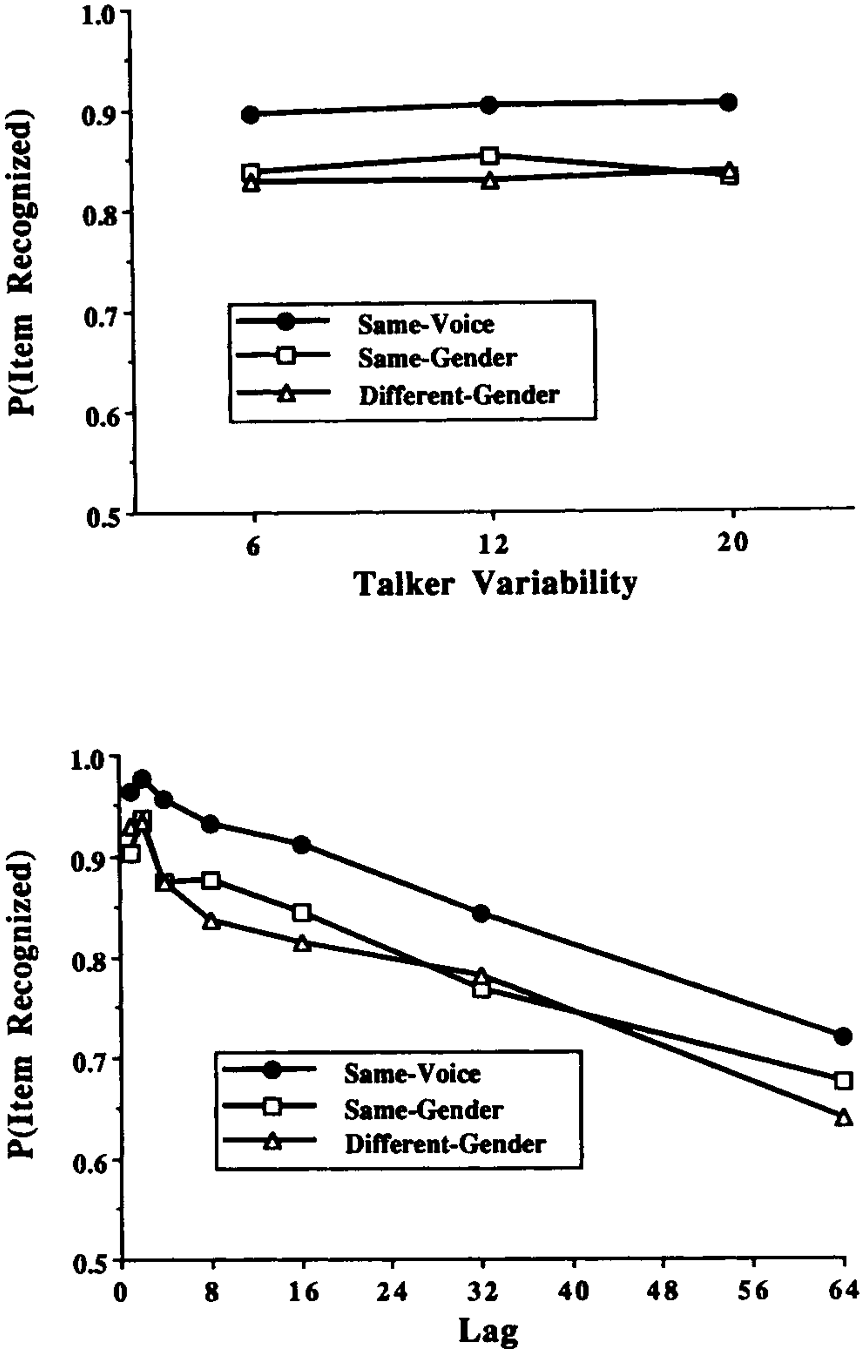

Figure 2.

Probability of correctly recognizing old items from a subset of the multiple-talker conditions in Experiment 1. (In both panels, item recognition for same-voice repetitions is compared with item recognition for different-voice/same-gender and different-voice/different-gender repetitions. The upper panel displays item recognition as a function of talker variability, collapsed across values of lag; the lower panel displays item recognition as a function of lag, collapsed across levels of talker variability.)

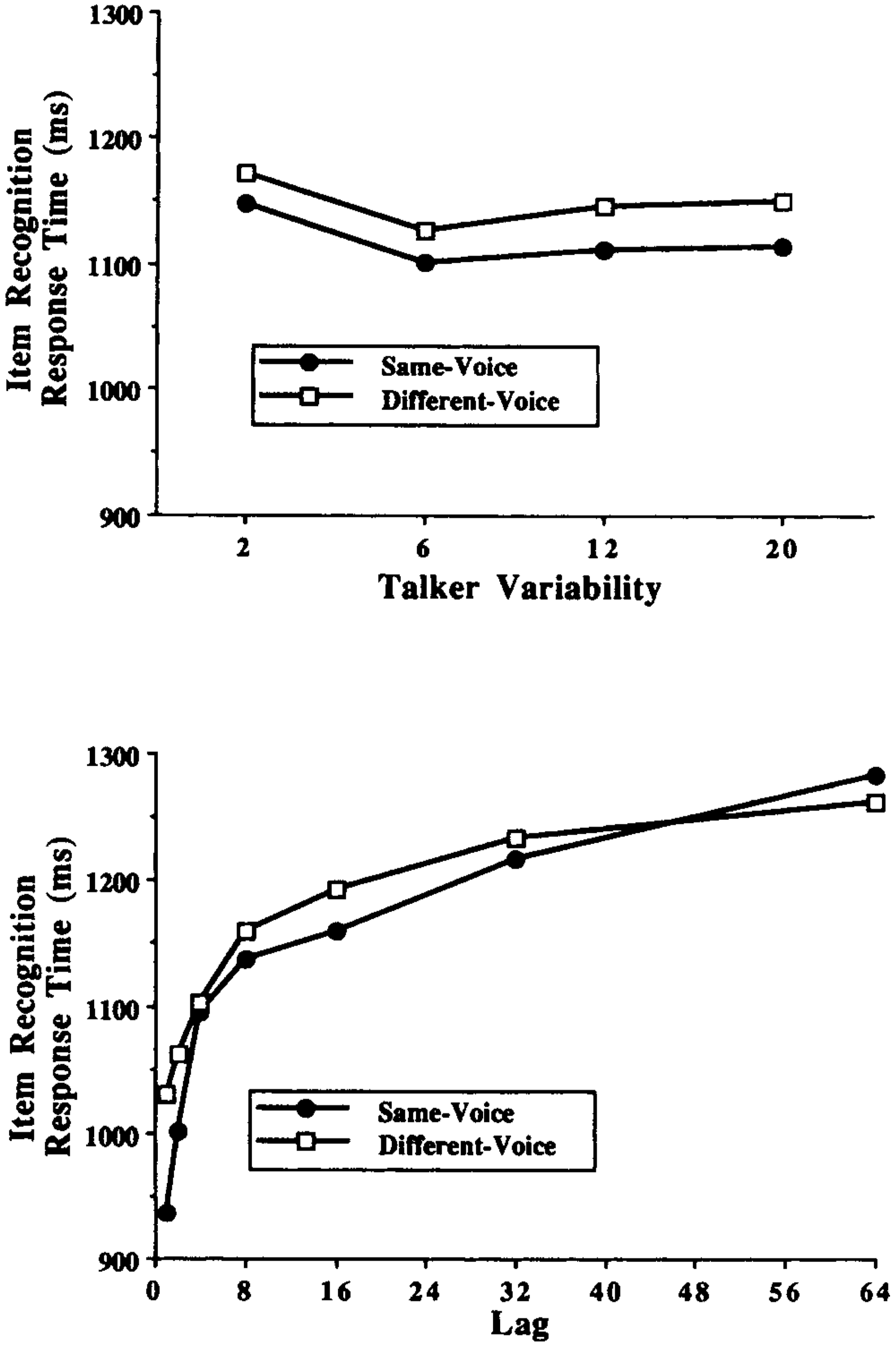

Figure 3.

Response times for correctly recognizing old items from all multiple-talker conditions in Experiment 1. (The upper panel displays response times for same-voice repetitions and different-voice repetitions as a function of talker variability, collapsed across values of lag; the lower panel displays the response times for same- and different-voice repetitions as a function of lag, collapsed across levels of talker variability.)

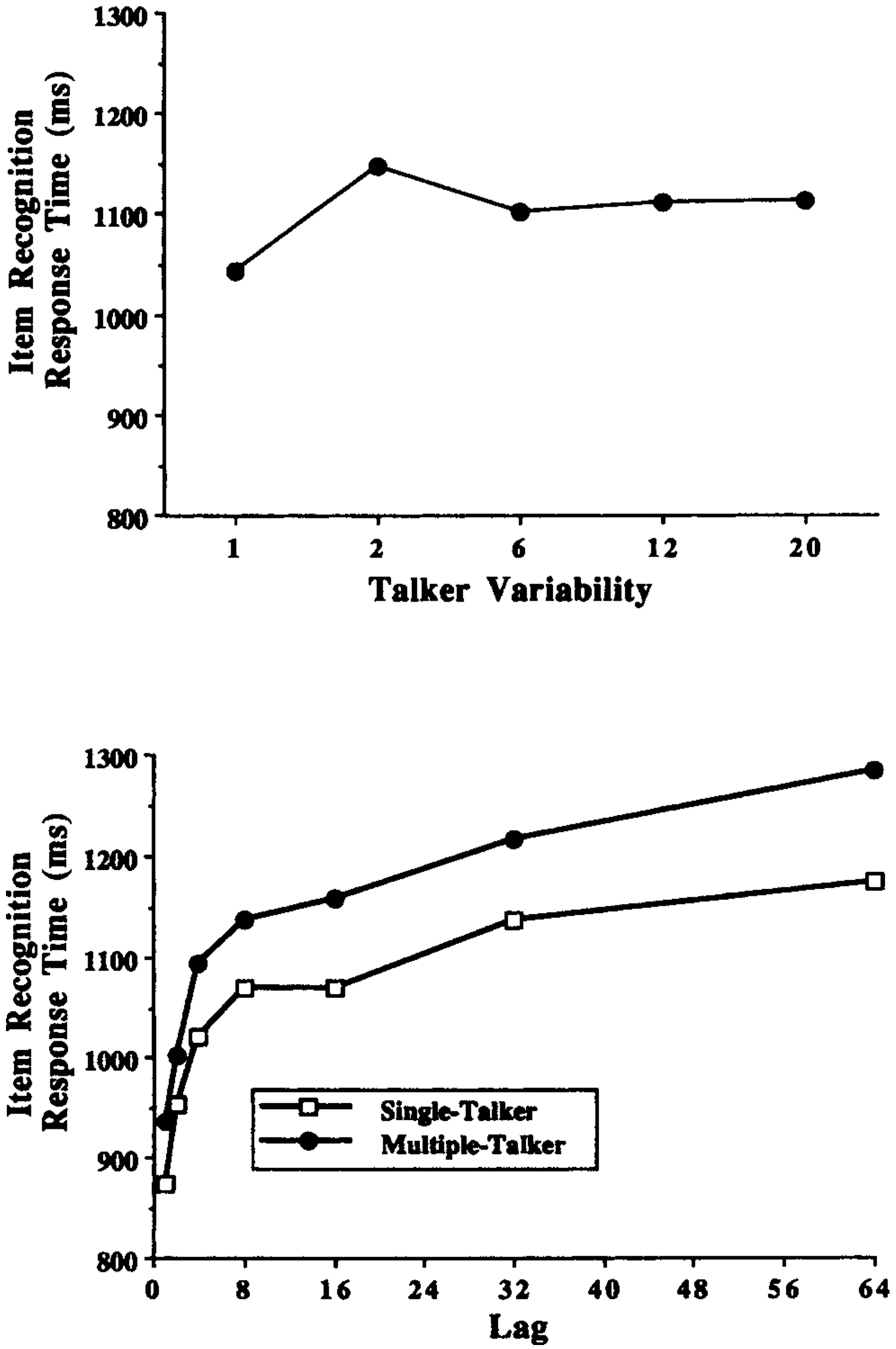

Figure 4.

Response times for correctly recognizing old items from the single-talker condition and the same-voice repetitions of multiple-talker conditions in Experiment 1. (The upper panel displays response times as a function of talker variability, collapsed across values of lag; the lower panel displays response times for the single-talker condition in comparison with the average response times for the same-voice repetitions of the multiple-talker conditions as a function of lag, collapsed across levels of talker variability.)

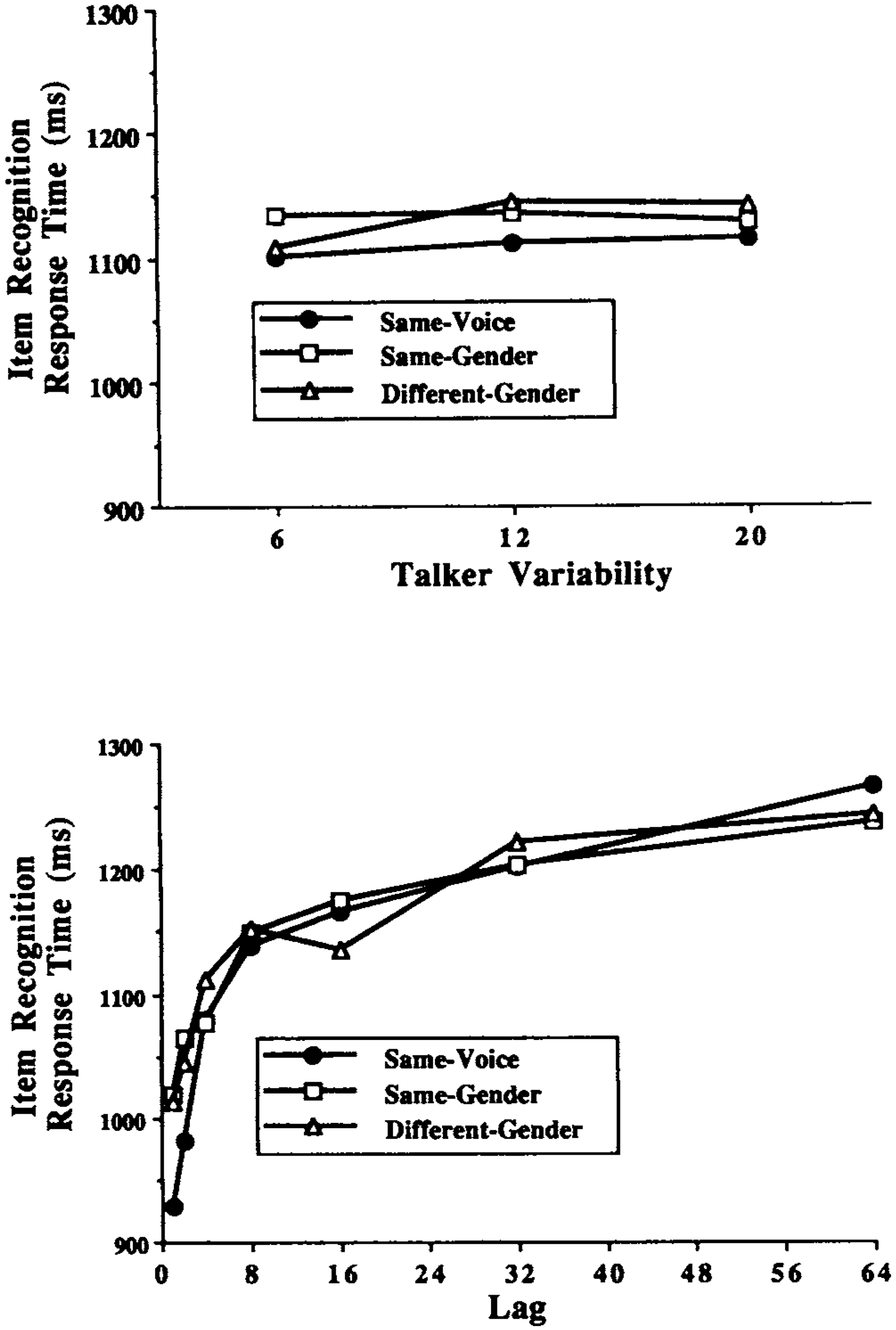

Figure 5.

Response times for correctly recognizing old items from a subset of the conditions in Experiment 1. (In both panels, response times for same-voice repetitions are compared with response times for different-voice/same-gender and different-voice/different-gender repetitions. The upper panel displays response times as a function of talker variability, collapsed across values of lag; the lower panel displays response times as a function of lag, collapsed across levels of talker variability.)

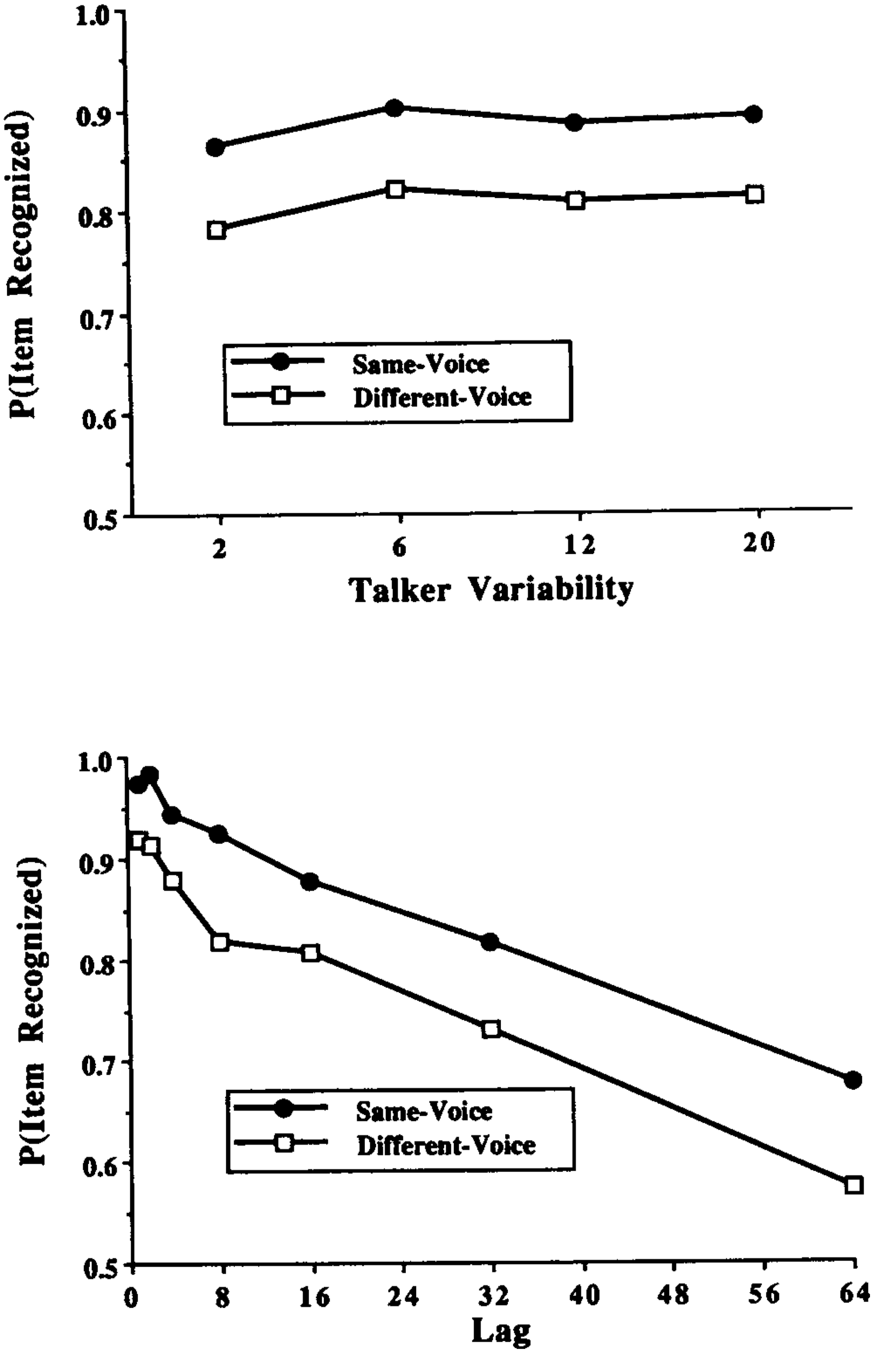

Figure 6.

Probability of recognizing old items from all conditions in Experiment 2. (The upper panel displays item recognition for same-voice repetitions and different-voice repetitions as a function of talker variability, collapsed across values of lag; the lower panel displays item recognition for same- and different-voice repetitions as a function of lag, collapsed across levels of talker variability.)

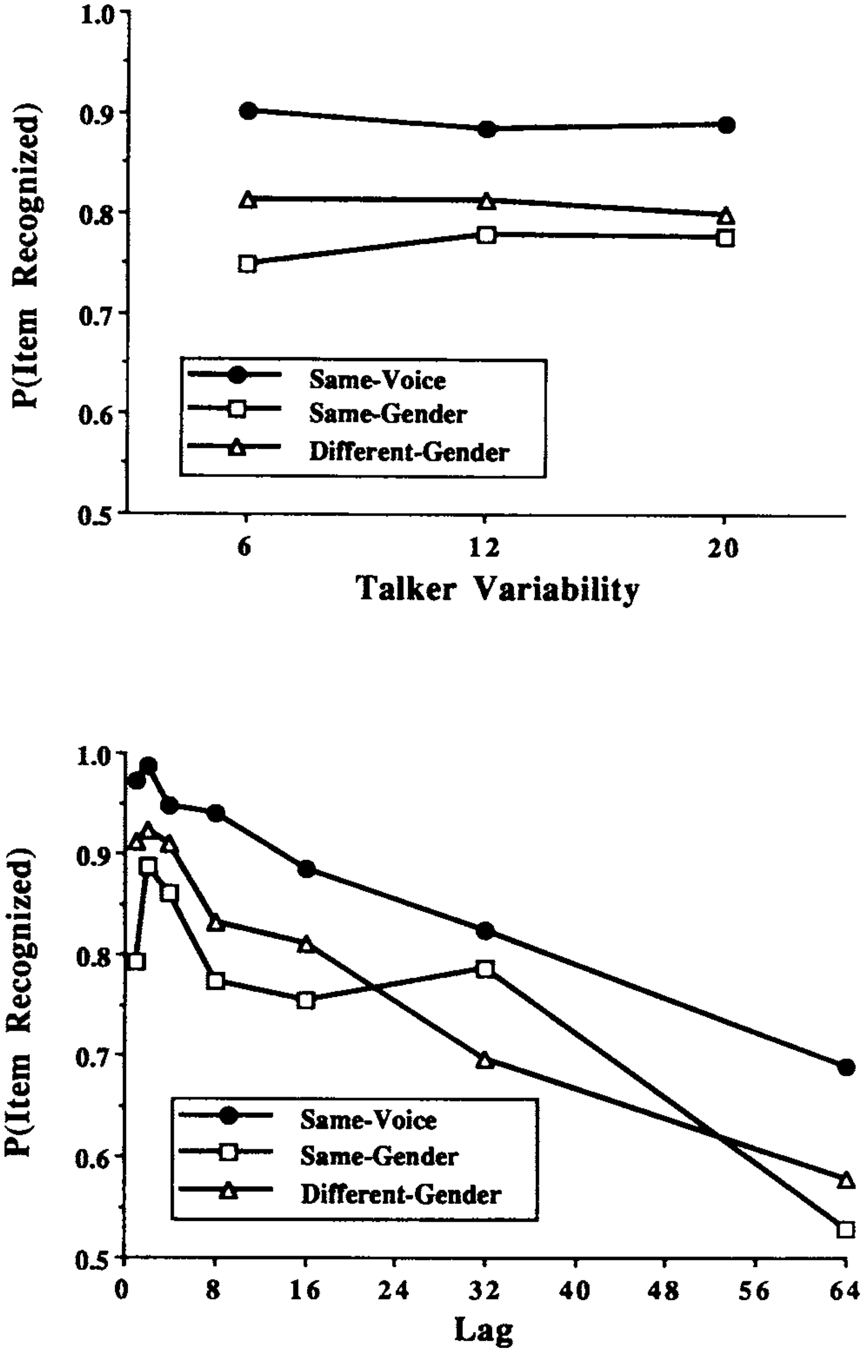

Figure 7.

Probability of correctly recognizing old items from a subset of the conditions in Experiment 2. (In both panels, item recognition for same-voice repetitions is compared with item recognition for different-voice/same-gender and different-voice/ different-gender repetitions. The upper panel displays item recognition as a function of talker variability, collapsed across values of lag; the lower panel displays item recognition as a function of lag, collapsed across levels of talker variability.)

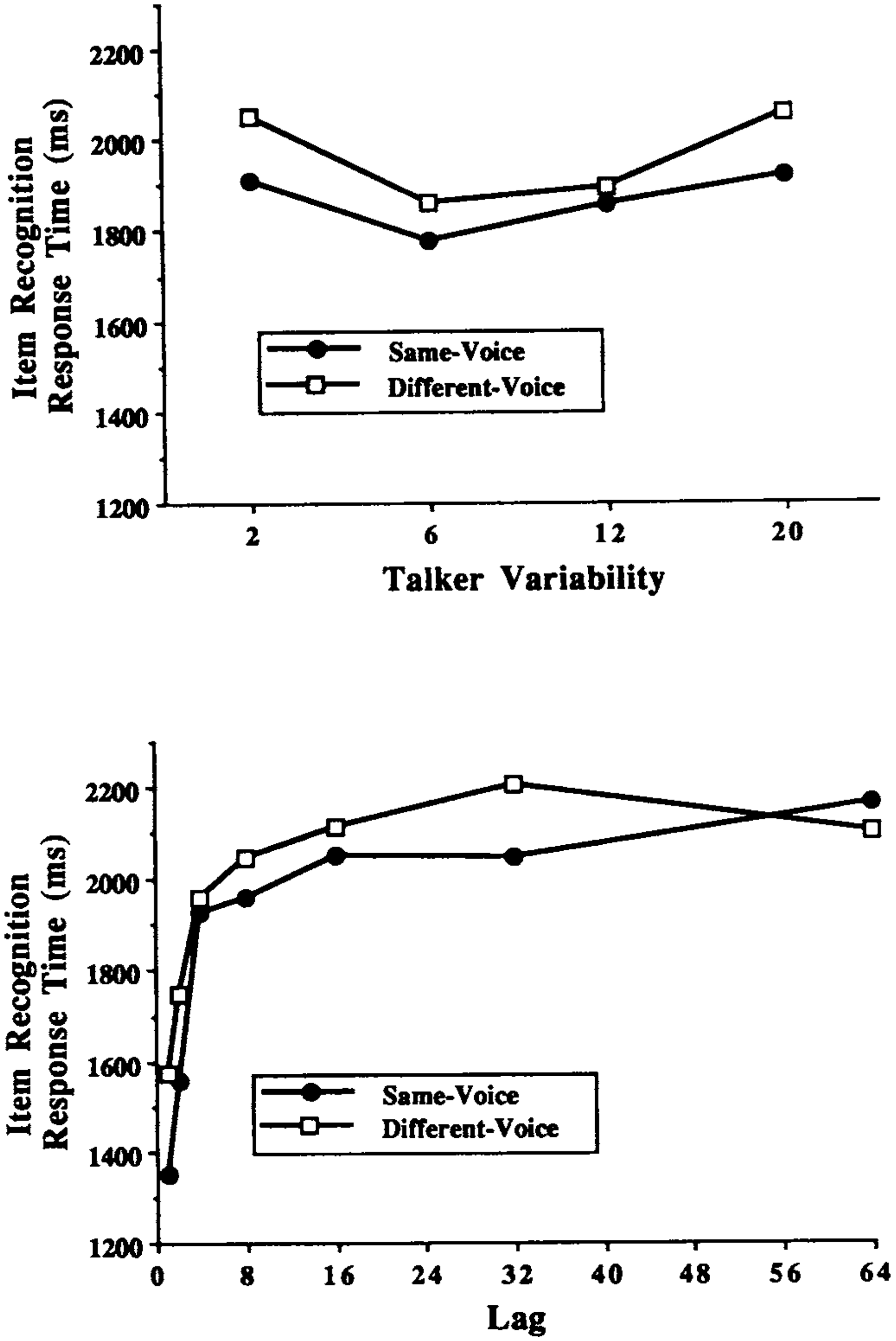

Figure 8.

Response times for correctly recognizing old items from all conditions in Experiment 2. (The upper panel displays response times for same-voice repetitions and different-voice repetitions as a function of talker variability, collapsed across values of lag; the lower panel displays response times for same- and different-voice repetitions as a function of lag, collapsed across levels of talker variability.)

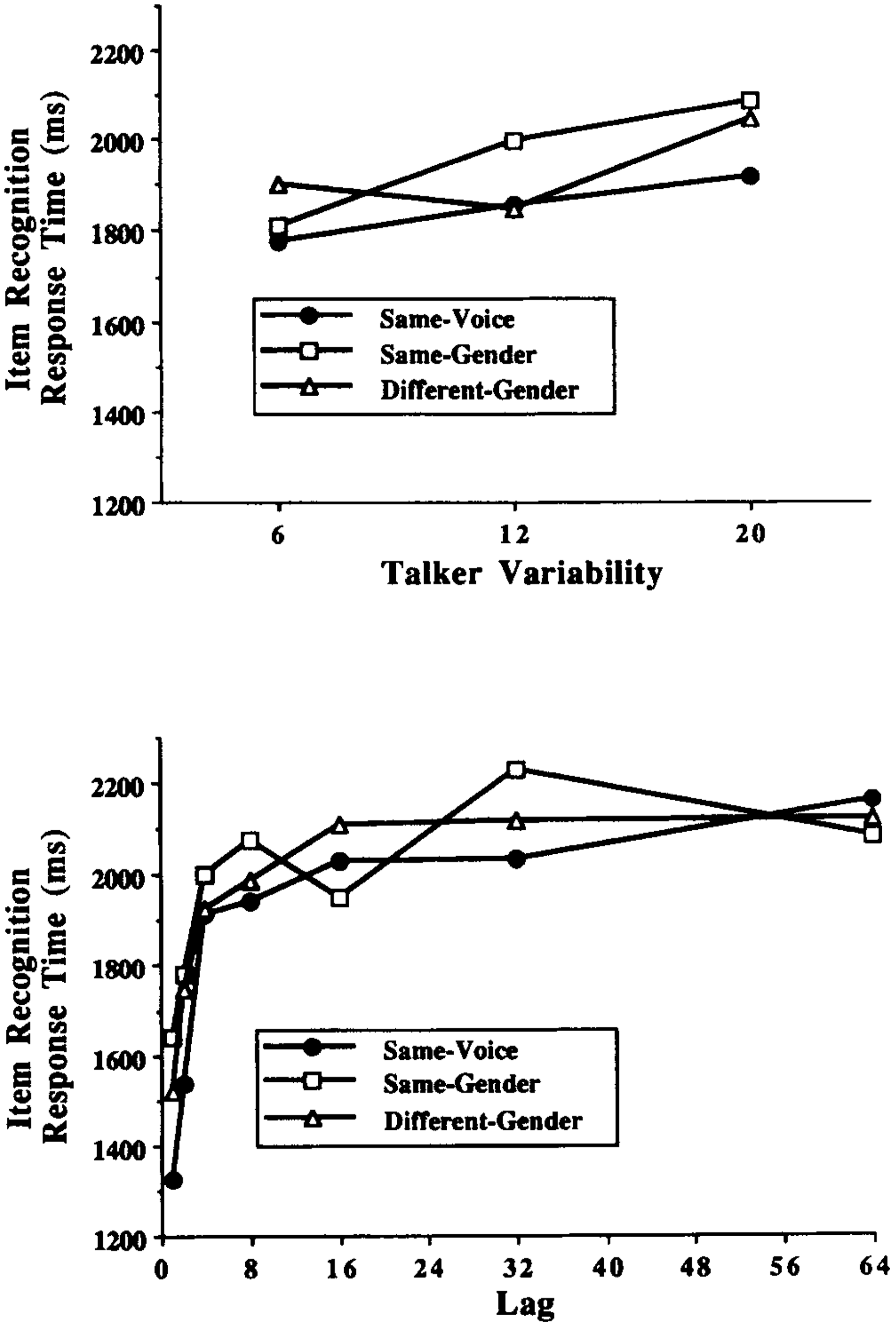

Figure 9.

Response times for correctly recognizing old items from a subset of the conditions in Experiment 2. (In both panels, response times for same-voice repetitions are compared with response times for different-voice/same-gender and different-voice/different-gender repetitions. The upper panel displays response times as a function of talker variability, collapsed across values of lag; the lower panel displays response times as a function of lag, collapsed across levels of talker variability.)

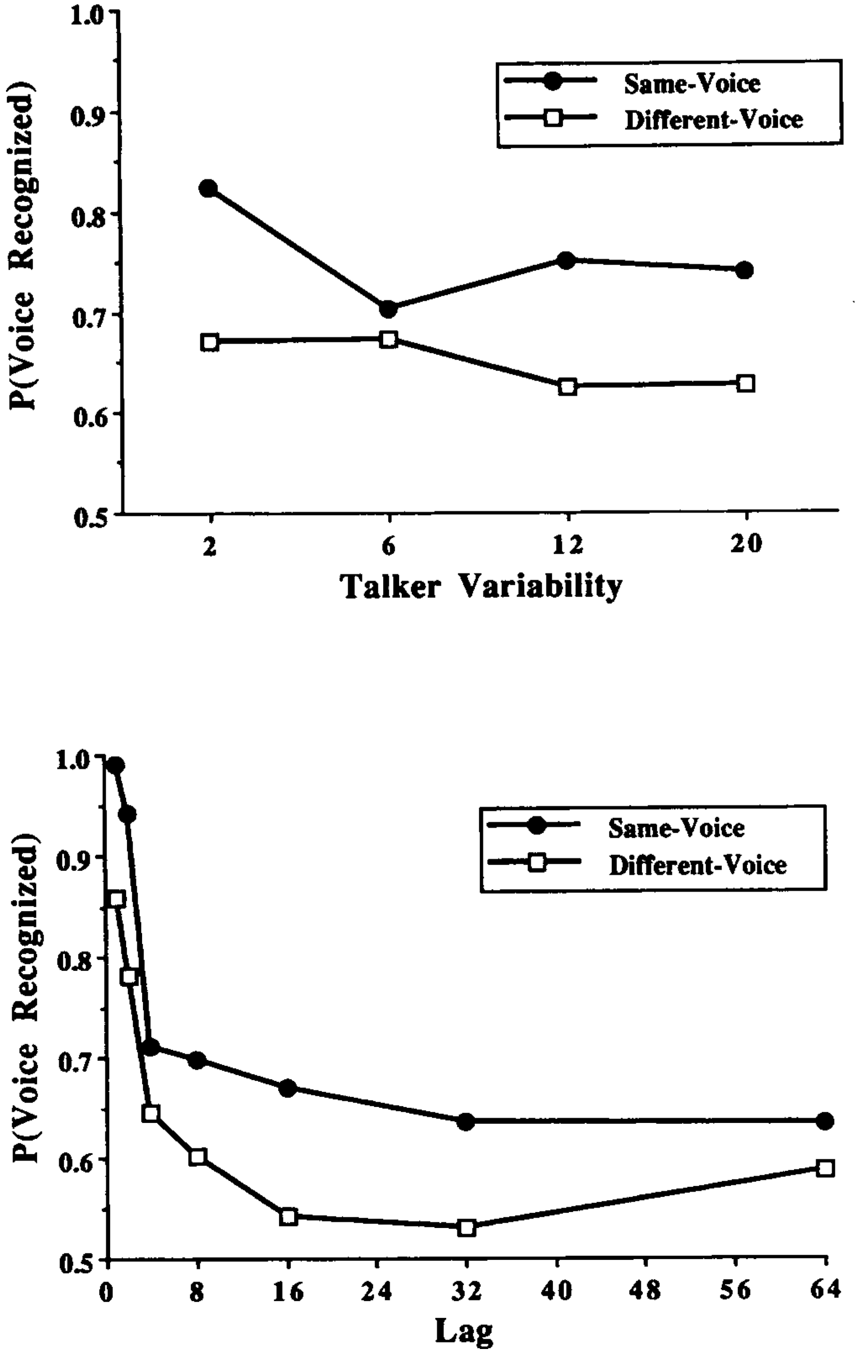

Figure 10.

Probability of correctly recognizing old items as a repetition in the same voice or as a repetition in a different voice from all conditions in Experiment 2. (The upper panel displays voice recognition for same- and different-voice repetitions as a function of talker variability, collapsed across values of lag; the lower panel displays voice recognition for same- and different-voice repetitions as a function of lag, collapsed across levels of talker variability.)

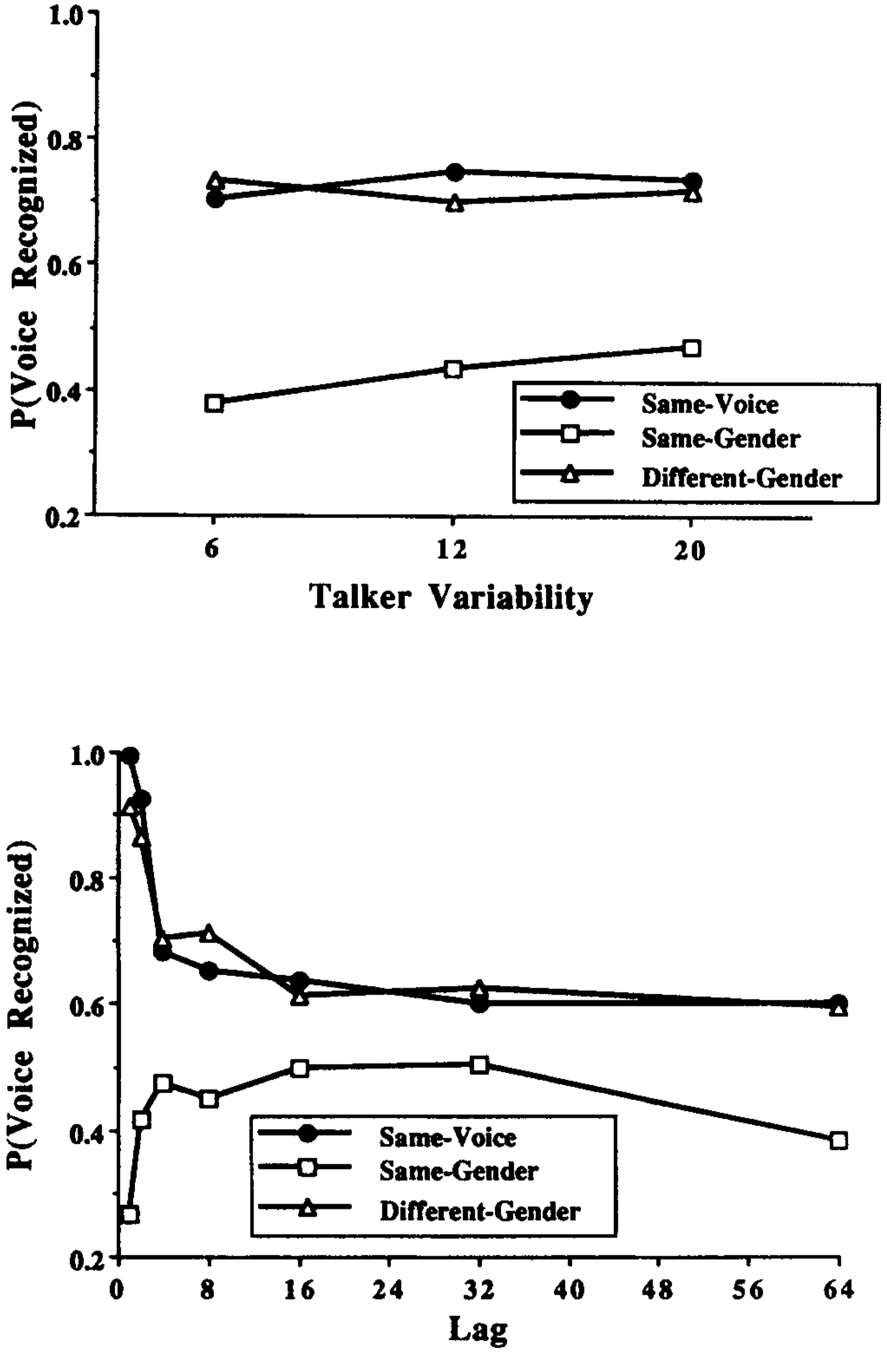

Figure 11.

Probability of correctly recognizing old items as a repetition in the same voice or as a repetition in a different voice from a subset of the conditions in Experiment 2. (In both panels, voice recognition for same-voice repetitions is compared with voice recognition for different-voice/same-gender and different-voice/different-gender repetitions. The upper panel displays voice recognition as a function of talker variability, collapsed across values of lag; the lower panel displays voice recognition as a function of lag, collapsed across levels of talker variability.)

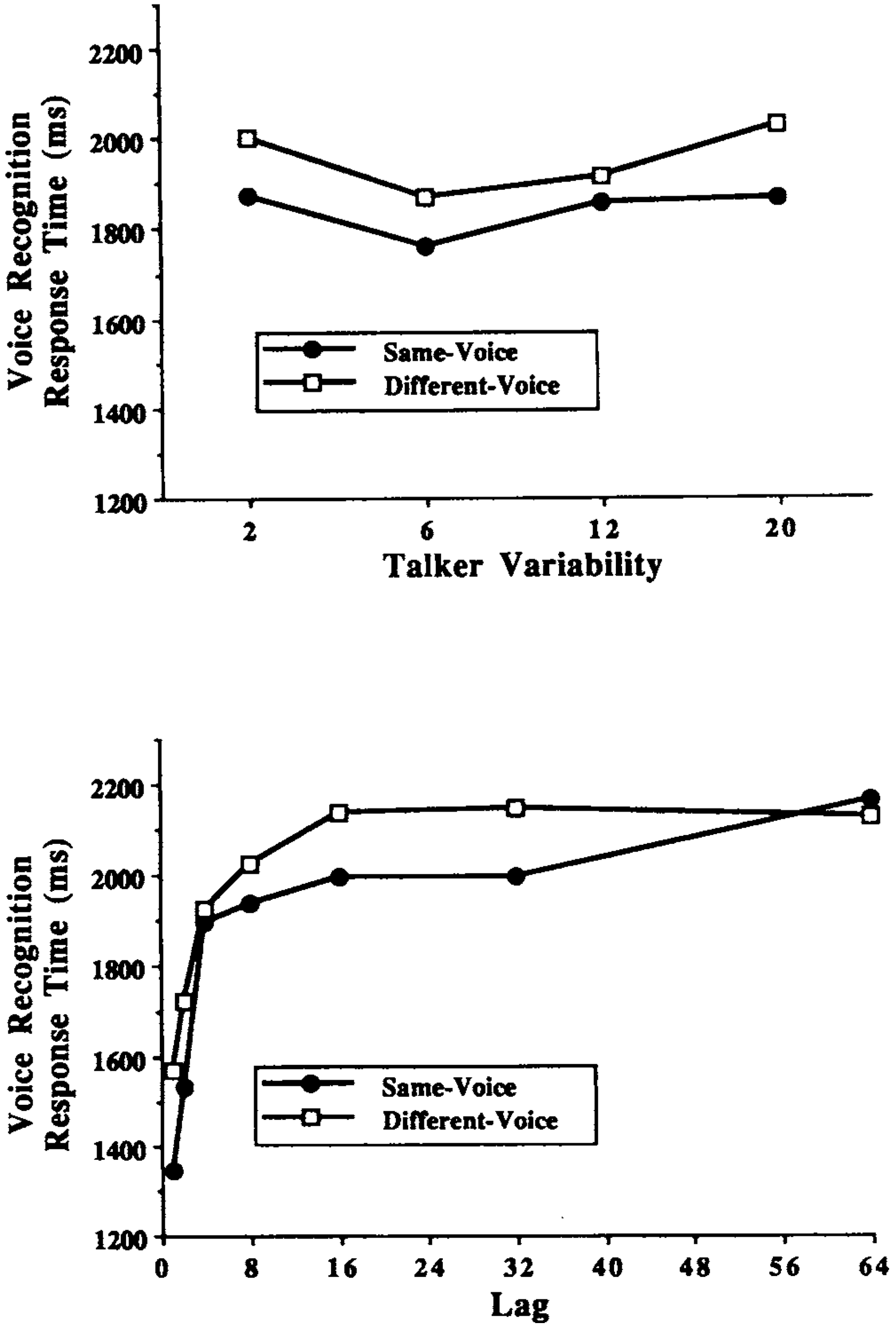

Figure 12.

Response times for correctly recognizing old items as a repetition in the same voice or as a repetition in a different voice from all conditions in Experiment 2. (The upper panel displays voice recognition response times for same- and different-voice repetitions as a function of talker variability, collapsed across values of lag; the lower panel displays response times for same- and different-voice repetitions as a function of lag, collapsed across levels of talker variability.)

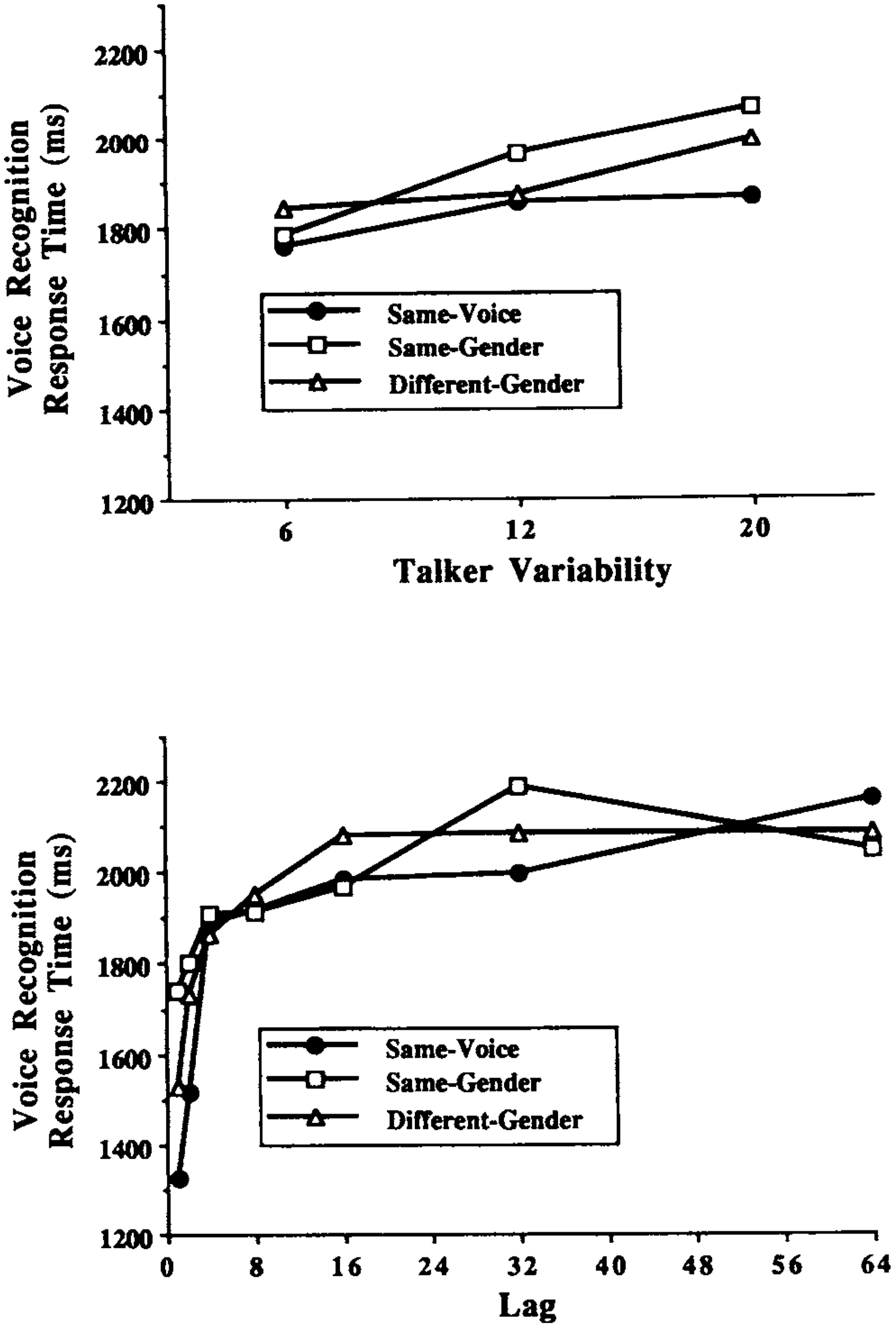

Figure 13.

Response times for correctly recognizing old items as a repetition in the same voice or as a repetition in a different voice from a subset of the conditions in Experiment 2. (In both panels, the response times for same-voice repetitions are compared with response times for different-voice/same-gender and different-voice/ different-gender repetitions. The upper panel displays voice recognition response times as a function of talker variability, collapsed across values of lag; the lower panel displays response times as a function of lag, collapsed across levels of talker variability.)

Table 1

False Alarm Rates and False Alarm Response Times in Experiment 1

Talker variability	Rate	Response time (ms)
Single talker	.20	1,296
Two talkers	.19	1,394
Six talkers	.20	1,361
Twelve talkers	.23	1,362
Twenty talkers	.22	1,370

Table 2

False Alarm Rates and False Alarm Response Times in Experiment 2

Talker variability	Rate	Response time (ms)
Two talkers	.15	2,404
Six talkers	.17	2,205
Twelve talkers	.18	2,126
Twenty talkers	.18	2,370