Perception & Psychophysics
2004, 66 (2), 352-362

Audiovisual speech facilitates voice learning

SONYA M. SHEFFERT and ELIZABETH OLSON
Central Michigan University, Mount Pleasant, Michigan

In this research, we investigated the effects of voice and face information on the perceptual learning of talkers and on long-term memory for spoken words. In the first phase, listeners were trained over several days to identify voices from words presented auditorily or audiovisually. The training data showed that visual information about speakers enhanced voice learning, revealing cross-modal connections in talker processing akin to those observed in speech processing. In the second phase, the listeners completed an auditory or audiovisual word recognition memory test in which equal numbers of words were spoken by familiar and unfamiliar talkers. The data showed that words presented by familiar talkers were more likely to be retrieved from episodic memory, regardless of modality. Together, these findings provide new information about the representational code underlying familiar talker recognition and the role of stimulus familiarity in episodic word recognition.

In this research, we were concerned with two different but related topics in spoken language processing: talker recognition and word recognition. In particular, we examined the contribution of audiovisual speech information to the learning and recognition of voices and the subsequent transfer of this talker-specific knowledge to a different task situation (episodic word recognition). Guiding this research was the idea that auditory and visual speaker recognition might be based on features similar to those used to recognize audiovisual speech and, hence, might show similar performance characteristics.

Psycholinguists and speech scientists have known for some time that speech perception is not simply an auditory process. Instead, the mouth movements we see influence what we hear. The availability of visual articulation from a talker's face serves to disambiguate acoustically confusable speech elements and improve word identification, particularly if the environment is noisy, if the listener has a hearing impairment, or if the auditory message is grammatically complex (Bernstein, Demorest, & Tucker, 2000; Erber, 1969; Reisberg, McLean, & Goldfield, 1987; Sumby & Pollack, 1954). Even under ideal listening conditions, people automatically and unconsciously combine information from both modalities (Liberman, 1982; McGurk & MacDonald, 1976; Summerfield, 1987). In the *McGurk effect*, for example, a mismatch between the speech from a voice and a face can cause an observer to report "hearing" a speech sound that represents a combination of phonetic features from each source. Auditory and visual speech integrate because each modality provides information about the same articulatory event.

Correspondence concerning this article should be addressed to S. M. Sheffert, Psychology Department, Central Michigan University, Mount Pleasant, MI 48859 (e-mail: sonya.sheffert@cmich.edu).

Note—This article was accepted by the previous editorial team, headed by Neil Macmillan.

The Relationship Between Speaker Recognition and Speech Recognition

Audiovisual speech also conveys socially relevant information about the individual's identity and other personal qualities, as well as enhancing the availability of phonetic information. Theorists have traditionally assumed that the features of the speech signal (whether heard or seen) that carry the linguistic message are independent of the features that carry a talker's identity (Bruce, 1988; Ellis, 1986; Halle, 1985; Laver & Trudgill, 1979). Because phonetic attributes are used for linguistic processing, it was assumed that nonlinguistic attributes are used for talker identification. Consequently, the literature on speaker recognition has been devoted almost exclusively to cues that reflect a talker's unique anatomical properties but are phonetically irrelevant (Bricker & Pruzansky, 1976). As one example, different modes of laryngeal vibration and vocal tract anatomy give rise to qualitative differences in voice quality (i.e., pitch, timbre, nasality, creakiness, etc.) without altering the consonants and vowels substantially. Many studies have demonstrated that qualitative, anatomically based cues are important for speaker recognition. Idiosyncratic variations in phonetic features are also among the cues used to identify talkers. Different speakers pronounce speech segments in different ways. Idiosyncratic pronunciation habits, or *idiolect*, allow listeners to identify individual talkers, even those who have similar vocal tract sizes and shapes and who share the same dialect (e.g., identical twins).

There is growing interest in clarifying the role of phonetic attributes in talker identification, driven by recent empirical evidence that idiolectal variation is effective in the learning and recognition of voices and faces. In a series of studies by Remez and his co-workers (Remez, Fellowes, & Rubin, 1997; Sheffert, Pisoni, Fellowes, & Remez, 2002), phonetic information was isolated from qualitative attributes of voice quality by using sine wave replicas of natural speech. Remez et al. found that listeners

perform well above chance at identifying colleagues from sine wave replicas of their natural speech productions. Moreover, a hierarchical cluster analysis of listener's identification errors indicated that perceived similarity among the sine wave talkers was based largely on shared specific pronunciation habits (dialect or idiolect), independently of talker gender (Fellowes, Remez, & Rubin, 1997). Remez et al. postulated an "idiolectal identification" hypothesis, whereby linguistic and talker perception tap a common representational code composed of phonetic attributes. Using a perceptual training paradigm, Sheffert et al. (2002) went on to show that people can learn to recognize sine wave voices, despite the fact that the signals lack the qualitative attributes of vocal sound production that have traditionally been assumed to be indispensable for voice learning.

Point-light speech can be thought of as a visual analogue to sine wave speech, in the sense that both convey phonetic information though dynamic patterns reflecting vocal tract articulation (Rosenblum & Saldaña, 1998). Articulatory movements can be isolated from other aspects of the face by placing illuminated dots (point lights) on a talker's cheeks, lips, tongue, and teeth and then filming the face in the dark; the observer sees only the configuration of moving dots. Using this methodology, Rosenblum et al. (2002) discovered that observers can match a fully illuminated talking face to its point-light counterpart. Rosenblum et al. argued that a talker's specific style of mouth motion can facilitate face recognition (see also Bassili, 1979; Bruce & Valentine, 1988; Christie & Bruce, 1998; Lander, Christie, & Bruce, 1999).

Theoretically, these results challenge the traditional separation between linguistic and speaker processing by showing that articulatory/phonetic features are not used solely for recognizing linguistic information but can also be recruited for the recognition of a talker's voice or face. The notion that common or redundant representations may be used for linguistic and talker processing provides a fairly straightforward account of these effects. Moreover, this explanatory framework also offers a possible account for numerous reports of contingencies between linguistic and talker processing (see Pisoni, 1996). For example, voice and phonetic dimensions appear to be processed in an integral manner (Green, Tomiak, & Kuhl, 1997; Mullennix & Pisoni, 1990). Other evidence has shown that talker information is retained in a word's episodic memory trace and influences implicit and explicit word retrieval (Craik & Kirsner, 1974; Goldinger, 1996, 1998; Palmeri, Goldinger, & Pisoni, 1993; Schacter & Church, 1992; Sheffert, 1998a, 1998b). Dynamic visible speaker information has also been linked with word and auditory speaker memory representations (Saldaña, Nygaard, & Pisoni, 1996; Sheffert & Fowler, 1995). The contingency between talker familiarity and speech processing is most germane to the present project and will be described in the next section.

Talker Familiarity and Linguistic Processing

Listeners often encounter comprehension difficulties when listening to an individual with an unusual dialect or voice quality. However, understanding the talker becomes much easier after the listener becomes accustomed to the talker's idiosyncratic speaking style. For example, children with a hearing impairment find that the speech of familiar family and friends is somehow *clearer* than speech produced by strangers. The reverse is also true; individuals who spend a large amount of time with a hearing-impaired child who often has unclear speech can better understand what the child is attempting to communicate than can those individuals who come only into casual contact with the hearing-impaired child.

These anecdotal observations are in line with studies that have shown that knowledge of a talker's voice has a direct effect on the perceptual analysis of his or her speech (Nygaard & Pisoni, 1998; Nygaard, Sommers, & Pisoni, 1994). In these experiments, listeners were trained over several days to recognize a set of talkers from auditory words or sentences. After learning the talkers, listeners completed a speech intelligibility test in which he or she attempted to identify a new set of words that were masked by noise. Nygaard et al. found that words spoken by familiar talkers were easier to perceive than words spoken by unfamiliar talkers. According to Nygaard et al., voice recognition became increasingly "proceduralized" or automatic over the course of training. The increased efficiency of voice identification, in turn, increased sensitivity to the linguistic information in the signal.

There is some evidence that talker familiarity can affect visual speech processing. For example, classification of visual phonemes from pictures of faces mouthing "ee" or "oo" is slower when the identity of the face changes from trial to trial (Schweinberger & Soukup, 1998). Talker variability reduces an observer's ability to lip-read sentences, relative to a single-talker list (Yakel, Rosenblum, & Fortier, 2000). Finally, Walker, Bruce, and O'Malley (1995), in a study using dynamic audiovisual materials, discovered that familiarity with a set of audiovisual talkers reduced the McGurk effect (e.g., auditory speech perception was less biased by incongruent visual mouth movements). Although all the familiar audiovisual talkers used by Walker et al. were well known to the participants (co-workers, etc.), a post hoc study showed that voice identification rates were extremely poor (merely 11%), relative to face identification (which approached ceiling). Consequently, Walker et al. concluded that the effects of talker familiarity on audiovisual speech perception were due to face familiarity, rather than to voice familiarity.

The fact that face and voice features differed in their contribution to judgments of talker familiarity provides a strong justification for the use of a familiarization procedure that measures voice knowledge independently of face knowledge. Such a procedure has been used in the present study.

At issue here is not the relative memorability of faces and voices; the consensus is that faces are easier to learn and recall than voices (Cook & Wilding, 1997; Legge, Grosmann, & Pieper, 1984; Shepard, 1967; Woodhead, Baddeley, & Simmonds, 1979; Yarmey, 1986). What remains to be established is how voice information and visible speaker information interact during talker recognition and speech recognition. For example, the fact that seeing a talker's face can improve phonetic perception, coupled with the fact that phonetic features can support talker recognition, raises the possibility that voice learning might be enhanced by facial speech gestures.

Unfortunately, the empirical data are inconclusive and, in some cases, suggest a very different hypothesis. For example, Legge et al. (1984) reported that pictorial face information had no reliable effect on voice learning (see also Armstrong & McKelvie, 1996; Yarmey, 1993). Specifically, Legge et al. showed that a talker's voice was recognized nominally better when the target voice was presented with an associated face image at study (approximately 60% correct voice identification for short speech samples; see their Figure 3). Voice recognition was nominally lower when there was no face context at study (approximately 52% correct voice identification). However, the pattern was reversed for long speech samples (>40 sec): Voice recognition was lower when a voice was accompanied by a face image at study. Using an auditory voice line-up task and sentence-length utterances, Cook and Wilding (1997, 2001) reported that a familiar voice was less likely to be recognized if it was originally encoded in the context of a picture or videotape of the talker's face. Cook and Wilding (1997, 2001) referred to this as the "face overshadowing effect" and argued that it reflects an attentional bias toward face information for the task of person recognition and toward voice information for the task of spoken language comprehension. Accordingly, faces can be expected to interact more with voice recognition than with word recognition, and there is support for this in the empirical literature (e.g., Sheffert & Fowler, 1995). Cook and Wilding (2001) also suggested that the extent to which the face dominates or overshadows the voice may be reduced if observers have become habituated to the face.

Study Objectives and Motivation

The first objective of the present study was to determine whether multimodal speech would affect voice learning. Using a perceptual-learning task with feedback, we trained listeners to identify five individuals to a specific criterion from only the talker's voice (auditory condition [A]) or from the talker's voice and articulating face (audiovisual condition [AV]). Voice learning was measured after each training session, using a new set of auditory-only words. By testing transfer to a new set of auditory-only words, we could also determine whether the representations that developed during AV or A training displays were sufficiently abstract to allow listeners to identify talkers from different words and a different modality.

The second objective of the study was to assess the effects of talker familiarity on episodic word recognition. Although much of the existing research has focused on how talker familiarity enhances the perceptual analysis of speech, the research reported here went a step further by examining the long-term mnemonic consequences of talker familiarity. Ideally, the value of talker familiarity would lie not just in identifying speech events, but also in translating them into detailed long-term representations. To explore this issue, trained participants completed a standard *old/new* word recognition task in which familiar and unfamiliar talkers produced equal numbers of study and test words. If enhanced memory storage is a by-product of enhanced perception, words spoken by familiar talkers would be more accessible. Because auditory and audiovisual speech might differ in memorability, we also varied modality. For half the participants, the study and test lists were in the auditory modality; the remaining participants received audiovisual words.

METHOD

Participants

The participants consisted of 40 Central Michigan University students. All were native speakers of English and reported having normal hearing and normal or corrected vision. Volunteers received extra credit in an undergraduate psychology class or were paid $6 for each hour of participation. Seven participants failed to complete the experiment (5 were dropped during the training phase, due to schedule conflicts, and 2 appeared to be responding randomly on the recognition test).

Design

A mixed factorial design was used. Training mode (A or AV) and word recognition test mode (A or AV) were manipulated between subjects. Half the participants were randomly assigned to learn the talkers from only the talker's voice (A), and the other half learned the talkers from the talker's voice and the corresponding articulating face (AV). After learning the talkers, half the participants from each training condition were then randomly assigned to either the A or the AV word recognition test. Thus, there were four groups of 10 participants, each group representing a different talker training × word recognition test combination: auditory–auditory, auditory–audiovisual, audiovisual–auditory, and audiovisual–audiovisual. Talker familiarity (recognition test words presented by old, familiar talkers or by new, unfamiliar talkers) was manipulated within subjects.

Materials

The stimulus materials consisted of dynamic full-motion audiovisual tokens of 10 talkers (five males and five females), producing 262 individual words (Sheffert, Lachs, & Hernández, 1996–1997). Each talker produced the same words. The words were monosyllabic consonant–vowel–consonant words (see the Appendix) that were highly familiar (i.e., all the words were judged to be highly familiar by college students, with a familiarity rating of at least 6.7 on a 7-point scale, where 1 is the *lowest* and 7 is the *highest*) and highly intelligible (overall intelligibility exceeded 90% correct for each talker under auditory-only identification conditions).

To create the materials, we videotaped each talker's utterances in a sound-attenuated recording studio and then converted the recordings to digital format, using a Macintosh computer and Adobe Premier software. The video signal was digitally sampled at 30 frames per second, with 24-bit resolution at 640 × 480 pixel size. The

audio signal was digitally sampled at 22 kHz and equated for root-mean squared amplitude. Acoustically, the talker ensemble was relatively homogenous, representing dialects from different geographical areas in the Midwestern United States. Visually, all the talkers were Caucasian, ranging in age from 18 to 32 years. In all the displays, the talkers looked directly into the camera and were shown from the neck up. The onset of each word token began and ended with a closed mouth. Other factors that might affect stimulus discriminability, such as lighting, image size, background, pose, and expression, were controlled.

From this corpus of words, we created four different sets of materials by randomly selecting, without replacement, 50 training words, 50 generalization words, and 160 word recognition test words. Thus, a word used in the training phase for one participant might be used in the generalization or recognition phase for another participant. After selecting the words for each set, we randomly assigned a talker to each word. During training and generalization, a participant was exposed to 5 of the 10 talkers (e.g., two males and three females or vice versa). The remaining 5 talkers served as unfamiliar *control* talkers during the word recognition test phase. None of the talkers was personally familiar to the participants before the study.

The materials for the training and the generalization tasks consisted of a random ordering of five repetitions of 10 words from each of the five talkers (250 words total). Each talker spoke the same 10 words. After each training session, participants were tested on their ability to identify the five voices, using a different set of 10 words, each repeated five times (250 words total). The generalization task was always auditory only, regardless of the training condition. If a participant failed to reach at least 75% correct on the voice generalization test, he or she returned the next day for another training and generalization session, each task using a new set of 10 words. When a participant reached criterion, he or she returned the next day for a word recognition test.

The materials for the word recognition test consisted of 80 target words presented by 10 talkers (5 familiar, 5 unfamiliar) and 80 new distractors, also presented by 10 talkers. Each talker spoke an equal number of words. The talker who spoke a target word at study was also the talker who produced the word at test. For each stimulus set, four different versions of the word recognition list were created, each representing a different random assignment of words to talkers and a different random word order. Finally, two additional words were used for the familiarization task (see the Procedure section), and these words were the same across all training sessions. In all the tasks, the words were presented in a random order (not blocked by talker), with a 5-sec interstimulus interval (ISI), and repetitions were always identical tokens (e.g., no changes in voice, facial expression, pose, or any other aspect of the token).

The various randomized training and test orders used in the present study were generated on a Macintosh desktop computer, using multimedia presentation software (Macromedia Director), and were transferred to S-VHS videotape. The materials for the auditory and the audiovisual conditions were identical, except that the auditory conditions were presented without the video signal.

Procedure

The participants were tested individually in a quiet laboratory in the presence of an experimenter. The stimulus materials were presented using a S-VHS VCR and a 22-in. color television. Collection of the response data was carried out on an IBM-compatible personal computer. The participants were not personally familiar with any of the talkers prior to the experiment.

Each training session consisted of three phases: familiarization, talker training, and generalization. Assignment of participants to training conditions (A or AV) and stimulus sets was random. The modality of the familiarization task was matched to the training condition. However, the generalization task was always auditory only, in order to equate voice familiarity across conditions. The last part of the experiment was a word recognition memory test. The word recognition test consisted of two phases: study and test, with equal numbers of participants assigned to the A or the AV condition.

Familiarization phase. Prior to each talker training session, the participants completed a very brief talker familiarization task designed to help establish or reinstate the correspondence between the speakers and their names. The participants were presented with two words spoken by each of the five talkers, along with their associated names. The same two words were always used in the familiarization phase, and these words were not used in the training, generalization, or word recognition phases. We instructed the participants to attend to talker-specific attributes, rather than to the semantic content of the words. After each word, the experimenter provided the identity of the talker (e.g., "That was Tom"). All the names were common monosyllabic names, such as "Ann," "Jake," or "Steve." In addition, the familiarization presentation mode always matched the training mode. The familiarization task lasted approximately 2 min.

Talker training phase. Following the familiarization task, the participants were presented with a random ordering of five repetitions of 10 words from each of the five talkers (250 words total). Each talker spoke the same 10 words within a given training session. The participants were asked to listen carefully (or listen and watch) during each trial and to attempt to verbally name the talker who presented each word. Each time a participant responded, the accuracy of the response and the name of the correct talker were immediately provided by the experimenter (e.g., "Correct. That was Tom") and were recorded by the experimenter, using a computerized response form. The training task took approximately 30 min.

Generalization phase. After the talker training task, the participants completed a generalization test to determine the extent to which their talker-specific knowledge would transfer to a novel set of words (rather than being tied to the particular training words). The generalization test procedure was identical to the training procedure, with two exceptions. There was no feedback after each trial, and the generalization test was always auditory only, regardless of whether the training had been A or AV. The latter procedure was needed in order to ensure that the listeners in the AV condition were not learning just the talkers' faces (which was trivially easy) but were also learning the voices (which was rather difficult, given the use of short words). The generalization task took approximately 30 min.

If a participant failed to achieve an average of 75% correct voice recognition performance on the generalization test, he or she was asked to return within 24 h for another training session. Familiarization, training, and generalization testing continued until a participant reached criterion on the generalization test.

Word recognition test. After reaching criterion, the participants returned to the laboratory for a test of spoken word recognition. Before beginning the word recognition test, we reassessed the listeners' talker-specific knowledge, to confirm that they were still highly familiar with the talkers at the time of the word recognition test. To this end, the participants completed the brief familiarization task described previously, followed by an abbreviated version of the generalization test. The abbreviated generalization test presented one instance of 10 words from each of the five talkers (50 items total). Corrective feedback was not given, and criterion was again 75% correct. This task took approximately 7 min. All the participants met criterion.

During the study phase of the word recognition task, the participants were presented with 80 words spoken by 10 talkers (5 familiar, 5 unfamiliar), with an equal number of words spoken by each talker (5-sec ISI). They were instructed to listen carefully to the words and to try to remember them in anticipation of a recognition test for the words presented in the study list. We informed the participants of the nature of the materials (number of talkers, words, and the random assignment of talkers to words) and encouraged

them to use whatever strategy they would naturally use if required to remember a fairly long list of unrelated words. No mention was made of the specific hypotheses. After the study phase, the participants engaged in a filler task for 5 min, in which they were given a list of letters and asked to list at least one name of a state from a letter cue (e.g., A–"Alaska," C–"Colorado," etc.). The participants wrote the state names beside the relevant beginning letter. After the filler task, the participants were presented with 160 words, half of which were studied words, with an equal number of words repeated by familiar or unfamiliar talkers (5-sec ISI). We again informed the participants that words would be presented by 5 familiar and 5 unfamiliar talkers, emphasizing that their word recognition judgments were to be based on word type information. The participants were instructed to listen to each word and decide whether it has been presented in the study phase or whether it was a new, unstudied word. The participants made their responses by circling either *old* or *new* on a prepared response form. The entire recognition session lasted less than 1 h, after which the participants were debriefed.

RESULTS

Training and Generalization Performance

Examination of the training data revealed that learning of unfamiliar voices was faster and more accurate when the voices were presented simultaneously with a dynamic articulating face (AV condition). Figure 1 displays the rate of voice learning (operationalized as accuracy on the generalization test) for the A and AV training conditions. The data revealed that by the 2nd day of training, 65% of the participants in the AV condition had reached criterion, as compared with 25% of the participants in the A condition. In fact, only 1 AV participant required four sessions, whereas 7 A participants required four or five sessions.

It is clear from Figure 1 that training performance improved considerably more quickly in the AV condition (2.1 sessions) than in the A condition (3.2 sessions), which was confirmed by a one-way analysis of variance (ANOVA) on the number of training sessions [$F(1,38) = 8.29$, $MS_e = 1.46$, $p = .007$]. Note that in all analyses, $\alpha = .05$.

Figure 2 displays the training and generalization data from the 1st day of training. Two aspects of the figure are noteworthy. The first aspect concerns the relationship between training and generalization within each participant group. For the participants in the A condition, the accuracy levels in the training and the generalization tasks were very similar. This indication of positive transfer across training and generalization shows that the listeners' knowledge of an individual voice was not tied to the particular training tokens, for it was sufficiently abstract to allow transfer to novel instances.

In contrast, the participants in the AV condition showed a markedly different pattern. Here, the participants exhibited a substantial drop between training (when the face of each talker was present) and generalization (when the face was absent). The near-perfect talker identification during training simply reflected the ease with which faces were learned and remembered, relative to voices. Interestingly, however, there was no evidence that the presence of a face during training impaired voice learning, relative to the A condition. In fact, generalization performance was higher after AV training than after A training (71% vs. 63% correct for AV and the A conditions, respectively). This result shows that the knowledge acquired from the AV displays generalized to different words and to a different test modality.

Several statistical tests confirmed this pattern. With respect to talker training, an ANOVA with training condition (A vs. AV) as a factor was conducted on the talker recognition training scores from Day 1 (which included all the participants), Day 2 ($n = 16$ in A, 13 in AV), and Day 3 ($n = 15$ in A, 7 in AV). Table 1 provides the mean talker recognition accuracy for Days 1–3, averaged across participants in each condition. The data from Days 4 and 5 were excluded from the analysis because of insufficient participant numbers. The analyses revealed a highly significant effect of training condition on Day 1 [$F(1,38) = 262.58$, $MS_e = 0.005$, $p < .0001$], Day 2

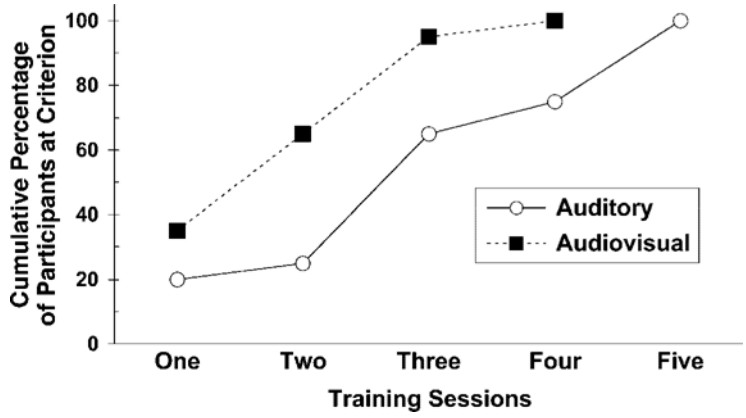

Figure 1. The cumulative percentage of participants at criterion for the auditory and audiovisual training conditions as a function of the number of training sessions.

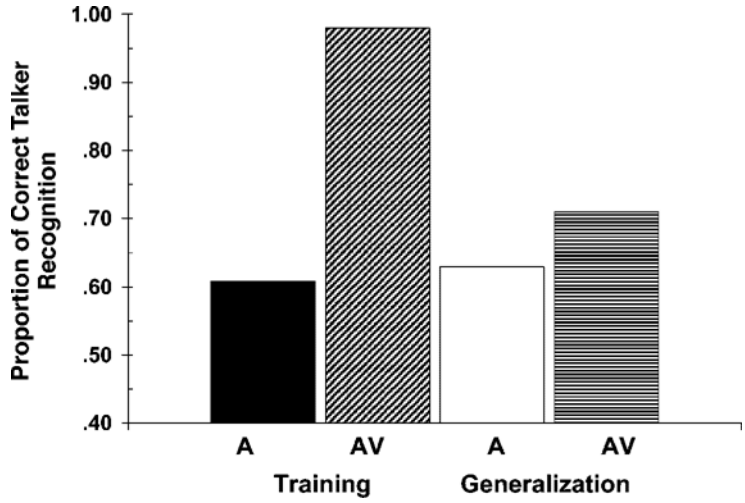

Figure 2. Mean proportions of correct talker recognition on the training and generalization tasks. A, auditory; AV, audiovisual.

[$F(1,27) = 220.06, MS_e = 0.003, p < .0001$], and Day 3 [$F(1,20) = 103.66, MS_e = 0.002, p < .0001$], confirming that talker recognition was far easier in the AV condition.

Examination of performance on the generalization test revealed that voice identification was more accurate after AV training than after A training, despite the fact that the talker's face was never present during the generalization test. ANOVAs comparing generalization across the two conditions (see Table 1) showed significantly higher voice recognition in the AV condition for Day 1 [$F(1,38) = 3.80$, $MS_e = 0.01$, $p < .05$] and Day 2 [$F(1,27) = 11.79, MS_e = 0.006, p < .002$]. Among the participants who were still learning the talkers on Day 3, the difference between the AV and the A conditions did not reach significance ($p = .4$).

In summary, the listeners in the AV training condition took fewer sessions to learn the voices and were more accurate at classifying voices from novel auditory words, relative to the participants in the A condition. This indicates that the presence of visual speaker information improved voice learning.

Table 1
Mean Proportion of Correct Talker Identification in Each Training and Generalization Condition for Days 1–3

	Talker Identification Accuracy	
Training Condition	Training	Generalization
Day 1		
Auditory	.61	.63
Audiovisual	.98	.71
Day 2		
Auditory	.69	.68
Audiovisual	.98	.78
Day 3		
Auditory	.77	.79
Audiovisual	.99	.82

Word Recognition Memory Performance

Table 2 presents an overview of word recognition performance. To obtain the most complete picture of recognition performance, we used three indices of accuracy: hits (*old* responses to studied words), false alarms (FAs; *old* responses to new words), and recognition scores (hits − FAs). For each measure, the values for familiar words always exceeded the values for unfamiliar words. This shows that the participants were more likely to respond *old* to a word presented in the context of a familiar voice. Each dependent measure was analyzed separately using ANOVAs with the between-subjects factors of training condition (A vs. AV) and word recognition condition (A vs. AV) and the within-subjects factor of talker familiarity (familiar talkers vs. unfamiliar talkers). In all cases, the effect of training condition was not significant, nor did it interact with any other factor. Bearing in mind that all the participants were equated on voice recognition (e.g., the same criterion on the generalization task) prior to the word recognition test, this result is not surprising. The data from the two training conditions were pooled, and subsequent analyses were based on the combined data from both groups.

The hit data (see Figure 3) were analyzed using an ANOVA with talker familiarity and word recognition condition as factors. The most important finding was that target words spoken by familiar talkers were recognized more accurately than target words spoken by unfamiliar talkers. The overall hit rate was 74% for words spoken by familiar talkers and 62% for words spoken by unfamiliar talkers, and this difference was highly significant [$F(1,38) = 42.74, MS_e = 0.007, p < .0001$]. Word recognition accuracy was only marginally higher in the auditory word recognition test condition [$F(1,38) = 3.38$, $MS_e = 0.05$, $p = .07$]. The interaction between recognition condition and talker familiarity was not sig-

Table 2
Mean Proportion of Correct Word Recognition (Hit Rate), False Alarm Rate, and Recognition Score in Each Experiment as a Function of Training Condition, Word Recognition Test Condition, and Word Context

	Word Recognition Performance					
	Familiar Talker Context			Unfamiliar Talker Context		
Training × Word Test	$P(C)_F$	FA_F	$H-F_F$	$P(C)_U$	FA_U	$H-F_U$
A × A	.77	.29	.49	.68	.25	.43
A × AV	.69	.25	.44	.57	.21	.36
AV × A	.78	.29	.49	.68	.25	.43
AV × AV	.74	.36	.38	.56	.24	.32
Experiment totals	.75	.30	.45	.63	.24	.39

Note—$P(C)$, proportion of correct *old* responses (hits); FA, false alarm; H−F, hits minus FAs.

nificant ($p = .15$). Planned comparisons confirmed that hit rates were significantly higher for familiar items in each test condition [A, $F(1,19) = 9.38, MS_e = 0.01, p < .006$; AV, $F(1,19) = 51.33, MS_e = 0.004, p < .0001$].

The FA data indicate that new, unstudied words spoken by a well-known talker were also perceived as more familiar and, consequently, were more likely to be judged (incorrectly) as *old*. This difference was significant [$F(1,36) = 4.68, MS_e = 0.02, p = .04$]. The effect of word recognition condition (e.g., A vs. AV) was not significant, and this factor did not interact with talker familiarity.

To control for differences in FA rates, we conducted an analysis using recognition scores (i.e., hits minus FAs for familiar items, hits minus FAs for unfamiliar items) as a dependent variable. The analysis yielded the same pattern of results as those obtained from the hit analysis. In particular, recognition scores were significantly higher for familiar items [$F(1,38) = 4.53, MS_e = 0.019, p < .05$]. The main effect of word recognition test condition and the word recognition test × talker familiarity interaction were not significant.

In summary, we obtained evidence that the participants responded differently to the familiar and the unfamiliar items. Words in the familiar talker context condition produced a greater proportion of correct *old* responses. However, there was also a tendency for new, unstudied words to be recognized as *old* if a familiar talker spoke them. With respect to modality effects, the experiment did not reveal any reliable differences in word recognition as a function of training modality or test modality.

DISCUSSION

The present experiment was designed to explore how speakers are learned and how their utterances are remembered. To summarize our main results, (1) the perceptual learning data showed that the opportunity to see a talker's articulating face during training improved the perceptual encoding of the talker's voice. We know of no other empirical demonstrations of the selective benefits of visible speaker information on the perceptual learning of voices and, thus, consider this to be the most important feature of our results. (2) Long-term memory for spoken words was enhanced by talker familiarity, and the magnitude of this effect was similar for A and AV word recognition conditions.

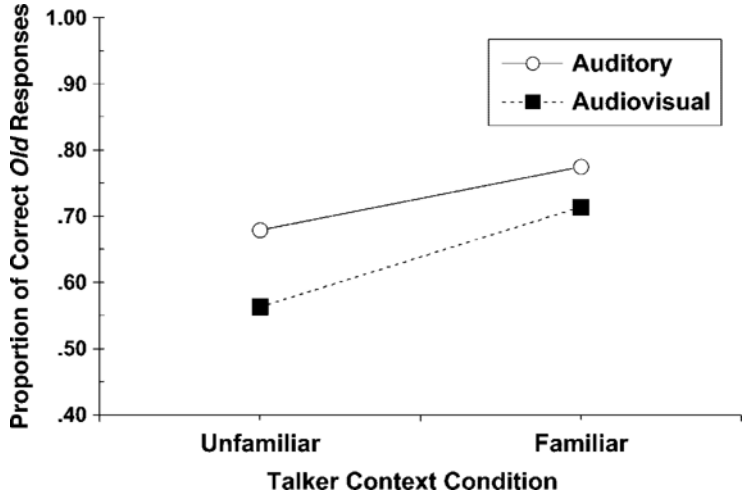

Figure 3. Mean proportions of correct *old* responses on the word recognition task for target words spoken by unfamiliar and familiar talkers.

The Perceptual Learning of Voices

As was discussed in the introduction, much recent interest has centered on the notion that auditory and visual speaker recognition might be based on features similar to those used to recognize audiovisual speech and, hence, show similar performance characteristics. Accordingly, we had predicted that learning to identify a voice might be enhanced by the addition of facial speech, and we tested this in the training phase of the experiment.

We found that the presence of visual speaker information improved voice learning. Listeners in the AV training condition were better at classifying voices from novel auditory words, relative to participants in the A condition. The generalization data extended the auditory perceptual-learning results reported by Nygaard et al. (1994) and Sheffert et al. (2002) by demonstrating that familiar voice recognition is mediated by representations that are sufficiently abstract to allow generalization over significant changes in phonological properties and sensory modality.

Our results differ from other reports that have shown either negative effects or null effects of face information on voice encoding (Armstrong & McKelvie, 1996; Cook & Wilding, 1997, 2001; Legge et al., 1984; Yarmey, 1993). Despite a number of procedural differences between the present study and previous studies that limit direct comparisons, including the materials and the training and testing tasks, we suspect that the amount of exposure to the talkers prior to the voice recognition test was the main reason for the discrepancy. For example, the participants in Cook and Wilding's (1997, 2001) "face overshadowing" studies had far less experience with the faces (one presentation of a sentence). In fact, the face interfered very little after three presentations of a sentence. Certainly, the findings described by Walker et al. (1995) are consistent with this idea, by showing that face information interfered less with auditory speech perception when the faces were highly familiar.

Alternatively, the AV training results might simply have reflected the observers' propensity to attend to faces, regardless of whether or not the talker's mouth movements were part of the visual display. To evaluate this possibility, we conducted a control experiment that was identical to the AV training condition, except that the participants were not able to see the talker's mouth (henceforth, the AV–mouth condition). This was accomplished by covering the portion of the video screen that displayed the mouth region (the area just below the nose, including the lower cheeks and the entire mouth and jaw).[1] These AV–mouth talkers were readily identified from other aspects of the face, such as the eyes, the nose, and the hairstyle. Indeed, the results from the training portion of this control study ($n = 20$) were parallel with those from the full-face AV condition (e.g., same high level of AV talker identification).

The key finding was that obscuring the visible speaker's mouth had a substantial effect on the participants' ability to learn the talker's voice. In fact, the outcome in terms of subsequent voice recognition was virtually indistinguishable from that for the auditory-alone training condition. For example, by the second day of training, only 35% of the participants in the AV–mouth group had reached criterion (vs. 25% of the participants in the A group and 65% in the AV group), and 4 participants required at least four sessions. Overall, the participants in the AV–no-mouth condition learned the voices in an average of 2.9 sessions, which was significantly longer than the time for the participants in the AV condition (2.1 days) but not reliably different from that for the participants in the A condition (3.2 days). In addition, the generalization scores (65%, 69%, and 76% for Days 1–3, respectively) did not differ from those of the participants in the A condition (all $ps > .4$).

This is preliminary evidence that the learning-promoting properties of visible speaker information obtained in the AV training condition were not the result of facial identity information. It is important to note that this manipulation removed both dynamic visible speech and static structural features of the mouth, leaving open the possibility that static mouth features were the locus of the AV benefit obtained in the present experiment.

Alternatively, our participants might simply have associated auditory information with visual information. Although we cannot rule this possibility out entirely, the data from the AV–no-mouth condition make this less plausible, because these participants easily could have associated visible and auditory features (and quite likely did so) but the association did not improve voice learning. Stronger evidence against the notion that cross-modal links are based on arbitrary associations can be found in Fowler and Dekle (1991). Although they measured phonetic perception, their results are pertinent to talker perception (at least to the extent that speech processing and talker processing exploit phonetic features). Fowler and Dekle obtained cross-modal integration effects by pairing acoustic syllables with mouth syllables that were perceived haptically (by touch, using the Tadoma method of speech reading). Although the surface forms of these physical signals differ, each specifies the same physical event, and therefore, the signals readily integrate. In contrast, visual orthographic syllables and acoustic speech do not integrate, despite the fact that the latter pairing represents a well-learned association. In keeping with these observations, we speculate that the null effects of face information on voice learning reported by Legge et al. (1984) and others might have been derived from the lack of intermodal invariants brought about through the use of arbitrary voice + face image pairings or tasks that emphasized nonlinguistic aspects of the talker's face (appearance/identity instead of mouth shape; cf. Schweinberger & Soukup, 1998).

If one accepts that performance for items in the AV condition is not merely a consequence of facial identify or arbitrary AV associations, it follows that some other factor is at work. We favor the view outlined in the introduction—namely, that visible speech gestures can provide addi-

tional information about a talker's idiosyncratic speaking style and that these features are compatible with auditory talker-specific features (Remez et al., 1997; Rosenblum et al., 2002). Furthermore, the same sort of articulatory features used to perceive audiovisual speech (Fowler, 1986; Liberman & Mattingly, 1985; Rosenblum et al., 2002) might eventually prove to be crucial for linking auditory and visual talker-specific information.

Talker Familiarity and Word Recognition

In the second phase of our study, we examined the extent to which familiarity with a talker would interact with long-term memory for spoken words. We predicted that spoken word memory would preserve attributes of the speaker who presented the word—the voice, face, and identity—and that these attributes would interact with word retrieval. Here, we focused on the effects of talker information as a type of context cue for spoken words, where talker context differed in preexperimental familiarity. For example, a word spoken by a familiar talker might be more memorable than a word spoken by an unfamiliar talker. We examined this possibility by using a standard *old/new* word recognition task. Half of the words were spoken by the five familiar talkers (from the training phase), and half were spoken by five new talkers. Using this method, we showed that talker familiarity affects word recognition.

Whenever a test word was spoken by a familiar talker, it was more likely to receive an *old* judgment than a test word spoken by a new talker. This increased the hit rate and the FA rate. Although the FA rate was higher for familiar context items, it did not completely offset the higher hit rate for familiar talker context words. In other words, the participants do not appear to have resorted to a strategy whereby they responded *old* to any word in a known voice. Interference among similar items arguably played a key role. Because familiar talkers were associated with many other words in memory that shared talker context features (accrued during the training procedure), they tended to activate more memory traces than did unfamiliar talkers. Consequently, an *old* test word spoken by a familiar talker might not have received enough unique activation to elicit recognition, and a similar distractor might have been incorrectly recognized (producing an FA). This sort of contextual interference was less likely to occur in the unfamiliar talker condition. In addition, it might have been difficult for some of the participants to switch from the training task (where only talker features were relevant) to the word recognition task (where talker features were irrelevant and not at all diagnostic of word identity). This could also explain why performance in the AV condition was slightly lower: Overriding task-irrelevant information was more difficult when the irrelevant stimulus was as salient as a face.

The integral nature of talker and linguistic perception raised the possibility that talker context information might be closely linked to word/item information and, therefore, function as a strong context cue (Maddox & Estes, 1997; Sheffert & Shiffrin, 2003). For example, Murnane and Phelps (1995) varied the environmental context of printed test words across study and test words. They found that reinstating context features that were integral to the interpretation of the item improved participants' ability to recognize targets without also inflating the FA rate. In our study, the joint presence of higher hits and higher FAs suggested that talker familiarity produced only modest improvements in the participants' ability to distinguish between *old* targets and *new* distractors. The talker familiarity effects were more characteristic of incidental context.

Familiarity acquired via extensive training using feedback could be critical for showing effects of talker familiarity. For example, Palmeri et al. (1993) varied voice (same or different across study and test) and talker variability (the number of talkers within a list). They found that word recognition performance was enhanced in *same-voice* trials, which is evidence that a word's context includes voice features. A question to ask, then, is whether the amount of experience with a talker's voice had an impact on word memory. The *2-voice* list provided listeners with over 170 opportunities to become familiar with each voice, whereas the *20-voice* list provided fewer than 20 such opportunities. Surprisingly, the number of talkers did not affect word memory accuracy or the magnitude of the same-voice advantage. It is possible that the nature of the talker knowledge acquired during the course of a list may be qualitatively different from that acquired in a training procedure.

More generally, these results suggest that the effects of familiarity observed on lower level perceptual tasks may not operate in the same way on tasks that tap long-term memory. The source and generality of talker familiarity effects are aspects of the talker–word contingency issue that have so far received very little attention but could have interesting theoretical implications for models of speech processing and memory.

REFERENCES

ARMSTRONG, H. A., & MCKELVIE, S. J. (1996). The effect of face context on recognition memory for voices. *Journal of Experimental Psychology: General*, **123**, 259-270.

BASSILI, J. N. (1979). Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face. *Journal of Personality & Social Psychology*, **37**, 2049-2058.

BERNSTEIN, L. E., DEMOREST, M. E., & TUCKER, P. E. (2000). Speech perception without hearing. *Perception & Psychophysics*, **62**, 233-252.

BRICKER, P. D., & PRUZANSKY, S. (1976). Speaker recognition. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics* (pp. 295-326). New York: Academic Press.

BRUCE, V. (1988). *Recognising faces*. Hove, U.K.: Erlbaum.

BRUCE, V., & VALENTINE, T. (1988). When a nod's as good as a wink: The role of dynamic information in facial recognition. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues. Vol. 1: Memory in everyday life* (pp. 169-174). New York: Wiley.

CHRISTIE, F., & BRUCE, V. (1998). The role of dynamic information in the recognition of unfamiliar faces. *Memory & Cognition*, **26**, 780-790.

COOK, S. A., & WILDING, J. M. (1997). Earwitness testimony: 2. Voices, faces and context. *Applied Cognitive Psychology*, **11**, 527-541.

COOK, S. A., & WILDING, J. M. (2001). Earwitness testimony: Effects of exposure and attention on the face overshadowing effect. *British Journal of Psychology*, **92**, 617-629.

CRAIK, F. I. M., & KIRSNER, K. (1974). The effect of speaker's voice on word recognition. *Quarterly Journal of Experimental Psychology*, **26**, 274-284.

ELLIS, H. D. (1986). Processes underlying face recognition. In R. Bruyer (Ed.), *The neuropsychology of face perception and facial expression*. Hillsdale, NJ: Erlbaum.

ERBER, N. P. (1969). Interaction of audition and vision in the recognition of oral speech stimuli. *Journal of Speech & Hearing Research*, **12**, 423-425.

FELLOWES, J. M., REMEZ, R. E., & RUBIN, P. E. (1997). Perceiving the sex and identity of a talker without natural vocal timbre. *Perception & Psychophysics*, **59**, 839-849.

FOWLER, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics* **14**, 3-28.

FOWLER, C. A., & DEKLE, D. J. (1991). Listening with eye and hand: Cross-modal contributions to speech perception. *Journal of Experimental Psychology: Human Perception & Performance*, **17**, 816-828.

GOLDINGER, S. D. (1996). Words and voices: Implicit and explicit memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 1166-1183.

GOLDINGER, S. D. (1998). Echoes of echoes: An episodic theory of lexical access. *Psychological Review*, **105**, 251-277.

GREEN, K. P., TOMIAK, G. R., & KUHL, P. K. (1997). The encoding of rate and talker information during phonetic perception. *Perception & Psychophysics*, **59**, 675-692.

HALLE, M. (1985). Speculations about the representation of words in memory. In V. A. Fromkin (Ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged* (pp. 101-114). New York: Academic Press.

LANDER, K., CHRISTIE, F., & BRUCE, V. (1999). The role of movement in the recognition of famous faces. *Memory & Cognition*, **27**, 974-985.

LANSING, C. R., & MCCONKIE, G. W. (1999). Attention to facial regions in segmental and prosodic visual speech perception tasks. *Journal of Speech, Language, & Hearing Research*, **42**, 526-539.

LANSING, C. R., & MCCONKIE, G. W. (2003). Word identification and eye fixation locations in visual and visual-plus-auditory presentations of spoken sentences. *Perception & Psychophysics*, **65**, 536-552.

LAVER, J., & TRUDGILL, P. (1979). Phonetic and linguistic markers in speech. In K. R. Scherer & H. Giles (Eds.), *Social markers in speech* (pp. 1-31). Cambridge: Cambridge University Press.

LEGGE, G. E., GROSMANN, C., & PIEPER, C. M. (1984). Learning unfamiliar voices. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 298-303.

LIBERMAN, A. M. (1982). On finding that speech is special. *American Psychologist*, **37**, 148-167.

LIBERMAN, A. M., & MATTINGLY, I. G. (1985). The motor theory of speech perception revised. *Cognition*, **21**, 1-36.

MADDOX, W. T., & ESTES, W. K. (1997). Direct and indirect stimulus-frequency effects in recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 539-559.

MCGURK, H., & MACDONALD, J. W. (1976). Hearing lips and seeing voices. *Nature*, **264**, 746-748.

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

MUNHALL, K. G., & VATIKIOTIS-BATESON, E. (1998). The moving face during speech communication. In R. Campbell, B. Dodd, & D. Burnham (Eds.), *Hearing by eye II: Advances in the psychology of speechreading and auditory–visual speech* (pp. 123-139). Hove, U.K.: Psychology Press.

MURNANE, K., & PHELPS, M. P. (1995). Effects of changes in relative cue strength on context-dependent recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 158-172.

NYGAARD, L. C., & PISONI, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, **60**, 355-376.

NYGAARD, L. C., SOMMERS, M. S., & PISONI, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, **5**, 42-46.

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

PISONI, D. B. (1996). Some thoughts on "normalization" in speech perception. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 9-32). San Diego: Academic Press.

PREMINGER, J. E., LIN, H.-B., PAYEN, M., & LEVITT, H. (1998). Selective visual masking in speechreading. *Journal of Speech, Language, & Hearing Research*, **41**, 564-575.

REISBERG, D., MCLEAN, J., & GOLDFIELD, A. (1987). Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli. In R. Campbell & B. Dodd (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 97-114). Hillsdale, NJ: Erlbaum.

REMEZ, R. E., FELLOWES, J. M., & RUBIN, P. E. (1997). Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception & Performance*, **23**, 651-666.

ROSENBLUM, L. D., & SALDAÑA, H. M. (1998). Time-varying information for speech perception. In R. Campbell, B. Dodd, & D. Burnham (Eds.), *Hearing by eye II: Advances in the psychology of speechreading and auditory–visual speech* (pp. 61-81). Hove, U.K.: Psychology Press.

ROSENBLUM, L. D., YAKEL, D. A., BASEER, N., PANCHAL, A., NODARSE, B. C., & NIEHUS, R. P. (2002). Visual speech information for face recognition. *Perception & Psychophysics*, **64**, 220-229.

SALDAÑA, H. M., NYGAARD, L. C., & PISONI, D. B. (1996). Encoding of visual speaker attributes and recognition memory for spoken words. In D. Stork & M. E. Hennecke (Eds.), *Speechreading by man and machine: Models, systems, and applications* (1995 NATO ASI Workshop, pp. 275-281). Berlin: Springer-Verlag.

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

SCHWEINBERGER, S. R., & SOUKUP, G. R. (1998). Asymmetric relationships among perceptions of facial identity, emotion, and facial speech. *Journal of Experimental Psychology: Human Perception & Performance*, **24**, 1748-1765.

SHEFFERT, S. M. (1998a). Contributions of surface and conceptual information to recognition memory. *Perception & Psychophysics*, **60**, 1141-1152.

SHEFFERT, S. M. (1998b). Voice-specificity effects on auditory word priming. *Memory & Cognition*, **26**, 591-598.

SHEFFERT, S. M., & FOWLER, C. A. (1995). The effects of voice and visible speaker change on memory for spoken words. *Journal of Memory & Language*, **34**, 665-685.

SHEFFERT, S. M., LACHS, L., & HERNÁNDEZ, L. (1996–1997). The Hoosier Audiovisual Multi-Talker Computer Database. In *Research on Spoken Language Processing* (Progress Rep. No. 21, pp. 578-583). Bloomington: Indiana University, Speech Research Laboratory.

SHEFFERT, S. M., PISONI, D. B., FELLOWES, J. M., & REMEZ, R. E. (2002). Learning to recognize talkers from natural, sinewave, and reverse speech samples. *Journal of Experimental Psychology: Human Perception & Performance*, **28**, 1447-1469.

SHEFFERT, S. M., & SHIFFRIN, R. M. (2003). Auditory registration without learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 10-21.

SHEPARD, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning & Verbal Behavior*, **6**, 156-163.

SUMBY, W. H., & POLLACK, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, **26**, 212-215.

SUMMERFIELD, A. Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 3-51). Hillsdale, NJ: Erlbaum.

WALKER, S., BRUCE, V., & O'MALLEY, C. (1995). Facial identity and facial speech processing: Familiar faces and voices in the McGurk effect. *Perception & Psychophysics*, **57**, 1124-1133.

WOODHEAD, M. M., BADDELEY, A. D., & SIMMONDS, D. C. (1979). On training people to recognize faces. *Ergonomics*, **22**, 333-343.

YAKEL, D. A., ROSENBLUM, L. D., & FORTIER, M. A. (2000). Effects of talker variability on speechreading. *Perception & Psychophysics*, **62**, 1405-1412.

YARMEY, A. D. (1986). Verbal, visual, and voice identification of a rape suspect under different levels of illumination. *Journal of Applied Psychology*, **71**, 363-370.

YARMEY, A. D. (1993). Stereotypes and recognition memory for faces and voices of good guys and bad guys. *Applied Cognitive Psychology*, **7**, 419-431.

NOTE

1. We acknowledge that our manipulation probably did not eliminate *all* visual phonetic properties. Indeed, this would be impossible to do without covering most of the face, given the fact that the motions of the lips and the jaw can produce simultaneous changes in more peripheral regions of the face, such as the upper cheeks and the eyebrows (Lansing & McConkie, 1999, 2003; Munhall & Vatikiotis-Bateson, 1998; Preminger, Lin, Payen, & Levitt, 1998). Nevertheless, the few visual linguistic features preserved in our displays were insufficient for identifying individual visemes or words.

APPENDIX
Stimulus Words

back	cite	girl	leave	pool	sign
badge	coat	give	leg	pup	size
bait	cod	goal	less	push	soak
bake	comb	goat	light	put	soil
ball	con	gone	loan	rain	south
ban	cone	gown	long	raise	tack
bang	cool	guide	loose	rake	take
bar	cot	gum	love	rang	talk
base	curve	hack	luck	rat	tan
bat	dare	hag	mace	rate	tape
beach	date	hall	mail	reach	taught
bead	dawn	ham	main	read	teach
beak	deal	heat	mall	real	team
bean	death	hen	map	rhyme	teeth
bed	debt	hick	mat	rich	thick
been	deep	hid	meat	rim	thing
boar	den	hike	mid	ring	thought
boat	dig	hole	mile	rise	thumb
bone	dirt	hood	mine	road	tile
boot	dog	hoot	mitt	roar	tin
both	doom	hope	moan	rock	ton
bug	doubt	house	mole	roof	toot
bum	down	hung	mood	root	top
bun	dune	jack	mouse	rose	town
cage	fade	job	move	rote	vice
cake	fair	join	neck	rough	vote
call	faith	judge	net	rule	wade
cane	fall	keep	noise	rum	wail
car	fan	kill	nose	sad	wait
case	fat	kin	note	sail	wash
cat	faze	king	one	sane	watch
caught	fear	kiss	pace	scene	weak
cause	feel	kit	pad	seat	wed
cave	fig	knead	page	seek	white
chain	fin	knob	pail	serve	wick
chair	fine	knot	pain	shade	wife
chat	firm	known	pan	shape	work
check	fit	lace	pat	shed	wrong
cheer	five	lake	path	sheet	young
cheese	food	lame	pawn	shell	
chief	fool	late	peace	ship	
chin	gain	lawn	pen	shop	
chore	gas	league	pet	shore	
church	gave	learn	pick	sick	

(Manuscript received August 5, 2002;
revision accepted for publication August 9, 2003.)