Journal of Experimental Psychology:
Learning, Memory, and Cognition
1996, Vol. 22, No. 5, 1166–1183

0278-7393/96/$3.00

Words and Voices: Episodic Traces in Spoken Word Identification and Recognition Memory

Stephen D. Goldinger
Arizona State University

Most theories of spoken word identification assume that variable speech signals are matched to canonical representations in memory. To achieve this, idiosyncratic voice details are first *normalized,* allowing direct comparison of the input to the lexicon. This investigation assessed both explicit and implicit memory for spoken words as a function of speakers' voices, delays between study and test, and levels of processing. In 2 experiments, voice attributes of spoken words were clearly retained in memory. Moreover, listeners were sensitive to fine-grained similarity between 1st and 2nd presentations of different-voice words, but only when words were initially encoded at relatively shallow levels of processing. The results suggest that episodic memory traces of spoken words retain the surface details typically considered as noise in perceptual systems.

In a now-classic article, Oldfield (1966) first described the *mental lexicon,* a collection of words in long-term memory that mediates perceptual access to lexical knowledge. The lexicon has since been a focus of extensive investigation and theorizing. Painting in broad strokes, there are two basic views on lexical representation: *Abstractionist theories* view the lexicon as a set of ideal, modality-free units, and *episodic theories* assume that groups of detailed traces collectively represent individual words (see Roediger & McDermott, 1993; Tenpenny, 1995). The abstractionist view is prominent in current theories; perception is typically assumed to involve information reduction, which is the decoding of specific episodes (tokens) into canonical representations (types; Morton, 1969; Posner, 1964). However, some theories posit episodic representations and perception, bypassing such decoding. Global memory models (Eich, 1982; Gillund & Shiffrin, 1984; Hintzman, 1986; Underwood, 1969), exemplar categorization models (Nosofsky, 1991), and distributed memory models (McClelland & Rumelhart, 1985) all assume episodic traces (although processing assumptions clearly differ).

This article is based on a doctoral dissertation submitted to Indiana University. Support was provided by National Institutes of Health Research Grant DC-00111-16 and National Institute on Deafness and Other Communication Disorders (NIDCD) Training Grant DC-00012-13 to Indiana University, and NIDCD First Award R29-DC02629-01 to Arizona State University.

I thank Rich Shiffrin, Linda Smith, and Bob Port for early suggestions. Special thanks to David Pisoni for 5 years of guidance and support, Lisa Burgin for tireless data collection, and Scott Lively for perusing the data with a keen and critical eye. Larry Jacoby, Carol Fowler, Patricia Tenpenny, Dominic Massaro, and Dan Schacter all provided feedback on earlier versions. Comments were also kindly provided by Peter Jusczyk, Guy Van Orden, Tamiko Azuma, and Robert Remez.

Correspondence concerning this article should be addressed to Stephen D. Goldinger, Department of Psychology, Arizona State University, Box 871104, Tempe, Arizona 85287-1104. Electronic mail may be sent via Internet to goldinger@asu.edu.

Memory for Details in Word Recognition

In abstractionist theories, idiosyncratic features of words are *normalized* with respect to canonical mental representations; the perceptual system filters surface details that are tangential to word identity (Brown & Carr, 1993; Carr, Brown, & Charalambous, 1989; Green, Kuhl, Meltzoff, & Stevens, 1991; Jackson & Morton, 1984). In strong form, the normalization assumption suggests that surface features of words will be absent from long-term memory. Given this hypothesis, both implicit and explicit memory for such surface details has been previously investigated. Whereas explicit memory entails conscious recollection, implicit memory entails facilitation of task performance, possibly without conscious recollection (Musen & Treisman, 1990; Schacter, 1987; Tulving & Schacter, 1990). A common measure of implicit memory is the *repetition effect,* which is improved perception of words with repeated presentations (Jacoby & Dallas, 1981; Jacoby & Hayman, 1987).[1]

With respect to memory for stimulus surface details, implicit and explicit memory data differ in several interesting respects: First, memory for surface details is more often revealed by implicit measures than by explicit measures (Schacter, 1987; Tenpenny, 1995). Second, the passage of time differentially affects each measure: Surface details rapidly fade from explicit memory, but persist in implicit memory (Cave & Squire, 1992; Musen & Treisman, 1990). Third, manipulating the level of stimulus processing (LOP) during study strongly affects explicit memory, but has little effect on implicit memory (Graf & Mandler, 1984; Jacoby & Dallas, 1981; but see Challis & Brodbeck, 1992). Because the goal of the present investigation was to assess memory for voice details of spoken words, convergent measures were examined: Implicit and explicit

[1] Common usage is to refer to repetition effects as priming. Following Jacoby and Brooks (1984; Jacoby, Marriott, & Collins, 1990), I prefer to avoid this term. Priming implies that abstract units, such as logogens (Morton, 1969), are temporarily activated by stimulus presentation, suggesting a theoretical predisposition toward abstractionism. Also, it is doubtful that repetition effects observed after a full day (see Experiment 2) constitute priming in the logogen sense.

memory tests were administered after short and long retention intervals (Experiment 2) and after words were studied at several LOPs (Experiment 3).

Memory for surface details of printed words has been well documented (see Goldinger, 1992; Tenpenny, 1995). For example, Hintzman, Block, and Inskeep (1972; also Kirsner, 1973) found that recognition memory was superior for words studied and later tested in a constant typography, relative to words that changed typography between study and test. In a similar implicit memory test, Roediger and Blaxton (1987) found larger repetition effects in word fragment completion for fragments studied and later tested in a constant font, relative to a changed font. This was observed in an immediate test, and after a 1-week delay (see also Jacoby & Hayman, 1987; Manso de Zuniga, Humphreys, & Evett, 1991). However, some studies reveal minimal effects of such typographic changes. For example, Carr et al. (1989) had volunteers read texts aloud twice. Between readings, half the texts were switched from typed to handwritten format, or vice-versa. When equivalent savings were observed despite format changes, Carr et al. suggested that repetition effects reflect priming of abstract word units that are insensitive to surface variability. Similarly, Scarborough, Cortese, and Scarborough (1977) found equivalent repetition effects when words presented in lexical decision were later repeated in old or new typefaces (see also Feustel, Shiffrin, & Salasoo, 1983). From these null findings, the role of specific episodes in memory and perception seems minimal. However, despite such data, many positive results underscore the importance of episodes in lexical representation; null findings may reflect particular tasks or stimuli, not lexical abstraction (Tenpenny, 1995).

In research on spoken word identification, most theories assume the speech signal is converted to a sequence of discrete segments, which is then compared to abstract lexical entries (McClelland & Elman, 1986; Studdert-Kennedy, 1976; see Pisoni, 1993). These assumptions are motivated by the extremely variable nature of the speech signal (see Goldinger, Pisoni, & Luce, 1996; Klatt, 1979). Indeed, a key issue in speech perception concerns *speaker normalization.* Speakers differ in vocal tracts (Joos, 1948; Peterson & Barney, 1952), glottal characteristics (Monsen & Engebretson, 1977), strategies for producing phonemes (Ladefoged, 1980), and native dialects. Yet, people understand most speakers without difficulty. Speaker normalization presumably allows listeners to follow the lexical–semantic content of speech; voice details used in early phonetic perception are discarded after lexical access (Jackson & Morton, 1984; Krulee, Tondo, & Wightman, 1983). Thus, long-term memory for spoken words should reflect elements of meaning; elements of perception, such as voice details, should be lost.

Unfortunately, the speaker normalization hypothesis appears unfalsifiable, at least by perceptual tests. Mullennix, Pisoni, and Martin (1989) found that speaker variability impairs identification of words in noise, and speeded shadowing of words in the clear (see also Nusbaum & Morin, 1992). They suggested that a mandatory, *capacity-demanding* normalization process usurps resources needed for primary task performance. However, null effects of speaker (and font) variability have led other researchers to suggest that *automatic* normalization supports fluent performance (Carr et al., 1989; Green et al., 1991; Jackson & Morton, 1984; Krulee et al., 1983). Apparently, both positive and null effects reflect normalization. Indeed, given the basic assumption that variable signals are matched to canonical representations, perception implies normalization by fiat. To reexamine normalization, an alternative to perceptual tasks is needed. The best possibility is memory, as investigated in the present study.

Despite the normalization hypothesis, voice information is clearly not tangential to communication. Voices convey "personal information" about speakers, such as age, gender, regional origin, and emotional state. Such information is extraneous to most views of speech perception, but is clearly used in communication (Ladefoged & Broadbent, 1957). Introspection suggests that voice memory is accurate (e.g., one can recognize friends on the telephone, etc.). More formally, although early research was discouraging (e.g., McGehee, 1937), later studies found reliable memory for voices (Carterette & Barnebey, 1975; Papçun, Kreiman, & Davis, 1989). Indeed, Van Lancker, Kreiman, and Emmorey (1985; Van Lancker, Kreiman, & Wickens, 1985) found that famous voices are well recognized, even when played backwards or at altered rates.

Examining memory for isolated spoken words (as in the present study), Martin, Mullennix, Pisoni, and Summers (1989) found impaired serial recall of 10-speaker lists, relative to 1-speaker lists. They suggested that 10-speaker lists engage a normalization process, usurping working memory resources needed for rehearsal. However, Goldinger, Pisoni, and Logan (1991) later found that speaker variability interacts with presentation rate; when words were presented relatively slowly, primacy recall from 10-speaker lists surpassed that from 1-speaker lists. As presentation rate affects rehearsal (Murdock, 1962; Rundus, 1971), these data suggest that attention is not used for normalization, but to help encode voice information into long-term memory, alongside lexical information. This is supported by other data on memory for voice attributes of spoken words (Cole, Coltheart, & Allard, 1974; Hintzman et al., 1972). For example, Craik and Kirsner (1974) investigated continuous recognition memory for words and voices. In this task, words are continuously presented, minimizing rehearsal (Shepard & Teghtsoonian, 1961). Listeners try to classify each word as new on its first presentation, and old on its repetition. In their study, words were first spoken by a male or female; when repeated, half the words switched voices. Same-voice repetitions were better recognized than different-voice repetitions, even after 32 intervening trials. Craik and Kirsner concluded that voice details persist in memory for at least 2–3 min. However, this conclusion was tenuous, as voice changes always entailed gender changes. Thus, the same-voice advantage could reflect either analog episodes or abstract "gender codes" (Geiselman & Crawley, 1983).

To resolve this issue, Palmeri, Goldinger, and Pisoni (1993) replicated and extended Craik and Kirsner's (1974) study. To provide a more stringent test, five levels of speaker variability were used: Listeners heard 1, 2, 6, 12, or 20 voices, with equal numbers of male and female speakers in all multiple-speaker conditions. Given this procedure, the automaticity of voice encoding could be assessed: If listeners strategically encode voices, increasing from 2 to 20 speakers should impair their ability to do so. The gender code hypothesis, which predicts that recognition should be gender dependent, not voice dependent, could also be tested. Thus, same-voice and different-voice/same-gender repetitions should yield equivalent recognition; both should exceed different-voice/different-gender repetitions. Instead, observed same-voice hit rates were higher than different-voice hit rates, regardless of gender or number of speakers, at all lags up to 64 intervening trials. Apparently, detailed word-plus-voice episodes are automatically encoded in perception; repetition of exact tokens facilitates later recognition.

Recent research by Schacter and his colleagues is particularly germane to the present study, as several common issues were investigated. Schacter and Church (1992) examined implicit and explicit memory for spoken words as a function of LOP at study, and whether voices matched across study and test. In two experiments, participants performed a shallow or deep processing task during study, followed by either perceptual identification of words in noise or recognition memory of words in the clear. Despite numeric trends, no reliable voice effects were observed. Hypothesizing that white noise was responsible, Schacter and Church conducted further experiments using auditory fragment completion and cued recall. Listeners processed bisyllabic words at shallow or deep LOPs during study; during test they heard old and new word-initial syllables. Old fragments were spoken in either old or new voices. Depending on condition, listeners completed fragments with the first words to come to mind, or used the fragments as cues to recall study words. Repetition effects in fragment completion were stronger for same-voice repetitions, regardless of LOP. No voice effect was observed in cued recall (see also Schacter, Church, & Osowiecki, 1994; Schacter, Church, & Treadwell, 1994).

Recently, Church and Schacter (1994) extended their earlier research: In their first experiment, listeners heard words produced by several speakers during study; during test they either identified low-pass filtered words (thus degraded without white noise), or performed recognition memory to clear words. For old words; half of the voices changed between study and test. As before, no voice effect emerged in explicit memory, but a reliable voice effect was observed in word identification. In further implicit memory experiments (using both word identification and fragment completion), Church and Schacter examined study–test changes of word intonation and fundamental frequency, both within voices. Changes of either intonation or fundamental frequency affected implicit memory, but had little effect on explicit memory.

The present investigation extended prior studies in several respects: Schacter and Church (1992) found that voice details persist in memory for 4–5 min. However, a robust voice effect at maximum lag in the study by Palmeri et al. (1993) suggests that episodic traces could last far longer. In the present study, episodic retention was assessed over longer time periods (Experiment 2), and across LOPs (Experiment 3). As in Schacter and Church (1992), memory was assessed in both explicit and implicit tests, but more voices were included in the design. Following Palmeri et al.'s (1993) methodology, this afforded a test of the automaticity of voice encoding, and it removed the confound of voice and gender changes. Also, the perceptual similarities among all stimulus voices were discovered by means of multidimensional scaling (Experiment 1; see Carterette & Barnebey, 1975). If traces of spoken words retain fine-grained voice details, as suggested by Church and Schacter (1994), identification of old words in new voices should be affected by the perceptual similarity of the voices, even within genders. Taken together, the present experiments further examined the specificity and durability of episodic memory traces and the role of attention in determining their content.

Experiment 1

In most research examining memory for font or voice details, test stimuli are denoted same format or different format, with little concern for entailed magnitudes of perceptual differences (although interesting studies by Brown & Carr, 1993, and Papçun et al., 1989, used typicality ratings to estimate memorability of fonts and voices). Given the theoretic importance of such perceptual differences, Experiment 1 assessed the similarity relations among all stimulus voices. Same–different response times were compiled into a matrix for multidimensional scaling, providing a perceptual space for all voices (Podgorny & Garner, 1979; Sergent & Takane, 1987; Weiner & Singh, 1974). This was used as an analytic tool in later experiments.

Method

Participants

One hundred eighty-three Indiana University students participated for partial course credit. All were native English speakers with no history of speech or hearing disorders.

Stimulus Materials

The stimuli were 300 monosyllabic English words, all recorded by 10 different speakers. Most words (272) came from the Modified Rhyme Test (House, Williams, Hecker, & Kryter, 1965), although 28 were selected from phonetically balanced word lists (Egan, 1948). All words were originally recorded on audiotape in a sound-attenuated booth with an Ampex AG500 tape deck and an Electro-Voice D054 microphone. The words were low-pass filtered at 4.8 kHz, digitized at a 10-kHz sampling rate with a 12-bit analog to digital converter, and were separated into digital files with a waveform editor. After the root-mean-square (RMS) amplitudes of all words were equated, 10 volunteers identified all words in a pilot test; tokens that did not yield 90% correct identification were replaced by more intelligible tokens.

Procedure

Participants were tested in groups of 6 or fewer in a sound-attenuated room. A PDP 11/34 computer controlled all experimental procedures. Each trial began with a 500-ms illumination of a cuelight on a 2-button response box. After the cuelight, a 500-ms silent interval elapsed and two words (presented on TDH-39 headphones at 75 dB, sound pressure level [SPL]) were spoken in succession, separated by 100 ms. Participants indicated, as quickly and accurately as possible, whether the same word was spoken twice. Every trial presented two different speakers; participants were advised to ignore speaker differences and to respond only to word identity. For all participants, the right- and left-hand buttons corresponded to same and different, respectively. After each pair of words, the computer waited up to 5 s for all responses; latencies were recorded from the onset of the second word in each pair.

The experiment consisted of 225 same and 225 different trials, randomly intermixed. The critical data came from the same trials, which comprised all interspeaker comparisons required for a full similarity matrix. With 10 speakers, 45 pairwise combinations constituted the matrix. Every participant received each of these 45 voice pairings 5 times in the same trials. Across participants, all 225 words were equally represented in every cell of the matrix, thus ensuring that differences across cells were only due to speaker differences. Because this control could never be maintained in the different trials, only same trials were used in scaling. Participants rested for 4–5 min halfway through the procedure.

Results and Discussion

Mean latencies of correct same responses were calculated to the nearest millisecond for all 45 cells in the design. The resultant similarity matrix is shown in the Appendix. As 183 students participated, each matrix value represents approximately 915 observations. Incorrect responses (1.35% of all trials) were not included. In their guidelines for setting the dimensionality of scaling solutions, Kruskal and Wish (1978) suggest that the number of dimensions multiplied by four should not exceed the number of objects scaled. As the stimulus array included 10 voices (objects), only a two-dimensional solution was derived. The KYST (Kruskal, Young, Sherry, and Torgeson) nonmetric scaling program was run several times with different starting configurations and random seeds, always producing relatively low stress values and consistent intervoice distances.

The two-dimensional solution is shown in Figure 1. The circles F1–F5 denote female speakers; M1–M5 denote male speakers. Although their exact nature was not relevant to the present investigation, the emergent dimensions reflected two stimulus properties: The horizontal axis clearly corresponds to gender; males and females fall to either side of the midpoint. Within sets of male and female speakers, increases along the vertical axis correspond to increases in vocal pitch. Because this could not reflect absolute pitch (all males had lower voices than females), it is denoted *relative pitch.* After the scaling solution was derived, object coordinates in the space were used to estimate perceptual distances between all speakers.[2] In Experiments 2 and 3, effects of perceptual similarity between study and test voices were assessed with these estimated distances.

Experiment 2

Experiment 2 examined implicit and explicit memory for spoken words as a function of voice, speaker variability (number of speakers), and delay between study and test. In some regards, Experiment 2 resembled Schacter and Church's (1992) experiments on perceptual identification and recognition memory. However, Experiment 2 included several manipulations to build upon Schacter and Church's results. First, whereas their study and test sessions were always separated by 4–5 minutes, Experiment 2 included delays of 5 min, 1 day, and 1 week. Thus, the *longevity* of voice-specific memory was assessed. Second, Experiment 2 included three levels of speaker variability (2, 6, or 10 voices). Thus, the *automaticity* of voice encoding was assessed (as in Palmeri et al., 1993). Third, whereas Schacter and Church confounded voice changes with gender changes (as in Craik & Kirsner, 1974), the different-voice trials of Experiment 2 included both same- and different-gender tokens (although only in the 6- and 10-voice conditions). Indeed, by removing this confound and applying the scaling data, the role of intervoice similarity could be assessed.

Experiment 2 also reexamined the effect of white noise on the expression of voice-specific memory. Schacter and Church (1992; also Jackson & Morton, 1984) showed that voice effects in perceptual identification are unreliable when clear study words are later identified in noise. Indeed, they hypothesized that voice-processing centers of the right cerebral hemisphere are selectively impaired by white noise (see Zaidel, 1978, and the General Discussion). It is possible, however, that potential voice effects were attenuated by the study–test format change (i.e., noise was used only at test). To increase the likelihood of observing voice effects, each condition of Experiment 2 maintained stimulus constancy across study and test sessions. Words were presented in the clear (at both study and test) for recognition memory, but were presented in noise (at both study and test) for perceptual identification. Noise is typically assumed to momentarily increase perceptual difficulty, and to then be quickly forgotten. However, if perception creates detailed memory traces, the mask may be encoded alongside other details. If so, repeating exact signal-plus-noise events at study and test may increase voice effects in perceptual identification (Hintzman, 1986; Jacoby, 1983).

Given the stimulus constancy constraint, the implicit and explicit memory conditions of Experiment 2 were not directly comparable (e.g., Musen & Treisman, 1990; Tulving, Schacter, & Stark, 1982), as both study and test sessions differed across conditions. This confound was considered necessary to observe voice effects. However, *confound* is defined arbitrarily in this regard. Schacter and Church's perceptual identification experiments were equally confounded in the opposite direction; noise was used only in implicit memory test sessions. Nevertheless, it is important to note that implicit–explicit memory comparisons are not ideally supported by Experiment 2. Indeed, this difficulty is exacerbated by another difference across conditions. Specifically, test sessions in recognition memory presented equal numbers of old and new words, but test sessions in perceptual identification presented all old words. This was somewhat unfortunate, as overall repetition effects could not be assessed, relative to new words. However, this comparison was not a central goal of this study; voice effects were of primary interest. Clearly, without inclusion of new words in perceptual identification, overall improvement at test could reflect general practice effects. However, because all words and voices were equally represented in the design (and were thus equally practiced), same- and different-voice trials were directly comparable. As noted earlier, the implicit–explicit memory comparison was not the current focus; the

---

[2] Estimates were derived with the Euclidean geometric equation for distance between two points in a plane:

$$\text{Distance} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2},$$

in which (x_1, y_1) and (x_2, y_2) are planar coordinates for Points 1 and 2, respectively.

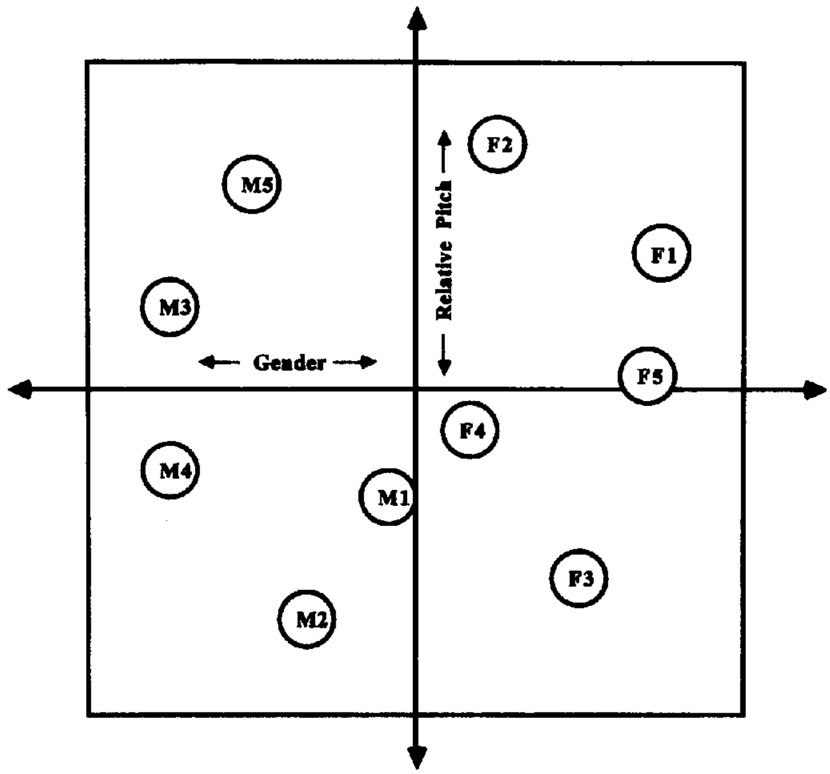

Figure 1. Experiment 1: Two-dimensional perceptual space derived from similarity matrix. Male and female speakers are denoted by symbols M1–M5 and F1–F5, respectively.

methods were used only to provide convergent data on voice-specific memory.

In summary, the implicit memory condition of Experiment 2 examined perceptual identification of words in noise. Listeners identified words in study and test sessions separated by 5 min, 1 day, or 1 week, with 2, 6, or 10 voices in the stimulus set. The key manipulation was voice: Each study word was later repeated in either its original voice or a new voice. Repetition effects (changes in identification accuracy across sessions) were examined. In the recognition memory conditions, listeners identified words in the clear during study, then received a surprise recognition memory test. The predictions followed prior research: Following Schacter and Church (1992) and Palmeri et al. (1993), voice effects were expected in both implicit and explicit memory (respectively). However, voice was expected to affect implicit memory over longer delays than explicit memory (Roediger & Blaxton, 1987). Following Palmeri et al. (1993), voices were expected to be encoded automatically, so voice effects were not expected to change across levels of speaker variability. Finally, the perceptual similarity of study and test voices was expected to affect the magnitude of different-voice repetition effects.

Method

Participants

Three hundred sixty Indiana University students participated; 226 received course credit and 134 received $10.00 each. All were native English speakers with no history of speech or hearing disorders. Half participated in perceptual identification (implicit memory), half in explicit recognition memory. Within these major conditions, participants were divided into subconditions, with 20 students each.

Stimulus Materials

The stimuli consisted of the words previously described. Referring to the perceptual space in Figure 1, the speakers in the 2-voice conditions were F1 and M1; the speakers in the 6-voice conditions were F1, F2, F3, M1, M2, and M3. With only 20 participants per subcondition, complete counterbalancing of words and voices was not possible. Thus, for all groups, quasirandom stimulus lists were prepared (following Palmeri et al., 1993). Each group received a different study list, with words randomly, but equally, assigned to voices and to old versus new trials (for recognition memory). Test lists for all groups were generated from their study lists, following one constraint: Approximately half of the old words were presented in their old voices; the remaining old words were presented equally often in all other voices. (In the 6- and 10-voice recognition memory conditions, 72 and 70 (respectively) of the 150 old words were presented in old voices.)

Procedure

Recognition memory. Participants were tested in groups of 5 or fewer, using the apparatus described earlier. In study sessions, listeners identified 150 randomly ordered words presented in the clear: Each trial began with a warning phrase (*get ready for next trial*) on the screen for 750 ms. Five hundred milliseconds after the warning, a word was presented at 75 dB (SPL) and listeners had up to 20 s to type it using a keyboard. In test sessions, participants typed *O* or *N* (old or new) in response to each of 300 randomly ordered words. The recognition memory test was a surprise; participants were not advised to memorize words during study. However, they were informed during test sessions that old words might be spoken in new voices.

Perceptual identification. Groups were tested in the same room with the same apparatus. In both study and test sessions, listeners identified 300 randomly ordered words in white noise. Each trial began with the warning phrase on the screen for 750 ms. Five hundred milliseconds later, continuous, band-limited white noise was presented at 70 dB (SPL). Fifty milliseconds after noise onset, a word was presented at 75 dB (SPL). Listeners had up to 20 s to type the word using a keyboard.

Results

Recognition Memory

Overall performance. Mean percentages of hits and false alarms were calculated for each participant and were used to estimate d' and β. Figure 2 displays overall accuracy (in terms of d') as a function of speaker variability and delay (no systematic β trends were observed). These data were analyzed in a 3 × 3 (Number of Voices × Delay) between-subjects analysis of variance (ANOVA). The main effect of number of voices was not significant, $F(2, 171) = 0.91$, $MSE = 1.10$, n.s., reflecting nearly identical performance across levels of speaker variability. (Throughout this article, all reported results are reliable at the $p < .05$ level or beyond, except for specifically denoted null results.) A significant main effect of delay, $F(2, 171) = 44.78$, $MSE = 1.10$, reflected the decreased accuracy at longer delays (this d' effect was due to both decreased hits and increased false alarms over time). The Number of Voices × Delay interaction was not significant, $F(2, 171) = 0.04$, $MSE = 1.10$, n.s.

Voice effects. Mean hit rates to same- and different-voice repetitions were calculated for all participants. Because false alarms cannot be analyzed in terms of repetition voice (they are responses to new words), only hit rates were analyzed. Table 1 shows hit rates as a function of voice, number of voices, and delay. Also shown are voice effects derived by subtraction.

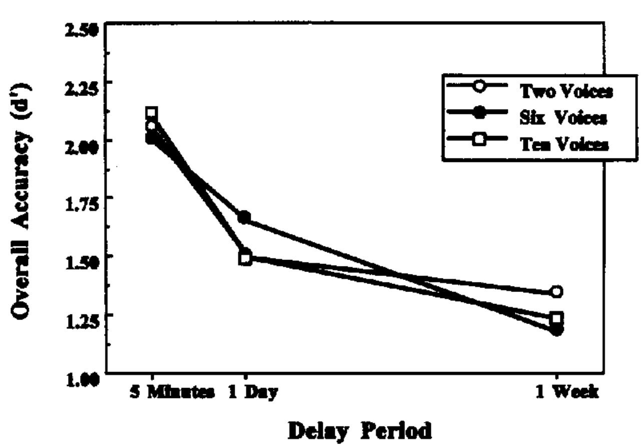

Figure 2. Overall recognition memory as a function of delay and number of voices in Experiment 2.

A 2 × 3 × 3 (Voice × Number of Voices × Delay) mixed-design ANOVA was conducted on the mean hit rates. A significant main effect of voice was observed, $F(1, 171) = 18.59$, $MSE = 1.05$; same-voice repetitions produced generally higher hit rates. Post hoc Tukey's honestly significant difference (HSD) analyses confirmed reliable voice effects in all 5-min delay conditions, the 2-voice–1-day condition, and the 10-voice–1-day condition. No voice effects were observed at the 1-week delay. A significant main effect of delay, $F(2, 171) = 21.31$, $MSE = 1.05$, reflected lower hit rates at longer delays, and a Voice × Delay interaction, $F(2, 171) = 11.61$, $MSE = 1.05$, reflected the decreased voice effect at longer delays. The main effect of number of voices was not reliable, $F(2, 171) = 1.65$, n.s., nor did it interact with other factors.

To ensure that voice effects were not completely due to cross-gender voice changes, hit rates to different-voice–same-gender (DV–SG) and different-voice–different-gender (DV–DG) repetitions were compared. (A finer-grained analysis of perceptual similarity effects is also presented below.) Table 2 shows these hit rates from the 6- and 10-voice conditions, along with implied gender effects. At delays that produced reliable voice effects, there were clear gender effects. At the 5-min delay, the 3.8% gender effect in the 6-voice condition was reliable, $F(1, 19) = 6.20$, $MSE = 5.13$, as was the 3.4% effect in the 10-voice condition, $F(1, 19) = 5.01$, $MSE = 4.99$. These gender effects show that overall voice effects were largely determined by gender changes across study and test tokens. However, in both conditions, DV–SG repetitions also produced reliable voice effects. The 4.5% effect in the 6-voice condition, $F(1, 19) = 9.19$, $MSE = 7.08$, and the 6.3% effect in the 10-voice condition, $F(1, 19) = 11.64$, $MSE = 8.20$, both show that voice effects are observable, even when voice changes are within gender. At the 1-day delay, gender effects were not reliable in either the 6- or 10-voice conditions, nor was the DV–SG voice effect reliable in the 6-voice condition. However, the 4.6% DV–SG voice effect in the 10-voice condition was reliable, $F(1, 19) = 4.51$, $MSE = 7.88$.

Perceptual Identification

Overall performance and voice effects. The percentage of words correctly identified in each session was calculated for each participant. Correct responses either matched target words exactly, or were homophones (e.g., *pear, pair*). Responses were corrected for simple spelling or typing errors, such as letter transpositions, prior to analysis. Table 3 displays word identification rates from both study and test sessions as a function of voice, number of voices, and delay. Also shown are repetition effects and voice effects, derived by subtraction.[3]

A 2 × 3 × 3 (Voice × Number of Voices × Delay) ANOVA was conducted on mean repetition effects for same- and different-voice trials.[4] A significant main effect of voice, $F(1,$

[3] As a result of a previous error in data analysis, these results differ slightly from those reported by Goldinger (1992).

[4] In most research on repetition effects (e.g., Church & Schacter, 1994), it is safely assumed that words are accurately perceived in study sessions. However, with words presented in noise during study, this may not be true. Another set of analyses on *conditionalized* identification

Table 1
Percentage of Hits in Recognition Memory as a Function of Voice, Number of Voices, and Delay in Experiment 2

	5 min			1 day			1 week		
Voices	Same	Different	VE	Same	Different	VE	Same	Different	VE
2 voices	83.5	75.9	7.6	71.7	66.0	5.7	68.4	67.1	1.3
6 voices	82.4	76.0	6.4	74.2	71.6	2.8	66.0	66.9	−0.9
10 voices	83.1	75.1	8.0	72.6	67.3	5.3	67.8	66.9	0.9
M	83.0	77.0	7.3	72.8	68.3	4.5	67.4	67.0	0.4

Note. VE = voice effect (same − different). False-alarm rates were .11 for 5 min., .15 for 1 day, and .19 for 1 week.

171) = 58.12, MSE = 1.99, reflected stronger repetition effects in same-voice trials (10.45%) than in different-voice trials (2.91%), although the small different-voice repetition effect was also reliable, $F(1, 171) = 7.31$, $MSE = 2.41$. Tukey's HSD analyses indicated that, in all but 3 cases (6-voice–1-day, 10-voice–1-day, 6-voice–1-week), same- and different-voice repetition effects reliably differed. No main effect of number of voices was observed, $F(2, 171) = 2.08$, n.s., but a Voice × Number of Voices interaction, $F(2, 171) = 5.05$, $MSE = 1.99$, reflected a small decrease in voice effects with increased speaker variability. Voice effects in the 2-, 6-, and 10-voice conditions were 9.30%, 7.50%, and 5.83%, respectively. No main effect of delay was observed, $F(2, 171) = 1.09$, n.s., but a Voice × Delay interaction, $F(2, 171) = 12.80$, $MSE = 1.99$, reflected larger voice effects at shorter delay intervals. Voice effects at the 5-min, 1-day, and 1-week delays were 9.17%, 7.43%, and 6.03%, respectively.

As in recognition memory, data from different-voice trials were next analyzed in terms of speaker gender (see Table 4). On average, DV–SG trials produced larger repetition effects (4.42%) than DV–DG trials (1.33%), $F(2, 111) = 24.16$, $MSE = 2.72$, as reflected by consistently positive gender effects in Table 4. Indeed, DV–DG trials did not yield reliable repetition effects (although see below). This pattern was consistent across number of voices and delays, producing null interactions with each variable (both Fs < 1.75).

Perceptual Similarity Effects

Recognition memory. After examination of categorical voice effects, the data were analyzed in terms of finer-grained perceptual similarity between study and test voices. Data from different-voice trials were divided into categories corresponding to pairwise combinations of speakers. The 6-voice conditions included 15 unique combinations of two speakers; the 10-voice conditions included 45 such combinations. In both 6- and 10-voice conditions, all pairwise combinations were equally represented across trials. Across participants, mean recognition hit rates were correlated with intervoice distances. Table 5 summarizes the results.

The negative omnibus correlations (upper portion of Table 5) show that hit rates decreased as intervoice distances increased. In both the 6- and 10-voice conditions, reliable omnibus correlations were observed in the 5-min and 1-day conditions, but not in the 1-week condition. To compare these correlations, slopes for all participants were compared in separate 1-way (delay) ANOVAs, revealing significant main effects of delay, 6-voice: $F(2, 59) = 121.21$, $MSE = 0.11$; 10-voice: $F(2, 59) = 24.01$, $MSE = 0.88$, confirming that delays mediated the perceptual similarity effect. To complement the correlations, simple binary analyses were also conducted. For each participant, it was determined whether repetition effects were generally larger for "closer" voices to each target voice. (For each voice in 6-voice conditions, 2 voices were deemed "close" and 3 "far." In 10-voice conditions, these values were 4 and 5 voices, respectively.) Participants whose repetition effects followed such perceptual similarity more than half the time were denoted +1; otherwise they were denoted 0. Sign tests were conducted on these scores, shown in the second tier of Table 5, showing patterns identical to the correlations.

In the 10-voice conditions, within-gender correlations (M–M and F–F) were attenuated, relative to the omnibus correlations, and were generally insignificant. However they were again negative, and became less so over delays, F–F: $F(2, 59) = 6.51$, $MSE = 0.30$; M–M: $F(2, 59) = 5.92$, $MSE = 0.34$. With only 3 speakers of each gender in 6-voice conditions, within-gender correlations were not feasible. As a substitute, binary analyses were again conducted. For each participant, it was determined whether mean within-gender repetition effects were larger for the more or less similar voice, relative to the study voice. Proportions of positives were analyzed by means of sign tests. As shown in Table 5, hit rates were affected by within-gender perceptual similarity after 5 min and somewhat after 1 day, but not after 1 week. Table 5 also shows within-gender sign tests for 10-voice data, generally verifying the relationships shown by correlations.

Perceptual identification. Table 6 shows perceptual similarity results for perceptual identification. Unlike recognition memory, correlations in perceptual identification were quite steady across delays. In the 6-voice conditions, omnibus correlations were reliable at all delays. In 10-voice conditions,

rates examined test data only for words that were correctly identified during study. Results closely matched the unconditionalized identification data: A voice effect was observed across levels of speaker variability and delay, with no Voice × Delay interaction. Note, however, that conditionalizing could exaggerate the voice effect, as the same-voice trials may overrepresent particularly easy study tokens in test sessions. (Note, however, that Goldinger, 1992, found similar results in a two-voice condition in which all test words were new tokens.)

Table 2
Percentage of Hits in Recognition Memory as a Function of Voice, Number of Voices, Delay, and Repetition Voice Gender in Experiment 2

	5 min				1 day				1 week			
Voices	Same	DV–SG	DV–DG	GE	Same	DV–SG	DV–DG	GE	Same	DV–SG	DV–DG	GE
6 voices	82.4	77.9	74.1		74.2	72.4	70.8		66.0	66.6	67.2	
VE		4.5	8.3	3.8		1.8	3.4	1.6		−0.6	−1.2	−0.7
10 voices	83.1	76.8	73.4		72.6	68.0	66.1		67.8	66.9	66.9	
VE		6.3	9.7	3.4		4.6	6.5	1.9		0.9	0.9	0.0
M	82.75	77.35	73.75	3.60	73.40	70.20	68.45	1.75	66.90	66.75	67.05	−0.35

Note. DV–SG = different-voice–same-gender; DV–DG = different-voice–different-gender; GE = gender effect (DV–DG minus DV–SG); VE = voice effect (same minus DV–DG or DV–SG).

omnibus correlations were reliable at the 5-min and 1-week delays, but not at the 1-day delay. Slopes for all participants were again tested in separate 1-way (delay) ANOVAs; neither main effect of delay was reliable, 6-voice: $F(2, 59) = 1.03$, $MSE = 1.14$, n.s.; 10-voice: $F(2, 59) = 0.76$, $MSE = 1.59$, n.s. As shown in Table 6, omnibus sign tests confirmed the patterns shown by the correlations.

In the 10-voice conditions, within-gender correlations generally resembled the omnibus correlations (albeit in attenuated degrees), being consistently negative across delays. For the 6- and 10-voice conditions, within-gender sign tests were again conducted. As shown in Table 6, although some proportions were unreliable, a general pattern of hit rates following perceptual similarity was observed across delays.

Discussion

The major results of Experiment 2 are easily summarized: In explicit recognition memory, a same-voice advantage was observed to equivalent degrees across levels of speaker variability, but not over delays: The same-voice advantage was 7.5% after 5 min, 4.1% after 1 day, and an unreliable 1.6% after 1 week. At shorter delays, however, listeners' sensitivity to voice was impressive. Beyond a basic same-voice advantage, recognition memory was affected by fine-grained similarity relations among voices. After a full day, with 10 voices in the stimulus set, listeners were sensitive to the magnitude of voice differences between study and test tokens of nominally identical words.

In perceptual identification, voice effects were observed at all delays up to a week. This was expected, given other findings in implicit memory (Cave & Squire, 1992; Jacoby & Dallas, 1981; Musen & Treisman, 1990). The perceptual similarity effects were also relatively stable over a week, in contrast to recognition memory. Whether this asymmetry implies functionally separate implicit and explicit memory systems (Tulving & Schacter, 1990) is debatable, as the present tests were not equated in all aspects (see Neely, 1989). Nevertheless, the convergent data show that voice details persist in recognition memory for at least a day, and are detectable in perceptual identification for at least a week.

Experiment 3

Delays between study and test affect explicit memory more than implicit memory, as seen in Experiment 2 (also Cave & Squire, 1992; Musen & Treisman, 1990; Tulving et al., 1982). In Experiment 3, delays were held constant, but the LOP

Table 3
Correct Word Identifications (in Percentages) as a Function of Voice, Number of Voices, and Delay in Experiment 2

	5 min			1 day			1 week		
Voices	Same	Different	VE	Same	Different	VE	Same	Different	VE
2 voices									
Study	58.1	62.9		57.9	62.2		56.2	57.7	
Test	71.6	68.0		69.0	60.5		68.5	63.3	
RE	13.5	5.1	8.4	11.1	−1.7	12.8	12.3	5.6	6.7
6 voices									
Study	54.2	48.7		49.1	46.9		50.3	48.8	
Test	66.0	47.5		57.1	50.2		58.2	51.9	
RE	11.8	−1.2	13.0	8.0	3.3	4.7	7.9	3.1	4.8
10 voices									
Study	44.6	43.1		41.3	41.0		44.7	47.2	
Test	54.4	46.8		52.6	47.5		53.1	49.0	
RE	9.8	3.7	6.1	11.3	6.5	4.8	8.4	1.8	6.6
M	11.7	2.5	9.2	10.1	2.7	7.4	9.5	3.5	6.0

Note. VE = voice effect (same − different); RE = repetition effect (test − study).

Table 4
Correct Word Identifications (in Percentages) as a Function of Voice, Number of Voices, Delay, and Repetition Voice Gender in Experiment 2

	5 min			1 day			1 week		
Voices	DV–SG	DV–DG	GE	DV–SG	DV–DG	GE	DV–SG	DV–DG	GE
6 voices									
Study	50.1	47.3		46.8	47.0		47.9	49.7	
Test	51.6	43.4		51.5	48.9		52.8	51.0	
RE	1.6	−3.9	5.5	4.7	1.9	2.8	4.9	1.3	3.6
10 voices									
Study	43.7	42.5		41.3	40.7		47.8	46.6	
Test	48.5	45.1		48.8	46.2		50.8	47.2	
RE	4.8	2.6	2.2	7.5	5.5	2.0	3.0	0.6	2.4
M	3.2	−0.7	3.9	6.1	3.7	2.4	4.0	1.0	3.0

Note. DV–SG = different-voice–same-gender; DV–DG = different-voice–different-gender; GE = gender effect (DV–DG minus DV–SG); RE = repetition effect (test − study).

during study was manipulated (Craik & Tulving, 1975). Performance was compared across LOPs to assess whether voice effects are purely stimulus driven, or if they are partially determined by listeners' focus of attention during study (Whittlesea & Brooks, 1988; Whittlesea & Cantwell, 1987). Despite some null results (e.g., Schacter & Church, 1992), previous research suggests that token-specific repetition effects vary across LOPs. For example, Graf and Ryan (1990) had volunteers rate words, printed in two unusual fonts, in terms of readability (shallow LOP) or pleasantness (deep LOP). Recognition memory was better for deep LOP words, but font effects were stronger for shallow LOP words. Challis and Brodbeck (1992) demonstrated clear LOP effects in implicit memory, and Masson and Freedman (1990) suggested that font effects in implicit memory are only observed when a shallow LOP is used (see Blaxton, 1989; Tenpenny, 1995; Weldon, 1991).

In Experiment 3, all listeners heard 6 voices in the stimulus set, and all study and test sessions were separated by 5 min. During study, listeners performed one of three speeded classification tasks to words presented in the clear. Words were classified according to speaker gender, initial phonemes, or syntactic classes. Classifying gender was expected to be the easiest (shallowest) task. Mullennix and Pisoni (1990) had listeners classify spoken words according to gender or initial phonemes in a Garner (1974) task. Voices and phonemes emerged as integral dimensions of spoken words, but the interference pattern was asymmetric—voice variations impaired phoneme classification more than the reverse. This suggests that voice information is available for perceptual classification before phonemic information is available. Logically, classifying words by syntax requires deeper processing than classifying by gender or phonology.

Table 5
Perceptual Similarity Effects in Recognition Memory in Experiment 2

Voices	5 min	1 day	1 week
Omnibus correlations			
6 voices	−.611**	−.354*	−.219
10 voices	−.530**	−.432*	−.146
Proportion of participants following overall perceptual similarity			
6 voices	20/20**	15/20*	8/20
10 voices	18/20**	16/20*	9/20
Within-gender correlations (10 voices)			
F–F	−.416*	−.330*	−.069
M–M	−.398*	−.306	−.112
Proportion of participants following within-gender perceptual similarity			
6 voices			
F–F	18/20**	14/20*	9/20
M–M	15/20*	12/20	12/20
10 voices			
F–F	17/20**	12/20	11/20
M–M	15/20*	10/20	7/20

Note. F–F = female–female; M–M = male–male.
*$p < .05$. **$p < .01$.

Table 6
Perceptual Similarity Effects in Perceptual Identification in Experiment 2

Voices	5 min	1 day	1 week
Omnibus correlations			
6 voices	−.361*	−.377*	−.349*
10 voices	−.312*	−.297	−.324*
Proportion of participants following overall perceptual similarity			
6 voices	15/20*	16/20*	13/20
10 voices	14/20*	14/20*	15/20*
Within-gender correlations (10 voices)			
F–F	−.305	−.318*	−.319*
M–M	−.280	−.305	−.339*
Proportion of participants following within-gender perceptual similarity			
6 voices			
F–F	16/20**	12/20	14/20*
M–M	15/20*	15/20*	13/20
10 voices			
F–F	16/20**	14/20*	14/20*
M–M	14/20*	13/20	17/20**

Note. F–F = female–female; M–M = male–male.
*$p < .05$. **$p < .01$.

During test, half of the participants received a recognition memory test, and half performed a semantic classification task in which two printed words (e.g., *stove, door*) preceded a spoken word (e.g., *hot*). On hearing the spoken word, participants quickly selected the better semantic associate. This conceptually driven test was examined, in lieu of perceptual identification, for two reasons: First, to maintain stimulus constancy across study and test, it was necessary to avoid noise. Second, prior research (Challis & Brodbeck, 1992; Jacoby, Levy, & Steinbach, 1992) suggests that LOP effects may be more easily observed in such procedures, relative to more data-driven tests (although see the General Discussion).

Method

Participants

In the recognition memory condition, 105 Indiana University students participated for course credit. In the semantic classification condition, 105 Arizona State University students participated for course credit. Within these major conditions, participants were divided into three subconditions determined by LOP, with 35 in each subcondition. All participants were native English speakers with no reported history of speech or hearing disorders.

Stimulus Materials

The stimulus materials were the words previously described, but only 6-voice lists were presented. As seen in Figure 1, the speakers were F1, F2, F3, M1, M2 and M3. As in Experiment 2, every group of participants received randomly generated study lists of words, followed by appropriately structured test lists.

Procedure

Recognition memory. Participants were tested in groups of six or fewer, using the apparatus previously described. The study session entailed 150 randomly ordered trials of speeded classification. Each trial began with *get ready* on the computer screen for 500 ms, followed by response labels in the lower left- and right-hand corners of the screen, spatially corresponding to response buttons. Correct responses were randomly mapped to each button an equal number of times. In gender classification, response labels were *male* and *female.* In initial phoneme classification, response labels were minimal pairs of words (e.g., *bat, rat*). (Participants were carefully instructed to respond to letter sounds, not actual letters.) In syntax classification, response labels were syntactic categories (e.g., *noun–adjective*), with only one correct choice (e.g., for *ride,* which is both a noun and verb, choices were *verb* and *adjective*). Five-hundred milliseconds after the onset of response labels, a stimulus word was presented at 75 dB (SPL). Listeners had up to 5 s to classify the word; once all listeners responded, a 750-ms interval elapsed and a new trial began. In test sessions, participants received a surprise recognition memory test with 150 study words and 150 new words. Procedures followed the study session, but with constant response labels (*old* and *new*). Instructions in both sessions stressed speed and accuracy.

Speeded classification. Participants were tested in groups of 4 or fewer; each sat in a carrel with a pair of Sennheiser headphones, a computer monitor, and a response box. A Gateway 2000 computer controlled the experiment. Study sessions consisted of 150 randomly ordered trials of speeded classification, with the procedures just described. Test sessions consisted of 300 randomly ordered trials of semantic classification, including the 150 study words and 150 new words. Procedures were identical to those in the study session, except the response alternatives were now two printed words (e.g., *stove, door*). Upon hearing a spoken word (e.g., *hot*), participants quickly selected the better semantic associate. The available options on each trial provided one clear associate, as determined by a pretest with 5 volunteers.

Results

Overall Performance

Recognition memory. To assess the efficacy of the LOP manipulation in study sessions, percentages and mean latencies of correct responses were calculated for each participant. For the recognition memory test, percentages of hits and false alarms were calculated for each participant, as were mean hit latencies. Table 7 presents the results of both sessions as a function of LOP. In study sessions, significant main effects of LOP were observed in latency, $F(2, 102) = 312.40$, $MSE = 116.80$, and accuracy, $F(2, 102) = 133.44$, $MSE = 8.86$. In the recognition memory test, significant main effects of LOP were again observed in accuracy, $F(2, 102) = 73.44$, $MSE = 11.91$, and latency, $F(2, 102) = 81.30$, $MSE = 97.01$. These results verify that the LOP manipulation was effective.

Semantic classification. For both study and test sessions, percentages and mean latencies of correct responses were calculated for each participant. Table 8 shows data from both sessions as a function of LOP. In the study session, significant main effects of LOP were observed in accuracy, $F(2, 102) = 139.85$, $MSE = 7.72$, and latency, $F(2, 102) = 332.75$, $MSE = 141.06$. Test session results were similar (examining only old words): Significant main effects of LOP were observed in accuracy, $F(2, 102) = 11.96$, $MSE = 15.80$, and latency, $F(2, 102) = 113.00$, $MSE = 91.71$. Old words were classified more accurately, $F(1, 102) = 14.82$, $MSE = 10.31$, and faster, $F(1, 102) = 18.55$, $MSE = 159.19$, than new words. Post hoc comparisons showed that these differences were primarily due to the deep LOP condition. Relative to new words, classification accuracy was reliably higher only for words in syntax classification; latencies were reliably faster for words in both phoneme and syntax classification.

Table 7
Overall Results of Study and Test Sessions as a Function of Level of Processing in the Recognition Memory Condition of Experiment 3

Session	Level of processing		
	Gender	Phoneme	Syntax
Study session (speeded classification)			
Percentage correct	97.5 (1.37)	94.0 (1.40)	83.8 (1.93)
Response latency (ms)	477 (11.16)	515 (8.79)	1,243 (36.68)
Test session (recognition memory)			
Accuracy (d')	0.78 (0.02)	1.13 (0.03)	1.99 (0.02)
Response latency (ms)	1,104 (30.59)	1,018 (23.50)	881 (13.52)

Note. Standard errors are shown in parentheses. False-alarm rates were .233 for gender classification, .220 for phoneme classification, and .085 for syntax classification.

Table 8
Overall Results of Study and Test Sessions as a Function of Level of Processing in the Semantic Classification Condition of Experiment 3

Session	Level of processing			
	Gender	Phoneme	Syntax	New words
Study session (speeded classification)				
Percentage correct	94.1 (0.56)	87.8 (1.15)	81.0 (2.23)	
Response latency (ms)	490 (8.62)	531 (9.47)	1,212 (28.74)	
Test session (semantic classification)				
Percentage correct	83.9 (1.50)	87.8 (1.01)	90.0 (1.03)	85.1 (1.56)
Response latency (ms)	1,369 (35.67)	1,121 (31.10)	920 (17.41)	1,349 (32.45)

Note. Standard errors are shown in parentheses.

Voice Effects

Recognition memory. Mean hit rates and response latencies for same- and different-voice repetitions were calculated for each participant. Because false alarms cannot be analyzed in terms of voice, only hit rates and latencies were analyzed. The upper half of Figure 3 shows recognition memory hits and latencies as a function of voice and LOP. Separate 2 × 3 (Voice × LOP) ANOVAs were conducted on the mean hit rates and latencies. In hit rates, a significant main effect of voice, $F(1, 102) = 85.11$, $MSE = 33.95$, reflected higher accuracy for same-voice repetitions. Tukey's HSD analyses confirmed reliable voice effects at all LOPs. However, the Voice × LOP interaction was also significant, $F(2, 102) = 10.07$, $MSE = 33.95$, reflecting the smaller voice effect at deeper LOPs. A significant voice effect was also observed in latency, $F(1, 102) = 13.60$, $MSE = 97.42$, reflecting faster recognition of same-voice repetitions. Tukey's HSD analyses confirmed reliable voice effects in the gender and phoneme conditions, but not in the syntax condition. As in accuracy, a significant Voice × LOP interaction, $F(2, 102) = 12.38$, $MSE = 97.42$, reflected the smaller voice effect at deeper LOPs.

Semantic classification. Mean correct classification rates and latencies from test sessions were calculated for each participant. The lower half of Figure 3 displays these data as a function of LOP and voice. In accuracy, a voice effect, $F(1, 102) = 4.20$, $MSE = 14.94$, reflected superior classification of same-voice repetitions. However, Tukey's HSD analyses showed that same- and different-voice classification accuracy reliably differed only at the shallow LOP. The Voice × LOP interaction only approached significance, $F(2, 102) = 2.52$, $MSE = 14.94$, n.s. In latency, a significant voice effect, $F(1, 102) = 62.55$, $MSE = 89.31$, reflected faster classification of same-voice repetitions. Although Tukey's HSD analyses confirmed reliable voice effects at all LOPs, the Voice × LOP interaction was also significant, $F(2, 102) = 29.58$, $MSE = 89.31$, reflecting the larger voice effects at deeper LOPs.

Perceptual Similarity Effects

Recognition memory. As before, the data were analyzed in terms of perceptual similarity between study and test voices. Data from different-voice trials were divided into 15 categories comprising all pairwise combinations of speakers; all combinations were equally represented across trials. The mean hit rates and latencies across categories were correlated with their corresponding perceptual distances, as shown in the upper portion of Table 9.

Correlations of recognition hit rates and perceptual distances were generally negative, indicating that hits decreased as perceptual distances between old and new voices increased. In response latencies, correlations were generally positive, indicating that latencies increased as perceptual distances increased. However, these correlations varied across LOPs. To examine this, slopes from all participants were compared in one-way (LOP) ANOVAs. Main effects of LOP were significant in both the hit rate, $F(2, 102) = 61.71$, $MSE = 11.21$, and latency data, $F(2, 102) = 49.02$, $MSE = 31.19$, verifying that perceptual similarity effects varied across LOPs. As shown in the lower portion of Table 9, the within-gender data (analyzed by means of sign tests, as in Experiment 2) showed similar qualitative patterns. In both accuracy and latency, sensitivity to within-gender voice similarity was only reliable in gender classification, diminishing at deeper LOPs.

Semantic classification. Table 9 also shows perceptual similarity results for the semantic classification condition. In latency, omnibus correlations were uniformly negative across LOPs; responses were slower when perceptual distances between study and test voices were greater. The correlations were again compared in a one-way (LOP) ANOVA. The main effect of LOP did not approach significance, $F(2, 102) = 1.71$, $MSE = 19.08$, n.s., showing that perceptual similarity effects were steady (albeit small) across LOPs. In the accuracy data, all correlations were small and insignificant. Unlike the recognition memory condition, listeners' sensitivity to within-gender voice similarity was generally unreliable.

Discussion

As expected, the manipulation of LOP during study strongly affected test performance, in both the semantic classification and recognition memory conditions. More important, reliable voice effects were observed in Experiment 3. However, voice and LOP effects were not independent: In recognition memory, although the voice effect was reliable (by hit rates) in the deepest processing condition, it was clearly stronger at shallower LOPs. Also, as in Experiment 2, the voice effect was not categorical: A generalization gradient was observed following

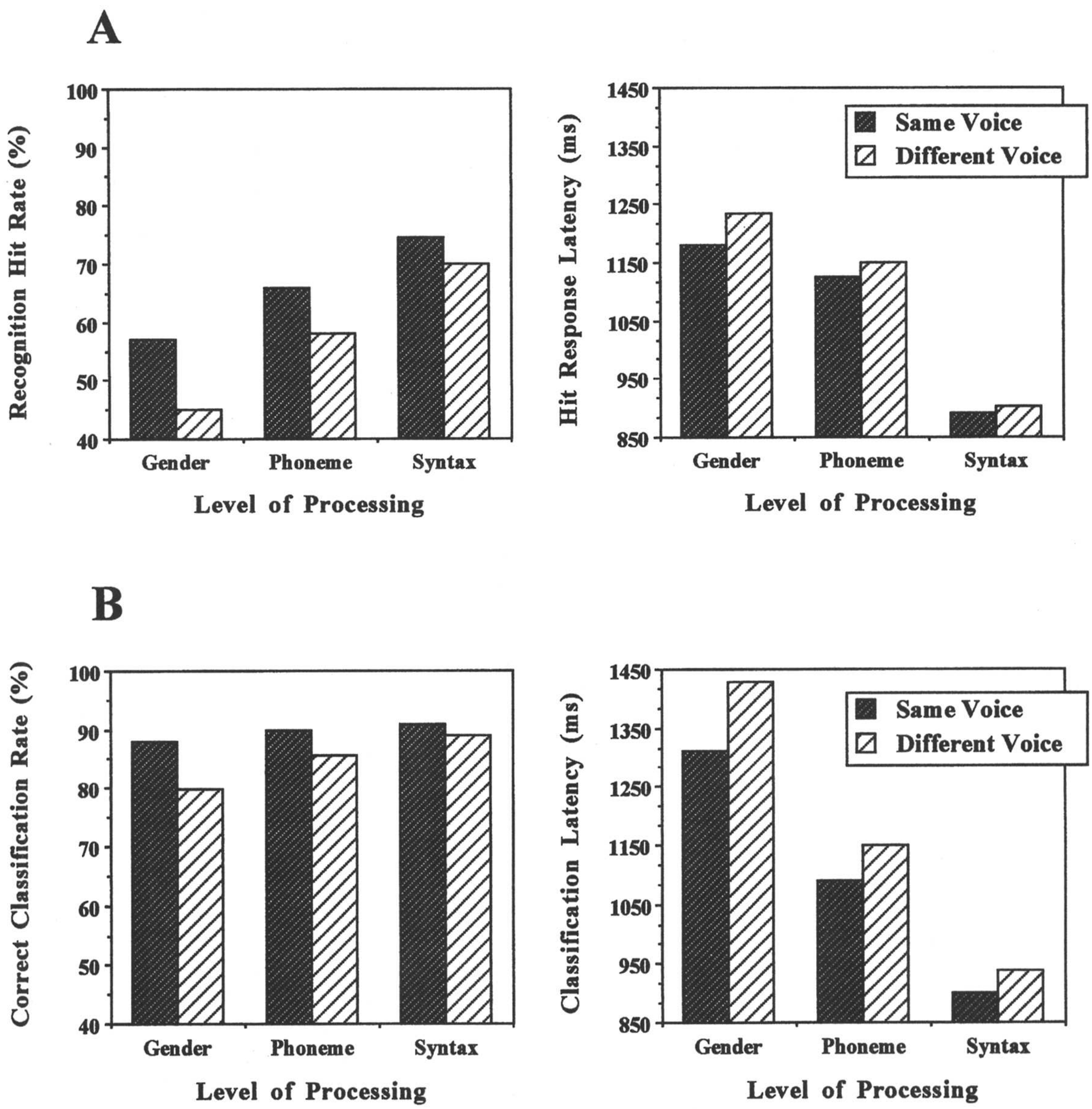

Figure 3. A: Recognition memory hit rates and response latencies as a function of level of processing and voice in Experiment 3. B: Semantic classification rates and response latencies for old words as a function of level of processing and voice in Experiment 3.

perceptual similarity, although not equally across LOPs. (However, given the excellent performance in the deep LOP condition, it is possible that the Voice × LOP interactions in accuracy and latency are partially due to ceiling and floor effects, respectively.)

Given the recognition memory data, it appears that the content of study words' episodic traces is influenced by the focus of attention at study. When attention was directed to more superficial details at study, stronger voice and perceptual similarity effects were observed. To the degree that recognition memory reflects perceptual processes, the voice effects in Experiments 2 and 3 suggest that perception may rely on memory for episodes (Jacoby & Brooks, 1984; Jacoby & Hayman, 1987). This suggestion is bolstered by the semantic classification data, in which voice effects remained reliable at deeper LOPs. But an interaction of voice and LOP was still observed, showing that focus of attention at study can affect token-specific performance on an implicit memory–perceptual test (Masson & Freedman, 1990).[5]

[5] In this regard, it is noteworthy that the semantic classification data of Experiment 3 could be "contaminated" by explicit memory. Specifically, on viewing the response alternatives for classification, participants may have recalled related study words, thus leading to improved performance. As this tendency would be most evident following deep processing, the LOP effect in implicit memory could be due to explicit

Table 9
Perceptual Similarity Effects in Experiment 3

Test statistic	Level of processing during study: Gender	Phoneme	Syntax
Omnibus correlations of different-voice performance and perceptual distance			
Recognition memory			
Hit rates	−.415**	−.330*	−.181
Response latencies	+.413**	+.310*	+.139
Semantic classification			
Correct classification rates	−.112	+.040	−.089
Response latencies	−.236*	−.201*	−.193
Proportion of participants following within-gender perceptual similarity			
Recognition memory			
Hit rates	32/35**	24/35*	18/35
Response latencies	28/35**	21/35	19/35
Semantic classification			
Correct classification rates	16/35	14/35	17/35
Response latencies	21/35	23/35*	16/35

*$p < .05$. **$p < .01$.

General Discussion

In both Experiments 2 and 3, same-voice repetitions led to generally better performance than different-voice repetitions of the same words, in contrast to a strong speaker normalization hypothesis. These results extend previous findings on surface memory for printed and spoken words (Church & Schacter, 1994; Cole et al., 1974; Craik & Kirsner, 1974; Goldinger et al., 1991; Hintzman et al., 1972; Jacoby & Hayman, 1987; Roediger & Srinivas, 1992; Schacter & Church, 1992). In particular, Experiment 2 extended our earlier continuous recognition memory study (Palmeri et al., 1993), which showed listeners' sensitivity to voice changes, even within gender, between words up to 5 min apart. Experiment 2 showed that detailed episodic traces of spoken words can influence recognition memory for a day, and perceptual identification for a week. Experiment 3, however, revealed a dependency between voice effects and the focus of attention at study. As discussed later, this result may help rationalize the present findings with respect to previous research, intuition, and theories of language processing.

The present investigation also assessed the role of fine-grain perceptual details in repetition effects for different-voice words. Similar prior investigations have typically used two voices, usually a male's and a female's. Extending this method, the present study revealed a monotonic relation of perceptual similarity and voice effects (due largely, but not entirely, to cross-gender voice changes). Jacoby and Brooks (1984; Johnston, Dark, & Jacoby, 1985) attributed such perceptual and memory advantages to the *perceptual fluency* of stimuli, which is partly determined by similarity of stimuli to prior episodes in memory.

recall (although see Challis & Brodbeck, 1992). For the present investigation, the more important data are the voice effects, which could still reflect perceptual fluency, despite possible involvement of explicit memory.

The implicit memory results (and those reported by Schacter & Church, 1992; Church & Schacter, 1994) are particularly relevant to the hypothesis of episodic perception. Recognition memory data of Experiment 2 suggested that voice details fade over time, but perceptual identification suggested they are more persistent. In a similar study, Masson (1984) found that recognition memory for details of transformed sentences diminished over a week, but implicit memory was unaffected by delay. Musen and Treisman (1990; also Cave & Squire, 1992) reported similar results for visual patterns. Such data complement the present results, confirming that episodic details can persist in memory for at least a week.[6]

The Episodic Lexicon?

By some theories of spoken word identification, the present data may be considered irrelevant, as identification and memorial encoding of words are considered separate modular processes (Fodor, 1983; Forster, 1979). However, the results show that episodic traces not only affect memory, but may also influence later perception (Church & Schacter, 1994; Roediger & Blaxton, 1987; Salasoo, Shiffrin, & Feustel, 1985). Parsimony may therefore suggest a lexicon consisting of episodic traces. Indeed, Jacoby has suggested nonanalytic word identification by means of direct comparison to episodes, rather than translation into abstract units (Jacoby, 1983; Jacoby & Brooks, 1984; Jacoby & Dallas, 1981; Jacoby & Hayman, 1987). Jacoby (also Tenpenny, 1995) argued that long-term repetition effects (as in Experiment 2) are consistent with memory for episodes, not priming of abstract units, such as logogens.

For most information-processing models (Liberman & Mattingly, 1985; Pisoni & Sawusch, 1975), episodic perception violates basic assumptions. Although some theories (e.g., Klatt, 1979; Paap, Newsome, McDonald, & Schvaneveldt, 1982) could be modified to predict episodic influences on word identification, several extant models already predict the key results. For example, Hintzman's (1986) MINERVA 2 assumes that every perceptual experience creates an independent, detailed memory trace. Despite their separate storage and idiosyncratic attributes, collections of traces activated at retrieval represent categories as a whole. Thus, MINERVA 2 accounts for both specificity and generality of memory with a single set of traces (for an interesting historical precursor, see Semon, 1923/1909; Schacter, Eich, & Tulving, 1978). Simulations of MINERVA 2 replicate results considered hallmarks of abstract representation (Hintzman, 1986; Hintzman & Ludlam, 1980), such as deriving long-lasting prototypes from random dot patterns (Posner & Keele, 1968, 1970). Although MINERVA 2 was not conceived as a model of the lexicon,

[6] Again, note that potential explicit memory contamination of the implicit memory data cannot be assessed in Experiment 2. Clearly, some words in perceptual identification must have been explicitly recalled, especially with the 5-min delay. In our laboratory, we are currently examining memory for words and voices using Jacoby's (1991) process-dissociation method (Jacoby, Toth, & Yonelinas, 1993; Richardson-Klavenh, Gardiner, & Java, 1994). The data collected thus far indicate a stronger voice effect for implicit (unconscious) memory.

Hintzman (1986) discussed its potential solutions to problems in word identification. For example, the model naturally derives context-sensitive interpretations of stimuli, which could help resolve lexical ambiguity. With minor modifications, MINERVA 2 or a related model (Eich, 1982; Gillund & Shiffrin, 1984; Nosofsky, 1991) could be applied as a model of the lexicon.

As another example, Feustel et al. (1983; Salasoo et al., 1985) described a hybrid model, in which both abstract lexical codes and episodic traces contribute to word perception. According to this view, words become codified by means of repetition; multiple episodes coalesce into unitary codes, similar to logogens. Episodes mediate token-specific repetition effects, but abstract codes provide the lexicon stability and permanence. Similarly, Kirsner and Dunn (1985; Kirsner, Dunn, & Standen, 1987) proposed a lexicon of abstract representations and episodic procedural records. According to this view, word identification entails processes that match stimuli to lexical entries; detailed records of these processes are stored as episodic traces. Surface details, such as voice, lend structure to the record. Upon later word identification, past records are reapplied to the degree they resemble new inputs (see Kolers, 1976; Kolers & Ostry, 1974).

Predictions of the record-based model easily apply to the present data: Recognizing a word in an unfamiliar voice will entail normalization and matching procedures, stored in a record. Later identification of the same stimulus will use the record, creating a strong repetition effect. Or, if a similar stimulus is presented, its perceptual operations will partially overlap the previous record, creating residual savings. With increased exposure to a speaker's voice (or typeface, rotated text, foreign accent, etc.), the increasing episode collection will support asymptotic (totally normalized) performance. This prediction was recently supported by Nygaard, Sommers, and Pisoni (1994), who made listeners familiar with speakers' voices and found facilitated identification of new words produced by those speakers.

The Active Lexicon?

Although the present research showed voice effects in memory and perception of spoken words, such effects were not purely stimulus driven. Instead, voice effects were stronger when listeners focused their attention on more superficial attributes at study. This suggests that episodic traces are not perceptual analogues that are totally defined by stimulus properties. Rather, they appear as complex perceptual–cognitive objects, jointly specified by perceptual forms and linguistic functions (Van Orden & Goldinger, 1994). Indeed, Masson and Freedman (1990) proposed that repetition effects are based on context-specific episodes, jointly specified by stimulus details and encoding processes (see also Blaxton, 1989; Whittlesea & Cantwell, 1987). In particular, effects of stimulus details (such as voice) are stronger when both study and test tasks involve more data-driven processing, such as identifying rotated text or words in noise.[7]

The focus of attention during study may explain an aspect of the present data that conflicts with previous research. Schacter and Church (1992) found significant voice effects only with words presented clearly, not with words in noise. They noted that the left cerebral hemisphere primarily operates on abstract information (Safer & Leventhal, 1977) and the right hemisphere on perceptual information, such as voice (Van Lancker & Kreiman, 1987). Zaidel (1978) reported that speech-like noise impairs right-hemisphere processing, so voice effects should be minimal when words are masked by noise. However, Experiment 2 revealed voice effects, despite noise. A key difference in method may be responsible: In Schacter and Church's (1992) research, study words were always presented in the clear, and test words were presented in noise. In Experiment 2, noise was used in both sessions, which could magnify voice effects: As noted earlier, noise is typically assumed only to increase perceptual difficulty. However, if the masking noise is encoded with other stimulus details, repeating exact signal-plus-noise events during test could increase repetition effects (Hintzman, 1986; Jacoby, 1983).

A similar proposal arises from theories of transfer-appropriate processing (Blaxton, 1989; Graf & Ryan, 1990; Roediger, Weldon, & Challis, 1989), which state that episodic memory is most strongly expressed when study operations are later repeated at test. When a listener hears words in noise, some perceptual (data-driven) operations are likely required for correct identification. Upon hearing the same word in noise again, similar operations will be used, more fluently than before, yielding a repetition effect (Jacoby, 1983; Kirsner & Dunn, 1985; Kolers & Ostry, 1974). Indeed, this account was recently supported by Saldaña, Nygaard, and Pisoni (1996), who examined continuous recognition memory for same- and different-voice words (cf. Palmeri et al., 1993). Between subjects, words were presented in the clear, in soft noise, in moderate noise, or in loud noise. The results were systematic and striking—a modest voice effect emerged with words in the clear, which grew more robust as the noise grew louder. Because all words were presented in noise throughout sessions (as in Experiment 2), same-voice benefits (perhaps on perceptual operations) were observed.[8] An analogy to the present Voice $\times$ LOP interaction is apparent: Perceptual–cognitive operations on words during study (either overt semantic classification or automatic extraction from noise) influence their potential repetition benefit during test. Listeners who attended to voice attributes during study (in gender classification) were more sensitive to later voice changes.

Considering this role of attention may help rationalize episodic models in several respects. For example, episodic models provide an intuitive account of token-specific repetition effects, but have generally weak intuitive appeal. Even

---

[7] There are clear exceptions to this pattern. Jacoby et al. (1992) and Woltz (1990) observed font repetition effects for words and passages that were read for meaning.

[8] Another recent confirmation was provided by Sheffert (1995), who contrasted the studies by Schacter and Church (1992) and Goldinger (1992). The data patterns from both investigations were replicated, and the role of noise at study was directly assessed. As predicted, presenting words in noise at both study and test increased the voice effect. Sheffert attributed this result to the repeated, data-driven operations required when words must be perceptually extracted from noise.

with some forgetting assumed (e.g., Hintzman, 1986), it is difficult to imagine that countless instances of words are retained in memory. Another problem regards the ambiguous boundaries of linguistic events. In the laboratory, episodes naturally conform to experimental trials, but real language is a hierarchical system in which words are fairly subordinate entities. Another problem concerns common experience—in normal speech communication, people converse in the realm of ideas. Listeners are not typically aware of processing tangential information, such as voice details, environmental context, and so on. In short, perception seems intuitively normalized, as the present data suggest when listeners operate at deeper LOPs.

However, if the focus of attention in perception partially specifies episodic content, surface details of words may be encoded (or prominent) only to the extent that they matter in original processing (Whittlesea & Cantwell, 1987). If voice details are tangential to the linguistic function of speech, they may be minimized in episodic traces, as superficial details of pennies are typically absent from memory (Nickerson & Adams, 1979). Because speech is typically used to convey messages, episodic traces will typically emphasize elements of meaning, not perception. Ideas may be distributed over long or short utterances, so this also predicts flexible episodic boundaries. The episodic lexicon may be more than a word collection; it may contain a rich linguistic history, reflecting words in various contexts, nuances, fonts, and voices. The present investigation only examined voice details, but the logic of episodic representation easily extends to other contextual details, including semantics.

Conclusion

Jacoby (1983) noted that "there is a great deal of unexploited similarity between theories of episodic memory and theories of perception. . . . The difference . . . is largely removed if it is assumed both types of task involve parallel access to a large population of memories for prior episodes" (pp. 35–36; see also Logan, 1988). The results of the present investigation, taken together with related findings (e.g., Church & Schacter, 1994), support an episodic view of the lexicon, in which words are recognized against a background of countless, detailed traces. Speech is not a noisy vehicle of linguistic content; the medium may be an integral dimension of later representation.

References

Blaxton, T. A. (1989). Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 657–668.

Brown, J., & Carr, T. (1993). Limits on perceptual abstraction in reading: Asymmetric transfer between surface forms differing in typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 1277–1296.

Carr, T., Brown, J., & Charalambous, A. (1989). Repetition and reading: Perceptual encoding mechanisms are very abstract but not very interactive. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 763–778.

Carterette, E. C., & Barnebey, A. (1975). Recognition memory for voices. In A. Cohen & S. G. Nooteboom (Eds.), *Structure and process in speech perception* (pp. 246–265). New York: Springer-Verlag.

Cave, C. B., & Squire, L. R. (1992). Intact and long-lasting repetition priming in amnesia. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 509–520.

Challis, B. H., & Brodbeck, D. R. (1992). Level of processing affects priming in word fragment completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 595–607.

Church, B., & Schacter, D. L. (1994). Perceptual specificity of auditory priming: Memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 521–533.

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

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

Craik, F., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104,* 268–294.

Egan, J. P. (1948). Articulation testing methods. *Laryngoscope, 58,* 955–991.

Eich, J. M. (1982). A composite holographic associative recall model. *Psychological Review, 89,* 627–661.

Feustel, T., Shiffrin, R., & Salasoo, A. (1983). Episodic and lexical contributions to the repetition effect in word recognition. *Journal of Experimental Psychology: General, 112,* 309–346.

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

Forster, K. I. (1979). Levels of processing and the structure of the language processor. In W. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 27–86). Hillsdale, NJ: Erlbaum.

Garner, W. (1974). *The processing of information and structure.* Potomac, MD: Erlbaum.

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

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1–67.

Goldinger, S. D. (1992). *Words and voices: Implicit and explicit memory for spoken words* (Research on speech perception Tech. Rep. No. 7). Bloomington, IN: Indiana University Press.

Goldinger, S. D., Pisoni, D. B., & Logan, J. S. (1991). On the nature of talker variability effects on serial recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 152–162.

Goldinger, S. D., Pisoni, D. B., & Luce, P. A. (1996). Speech perception and spoken word recognition: Research and theory. In N. J. Lass (Ed.), *Principles of experimental phonetics* (pp. 277–327). St. Louis, MO: Mosby Year-Book.

Graf, P., & Mandler, G. (1984). Activation makes words more accessible but not necessarily more retrievable. *Journal of Verbal Learning and Verbal Behavior, 23,* 553–568.

Graf, P., & Ryan, L. (1990). Transfer-appropriate processing for implicit and explicit memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 978–992.

Green, K., Kuhl, P., Meltzoff, A., & Stevens, E. (1991). Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the McGurk effect. *Perception & Psychophysics, 50,* 524–536.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411–428.

Hintzman, D. L., Block, R., & Inskeep, N. (1972). Memory for mode of input. *Journal of Verbal Learning and Verbal Behavior, 11,* 741–749.

Hintzman, D. L., & Ludlam, G. (1980). Differential forgetting of prototypes and old instances: Simulation by an exemplar-based classification model. *Memory & Cognition, 8,* 378–382.

House, A., Williams, C., Hecker, M., & Kryter, K. (1965). Articulation-testing methods: Consonantal differentiation with a closed-response set. *Journal of the Acoustical Society of America, 37,* 158–166.

Jackson, A., & Morton, J. (1984). Facilitation of auditory word recognition. *Memory & Cognition, 12,* 568–574.

Jacoby, L. L. (1983). Perceptual enhancement: Persistent effects of an experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 21–38.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30,* 513–541.

Jacoby, L. L., & Brooks, L. (1984). Nonanalytic cognition: Memory, perception, and concept learning. In G. Bower (Ed.), *The psychology of learning and motivation, Vol. 18* (pp. 1–47). New York: Academic Press.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110,* 306–340.

Jacoby, L. L., & Hayman, C. A. G. (1987). Specific visual transfer in word identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 456–463.

Jacoby, L. L., Levy, B. A., & Steinbach, K. (1992). Episodic transfer and automaticity: Integration of data-driven and conceptually-driven processing in rereading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 15–24.

Jacoby, L. L., Marriott, M. J., & Collins, J. G. (1990). The specifics of memory and cognition. In T. K. Srull & R. S. Wyer Jr. (Eds.), *Advances in social cognition, Vol. III* (pp. 111–121). Hillsdale, NJ: Erlbaum.

Jacoby, L. L., Toth, J. P., & Yonelinas, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General, 122,* 139–154.

Johnston, W., Dark, V., & Jacoby, L. L. (1985). Perceptual fluency and recognition judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 3–11.

Joos, M. A. (1948). Acoustic phonetics. *Language 24* (Suppl. 2), 1–136.

Kirsner, K. (1973). An analysis of the visual component in recognition memory for verbal stimuli. *Memory & Cognition, 1,* 449–453.

Kirsner, K., & Dunn, J. (1985). The perceptual record: A common factor in repetition priming and attribute retention? In M. Posner & O. Marin (Eds.), *Attention and performance XI* (pp. 547–565). Hillsdale, NJ: Erlbaum.

Kirsner, K., Dunn, J., & Standen, P. (1987). Record-based word recognition. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 147–167). Hillsdale, NJ: Erlbaum.

Klatt, D. (1979). Speech perception: A model of acoustic–phonetic analysis and lexical access. *Journal of Phonetics, 7,* 279–312.

Kolers, P. A. (1976). Reading a year later. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 554–565.

Kolers, P. A., & Ostry, D. (1974). Time course of loss of information regarding pattern analyzing operations. *Journal of Verbal Learning and Verbal Behavior, 13,* 599–612.

Krulee, G., Tondo, D., & Wightman, F. (1983). Speech perception as a multilevel processing system. *Journal of Psycholinguistic Research, 12,* 531–554.

Kruskal, J., & Wish, M. (1978). *Multidimensional scaling.* London: Sage University Press.

Ladefoged, P. (1980). What are linguistic sounds made of? *Language, 56,* 485–502.

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

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21,* 1–36.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95,* 492–527.

Manso de Zuniga, C., Humphreys, G., & Evett, L. (1991). Additive and interactive effects of repetition, degradation, and word frequency in the reading of handwriting. In D. Besner & G. Humphreys (Eds.), *Basic processes in reading* (pp. 10–33). Hillsdale, NJ: Erlbaum.

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

Masson, M. (1984). Memory for the surface structure of sentences: Remembering with and without awareness. *Journal of Verbal Learning and Verbal Behavior, 23,* 579–592.

Masson, M., & Freedman, L. (1990). Fluent identification of repeated words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 355–373.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1–86.

McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General, 114,* 159–188.

McGehee, F. (1937). The reliability of the identification of the human voice. *Journal of General Psychology, 17,* 249–271.

Monsen, R. B., & Engebretson, A. M. (1977). Study of variations in the male and female glottal wave. *Journal of the Acoustical Society of America, 62,* 981–993.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review, 76,* 165–178.

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

Mullennix, J. W., Pisoni, D. B., & Martin, C. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America, 85,* 365–378.

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

Musen, G., & Treisman, A. (1990). Implicit and explicit memory for visual patterns. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 127–137.

Neely, J. H. (1989). Experimental dissociations in the semantic/episodic memory distinction. In H. L. Roediger III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 229–270). Hillsdale, NJ: Erlbaum.

Nickerson, R., & Adams, M. (1979). Long-term memory for a common object. *Cognitive Psychology, 11,* 287–307.

Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 3–27.

Nusbaum, H. C., & Morin, T. (1992). Paying attention to differences among talkers. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech perception, production, and linguistic structure* (pp. 55–83). Tokyo: IOS Press.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science, 5,* 42–46.

Oldfield, R. C. (1966). Things, words, and the brain. *Quarterly Journal of Experimental Psychology, 18,* 340–353.

Paap, K., Newsome, S., McDonald, J., & Schvaneveldt, R. (1982). An activation-verification model for letter and word recognition: The word-superiority effect. *Psychological Review, 89,* 573–594.

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 309–328.

Papçun, G., Kreiman, J., & Davis, A. (1989). Long-term memory for unfamiliar voices. *Journal of the Acoustical Society of America, 85,* 913–925.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24,* 175–184.

Pisoni, D. B. (1993). Long-term memory in speech perception: Some new findings on talker variability, speaking rate, and perceptual learning. *Speech Communication, 13,* 109–125.

Pisoni, D. B., & Sawusch, J. R. (1975). Some stages of processing in speech perception. In A. Cohen & S. Nooteboom (Eds.), *Structure and process in speech perception* (pp. 16–34). Berlin, Germany: Springer-Verlag.

Podgorny, P., & Garner, W. R. (1979). Reaction time as a measure of inter- and intraobject visual similarity: Letters of the alphabet. *Perception & Psychophysics, 26,* 37–52.

Posner, M. (1964). Information reduction in the analysis of sequential tasks. *Psychological Review, 71,* 491–503.

Posner, M., & Keele, S. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77,* 353–363.

Posner, M., & Keele, S. (1970). Retention of abstract ideas. *Journal of Experimental Psychology, 83,* 304–308.

Richardson-Klavenh, A., Gardiner, J., & Java, R. (1994). Involuntary conscious memory and the method of opposition. *Memory, 2,* 1–29.

Roediger, H. L., III, & Blaxton, T. A. (1987). Effects of varying modality, surface features, and retention interval on priming in word-fragment completion. *Memory & Cognition, 15,* 379–388.

Roediger, H. L., III, & McDermott, K. B. (1993). Implicit memory in normal human subjects. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology, Vol. 8* (pp. 63–131). Amsterdam: Elsevier.

Roediger, H. L., III, & Srinivas, K. (1992). Specificity of operations in perceptual priming. In P. Graf & M. Masson (Eds.), *Implicit memory: New directions in cognition, development, and neuropsychology.* Hillsdale, NJ: Erlbaum.

Roediger, H. L., III, Weldon, M. S., & Challis, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: A processing account. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 3–14). Hillsdale, NJ: Erlbaum.

Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology, 89,* 43–50.

Safer, M., & Leventhal, H. (1977). Ear differences in evaluating emotional tones of voice and verbal content. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 75–82.

Salasoo, A., Shiffrin, R., & Feustel, T. (1985). Building permanent memory codes: Codification and repetition effects in word identification. *Journal of Experimental Psychology: General, 114,* 50–77.

Saldaña, H. M., Nygaard, L. C., & Pisoni, D. B. (1996). *Episodic encoding of visual speaker attributes and recognition memory for spoken words.* Manuscript in preparation.

Scarborough, D., Cortese, C., & Scarborough, H. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 1–17.

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 501–518.

Schacter, D. L., & Church, B. (1992). Auditory priming: Implicit and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 915–930.

Schacter, D. L., Church, B., & Osowiecki, D. M. (1994). Auditory priming in elderly adults: Impairment of voice-specific memory. *Memory, 2,* 295–323.

Schacter, D. L., Church, B., & Treadwell, J. (1994). Implicit memory in amnesic patients: Evidence for spared auditory priming. *Psychological Science, 5,* 20–25.

Schacter, D. L., Eich, J. E., & Tulving, E. (1978). Richard Semon's theory of memory. *Journal of Verbal Learning and Verbal Behavior, 17,* 721–743.

Semon, R. (1923). *Mnemic psychology.* (B. Duffy, Trans.). London: George Allen & Unwin. (Original work published 1909.)

Sergent, J., & Takane, Y. (1987). Structures in two-choice reaction-time data. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 300–315.

Sheffert, S. M. (1995). *Implicit and explicit memory for words and voices.* Unpublished doctoral dissertation, University of Connecticut.

Shepard, R., & Teghtsoonian, M. (1961). Retention of information under conditions approaching a steady state. *Journal of Experimental Psychology, 62,* 302–309.

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

Tenpenny, P. L. (1995). Abstractionist versus episodic theories of repetition priming and word identification. *Psychonomic Bulletin & Review, 2,* 339–363.

Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science, 247,* 301–306.

Tulving, E., Schacter, D. L., & Stark, H. (1982). Priming effects in word fragment completion are independent of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 336–342.

Underwood, B. (1969). Attributes of memory. *Psychological Review, 76,* 559–573.

Van Lancker, D., & Kreiman, J. (1987). Voice discrimination and recognition are separate abilities. *Neuropsychologica, 25,* 829–834.

Van Lancker, D., Kreiman, J., & Emmorey, K. (1985). Familiar voice recognition: Patterns and parameters. Part I: Recognition of backward voices. *Journal of Phonetics, 13,* 19–38.

Van Lancker, D., Kreiman, J., & Wickens, T. (1985). Familiar voice recognition: Patterns and parameters. Part II: Recognition of rate-altered voices. *Journal of Phonetics, 13,* 39–52.

Van Orden, G. C., & Goldinger, S. D. (1994). The interdependence of form and function in cognitive systems explains perception of printed words. *Journal of Experimental Psychology: Human Perception & Performance, 20,* 1269–1291.

Weiner, F., & Singh, S. (1974). Multidimensional analysis of choice reaction time judgments on pairs of English fricatives. *Journal of Experimental Psychology, 102,* 615–620.

Weldon, M. S. (1991). Mechanisms underlying priming on perceptual tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 526–541.

Whittlesea, B., & Brooks, L. (1988). Critical influence of particular experiences in the perception of letters, words, and phrases. *Memory & Cognition, 16,* 387–399.

Whittlesea, B., & Cantwell, A. (1987). Enduring influences of the purpose of experiences: Encoding-retrieval interactions in word and pseudoword perception. *Memory & Cognition, 15,* 465–472.

Woltz, D. (1990). Repetition of semantic comparisons: Temporary and persistent priming effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 392–403.

Zaidel, E. (1978). Lexical organization in the right hemisphere. In P. Buser & A. Rougeul-Buser (Eds.), *Cerebral correlates of conscious experience* (pp. 177–197). Amsterdam: Elsevier.

Appendix

Similarity Matrix Derived in Experiment 1

Speaker	F1	F2	F3	F4	F5	M1	M2	M3	M4	M5
F1	—	716	794	694	670	769	874	852	896	822
F2		—	777	728	709	756	867	746	746	712
F3			—	719	719	812	789	714	733	882
F4				—	702	665	695	696	743	783
F5					—	748	796	887	833	858
M1						—	706	759	738	756
M2							—	733	719	748
M3								—	707	714
M4									—	700
M5										—

Note. Symbols F1–F5 denote female speakers; M1–M5 denote male speakers. Values shown are mean correct same-word classification times in milliseconds.

Received February 10, 1995
Revision received February 20, 1996
Accepted February 20, 1996 ■