109 6

Journal of Experimental Psychology:
Human Perception and Performance
1997, Vol. 23, No. 3, 651–666

Talker Identification Based on Phonetic Information

Robert E. Remez and Jennifer M. Fellowes
Barnard College

Philip E. Rubin
Haskins Laboratories

Accounts of the identification of words and talkers commonly rely on different acoustic properties. To identify a word, a perceiver discards acoustic aspects of an utterance that are talker specific, forming an abstract representation of the linguistic message with which to probe a mental lexicon. To identify a talker, a perceiver discards acoustic aspects of an utterance specific to particular phonemes, creating a representation of voice quality with which to search for familiar talkers in long-term memory. In 3 experiments, sinewave replicas of natural speech sampled from 10 talkers eliminated natural voice quality while preserving idiosyncratic phonetic variation. Listeners identified the sinewave talkers without recourse to acoustic attributes of natural voice quality. This finding supports a revised description of speech perception in which the phonetic properties of utterances serve to identify both words and talkers.

When a familiar voice speaks familiar words, a listener identifies both talker and message. Although impressions of a talker and a message are concurrent, it is commonly assumed that these two facets of speech perception derive from different auditory attributes. The properties of speech that distinguish consonants and vowels—the detailed pattern of vocal resonances associated with specific articulatory acts (e.g., Zue & Schwartz, 1980)—are unlike the acoustic properties that distinguish voices—often taken to be the acoustic correlates of vocal timbre, such as the range of frequency variation of glottal pulsing or the shape of the spectrum generated at the larynx (Bricker & Pruzansky, 1976; Hecker, 1971). It is self-evidently plausible to explain the perception of linguistic properties of utterances by appealing to acoustic constituents of speech that are consistent across individuals and unique to none and, conversely, to describe the perception of a talker's identity by appealing to acoustic properties of speech that are unique to an individual's voice and therefore are nonlinguistic. In this conceptualization, each kind of perceptual attribute derives from a different sensory cause.

Despite the credible segregation of perceptual paths leading to word recognition and talker identification, some studies of spoken words have undermined the hypothesis of functional independence. In these studies, investigators have observed an influence of nonlinguistic attributes of specific utterances on lexical decision, identification, and memory (Church & Schacter, 1994; Nygaard, Sommers, & Pisoni, 1994; Palmeri, Goldinger, & Pisoni, 1993). For example, performance on an implicit identification test was impaired relative to recognition performance when two occurrences of a word, first as a prime and then as a test item, differed in critical acoustic characteristics (Church & Schacter, 1994). Such evidence supports a proposal that utterance-specific acoustic attributes that may be unique to individual talkers moderate the perception of linguistic properties of speech. Although this suggestion is consistent with facts that preclude independence, the basis of the contingency nonetheless remains obscure and is the topic of this article.

We describe three experiments in which we found that phonetic attributes are a potential common code for lexical and individual identification. A *phonetic* grain of analysis pertains to the articulatory and perceptual effects that realize a specific utterance of a word, in contrast to the subordinate *auditory* impressions of the timbre of a complex speech spectrum and also in contrast to the superordinate *phonemic* representation of the sound pattern of a word that is abstracted from all of its specific instances. The evidence that supports our conclusion comes from tests that assessed the ability of listeners to recognize familiar voices solely from phonetic attributes when presented with signals that lacked acoustic correlates of natural voice quality. To present the case that the phonetic properties of utterances alone are useful for identifying talkers, we first review two aspects of

Robert E. Remez and Jennifer M. Fellowes, Department of Psychology, Barnard College; Philip E. Rubin, Haskins Laboratories, New Haven, Connecticut.

This research was supported by Grants DC00308 and HD01994 from the National Institutes of Health. During the development of this study, Jennifer M. Fellowes was a Research Intern of the Howard Hughes Biomedical Science Pipeline Project, a program of the Undergraduate Biological Sciences Education Initiative of the Howard Hughes Medical Institute.

We gratefully acknowledge the timely counsel and wise assistance of Arthur Abramson, Eva Blumenthal, Lila Braine, Jessica Lang, Lynne Nygaard, Jennifer Pardo, David Pisoni, Bella Schanzer, and Michael Studdert-Kennedy and the cooperation of the staff of Haskins Laboratories.

Correspondence concerning this article should be addressed to Robert E. Remez, Department of Psychology, Barnard College, 3009 Broadway, New York, New York 10027-6598. Electronic mail may be sent via Internet to remez@paradise.barnard.columbia.edu.

the contemporary account: the perception of words and the perception of voices.

Perception of Words

How does a listener know what a talker is saying? In contemporary accounts of comprehension, perceiving the meaning of a talker's message depends on recognizing the words. Accordingly, a listener is said to evaluate a talker's utterance by comparing its constituents to remembered words in a process of lexical access (Cutler, 1989; Forster, 1976; Marslen-Wilson, 1984). Although several different descriptions of lexical access appear in recent studies (Cutler & Norris, 1988; Luce, Pisoni, & Goldinger, 1990; McClelland & Rumelhart, 1981; see Lively, Pisoni, & Goldinger, 1994, for a review), a common assumption has shaped the account of contact between novel instances and word candidates stored in memory: Fine-grain representations are phonemic (see Halle, 1985, and Segui, 1984, for a discussion of hierarchical linguistic representation in word recognition). To use a phonemic address for recognizing a spoken word, a listener must analyze an unidentified speech signal to eliminate attributes unique to any specific utterance. Once a phonemic description is achieved, an unknown word is cast in the abstract form by which words within the mental lexicon are distinguished from one another, and the perceiver uses this representation as bait with which to fish for a known word stored in memory. In short, lexical processes represent linguistic properties common to all instances of a word, not the unique features of specific instances.

Accordingly, a lexicon that uses phoneme representations cannot represent differences among talkers who use the same words. Therefore, while the lexical system is responsible for identifying what is said, a separate faculty is typically held responsible for identifying who said it. In other words, talker recognition and word comprehension are dissociated in contemporary accounts (Bricker & Pruzansky, 1976; Hecker, 1971; Hollien & Klepper, 1984).

Perception of a Talker's Identity

How does the listener know who said the words? Much research has aimed to designate distinctive attributes of talkers, with investigations distributed across three lines of study: (a) talker recognition by visual inspection of spectrograms; (b) automatic recognition by computational analysis; and, (c) tests of identification of talkers by listening. Many and varied acoustic properties have been implicated, among them formant range (Fant, 1966; Sambur, 1975), nasal spectrum (Glenn & Kleiner, 1968; Sambur, 1975; Su, Li, & Fu, 1974), laryngeal spectrum (Monsen & Engebretson, 1977), long-term spectrum (Furui, 1978), pitch variability (Atal, 1972; Jassem, 1971; van Dommelen, 1987, 1990), spectral slope (Matsumoto, Hiki, Sone, & Nimura, 1973), registration of intensity (Lummis, 1973), and the metrics of the vowel space (Endres, Bambach, & Flösser, 1971; Goldstein, 1976). Perceptually, these acoustic properties presumably evoke aspects of voice quality, such as strength, melodiousness, or forcefulness (Carterette & Barnebey, 1975; Gelfer, 1988), though the means by which a diverse set of acoustic characteristics evokes complex impressions of vocal timbre is not understood; neither are the dimensions of the perceptual encoding of voice quality well specified. Whether the primary perceptual representation is auditory, anatomical, gestural, regional, or personal, a customary dynamic of recognition is presumed in which an unknown signal is assessed and its attributes are compared to candidates within a listener's memory of talkers (Bricker & Pruzansky, 1976; Hecker, 1971; Hollien & Klepper, 1984).

A consistent theme recurs in explaining recognition by human and machine. In this literature, information about a talker's identity is defined as an extra message apart from the linguistic constituents. The search for causes of the perception of personal identity, therefore, has turned consistently toward acoustic properties other than those that evoke the perception of phonemic contrasts. As Bricker and Pruzansky (1976) stated, the goal has been to explain the recognition of talkers by virtue of acoustic attributes that are reliably manifested in each voice yet which exhibit little intratalker variability, a requirement that is hardly met by highly variable and multiply cued linguistic attributes (e.g., the well-studied *voicing* contrast; see Lisker, 1978; Lisker & Abramson, 1964).

Dissociation of Word and Talker Identification

Some investigations have directly corroborated this prevailing dissociation of voice recognition and word comprehension. For instance, several tests have shown that information about a talker's identity is available from reversed speech, a condition that aimed to eliminate linguistic information while preserving the spectral attributes that typify a voice (Bricker & Pruzansky, 1966; Clarke, Becker, & Nixon, 1966; Van Lancker & Kreiman, 1985; Williams, 1964). Filtering has also been used to prevent linguistic properties of speech from contributing to voice recognition; the residue of acoustic structure can be used to identify talkers (Compton, 1963; Pollack, Pickett, & Sumby, 1954). Whispered speech, in contrast, often allows intelligibility to persist while impairing identifiability of a talker (Pollack et al., 1954; Williams, 1964). Overall, findings of the persistence of one function while the other function deteriorates have been taken as evidence of perceptual independence and of the reliance of each faculty on a different set of acoustic attributes.

Converging evidence of the functional dissociation of voice identification and word comprehension comes from the neuropsychological literature. Phonagnosia is a disorder that selectively impairs voice recognition while sparing speech perception (Van Lancker, Cummings, Kreiman, & Dobkin, 1988). A phonagnosic patient is able to comprehend an utterance but unable to recognize the familiar voice producing it. Recall that aphasia, a disorder of linguistic comprehension, is not typically accompanied by loss of the ability to identify a talker whose speech is not understood. Contrasting these deficits, we see that even though all of the relevant acoustic attributes are available, selective impairment of the perception of linguistic or personal attributes occurs as a function of the site of brain injury. Accordingly, these perceptual failures have been interpreted to indicate different and independent perceptual processes devoted to linguistic and personal attributes, regardless of their acoustic or auditory bases.

Exceptions to the Common View

Despite consensus about the independence of word and talker identification, several reports are difficult to explain according to this hypothesis. The first and best known of these is a classic study by Pollack et al. (1954), who found that a listener's ability to identify a talker depended on the phonemic variety of the speech sample, among other factors (also, see Bricker & Pruzansky, 1966; Coleman, 1973; Hollien, Majewski, & Doherty, 1982; Ladefoged & Ladefoged, 1980; Mullenix & Pisoni, 1990; see Atal, 1974, for a parallel case of automatic recognition). More recent studies (Church & Schacter, 1994; Goldinger, Pisoni, & Logan, 1991; Nygaard et al., 1994; Palmeri et al., 1993; Schacter & Church, 1992) suggest that experience with a talker's voice does affect the perception of spoken words.

Consider two clear cases. In one, Nygaard et al. (1994) trained listeners to name a set of 10 voices producing monosyllabic words. Listeners learned the characteristics of talkers well enough to identify the voices from new utterances of the word set that had been used in training and from a second word set that was used in a more stringent test under conditions of greater uncertainty. After training, listeners were asked to identify yet a third and different set of words masked by noise; some of the words were produced by talkers whom the listeners had learned to identify, whereas other words were produced by unfamiliar talkers. Listeners identified words produced by familiar talkers better than words produced by unfamiliar talkers. Contrary to the presumption of the independence of talker and word recognition, these experimenters hypothesized that familiarity with a vocal source facilitated recognition of words, although the precise cause of the benefit of familiarity was not identified (cf. Nygaard & Kalish, 1994).

In another study, Church and Schacter (1994) sought an explanation for a similar contingency. In five experiments aiming to identify instance-specific acoustic characteristics that affect the identification of words, they used a testing paradigm that assessed implicit memory. When critical acoustic characteristics differed between two occurrences of a word, once as a prime and again as a test item, performance on implicit identification tests was impaired; no similar influence of acoustic characteristics was observed in parallel tests of recognition performance. The effective acoustic manipulations included (a) presenting a word during the study phase of the test spoken in one voice and reprising the word in a different voice during the test phase; (b) varying the paralinguistic qualities of the speech samples (*happy* at prime and *sad* at test, and vice versa); and, (c) altering the fundamental frequency of phonation through speech synthesis. No effect was observed of a manipulation of gross signal power, an acoustically huge transformation with no evident consequences for memory. Like Nygaard et al. (1994), Church and Schacter described their findings as a contingency of lexical processes on extralinguistic attributes—in other words, as a case contrary to the appealing ideal of abstract phonemic addressing in the lexicon. However, they sketched a preliminary mechanism to explain such effects, one in which vast cognitive resources spanning both cerebral hemispheres apply linguistic and nonlinguistic attributes of speech to the formation of declarative and nondeclarative representations alike. By taking a broad perspective, we see that the pertinent findings and models are useful heuristically if not predictively and that the causes of the contingency of words and voices remain obscure. Nonetheless, the clues provided by Church and Schacter, Nygaard et al. (1994), and Pollack et al. (1954) are important for what they suggest. Namely, this particular contingency, at its simplest, indicates that a single form of representation may underlie perception of the disparate attributes of words and talkers.

The Present Experiments

A likely prospect for a common code is the phonetic level of analysis of speech sounds. The phonetic component of the speech chain occurs apart from the conceptualization of the message and the activation of its syntactic and lexical vehicles in speech production; it includes the selection, sequencing, and execution of the expressive postures and gestures of the organs of articulation and their acoustic consequences (Catford, 1988; also, see Fowler, Rubin, Remez, & Turvey, 1980). In perception, phonetic attributes correspond in part to the apprehension of specific speech sounds, in contrast to the abstracted and general phonemic form of an utterance. A phonetic description is required, for instance, to represent whether a talker said [sʊkʰjʊɹɪtʰi], [səkʰjʊɹəɾi], or [skjʊɹɾi] in realizing the word *security.*[1] Phonetic attributes would be useful for identifying talkers if variation in the phonetic realization of words is particular to individual talkers, if these properties of speech are durable

[1] The phonetic manifestations of a single word vary widely over instances, though not only by chance. Linguists identify several principles of variation that converge in the production of a particular instance, illustrated in part in the case of the word *security.* Differences in articulatory rate are often expressed by different allophones, such that rapid and slow forms of the same words use different consonants and vowels, and not simply by briefer or longer versions of the same segmental constituents. The phonetic quality and variety of vowels vary also as a function of style, alternating between formal and casual modes (Labov, 1972; Picheny, Durlach, & Braida, 1986). Dialects or accents express the same words in phonetic forms that differ regionally and socially across talkers (Labov, 1986). Accordingly, the phonetic grain of utterances constitutes a rich source of talker-specific information, one that is linguistically governed, independent of voice quality, and independent of the messages conveyed by speech.

perceptually, and if the perceiver is disposed to apply linguistic characterizations as well as representations of voice quality in detecting and remembering differences among familiar talkers.

Although there is general agreement that phonetic attributes play a role in the recognition of spoken words (Lively et al., 1994), there is no direct evidence that phonetic segmental constituents alone—the specific consonant and vowel allophones manifest in an utterance—can evoke an impression of a particular talker. In our experiments we sought to provide a test by exploiting an acoustic technique that preserves the phonetic properties of speech while discarding the acoustic attributes of voice quality and intonation. This method, sinewave replication (Remez, Rubin, Pisoni, & Carrell, 1981), is a form of copy synthesis in which a natural utterance is sampled, analyzed, and recreated by imitating its acoustic properties, with one prominent difference from typical speech synthesis. Most synthetic speech meticulously simulates the specific acoustic constituents of a speech signal, thereby achieving a natural vocal quality. In contrast, when composing a sinusoidal replica we make no attempt to fabricate the great variety of acoustic products of vocalization. Instead, the sinewave synthesizer is set to produce just three or four sinusoids to imitate the coarse-grain spectrotemporal properties of a speech signal. Such drastic reduction in the richness of the spectrum renders a sinewave replica completely unnatural in timbre (Remez et al., 1981; Remez & Rubin, in press), though most listeners readily transcribe a sinewave sentence replica as if it were the original natural speech sample from which it was derived (Remez, Rubin, Berns, Pardo, & Lang, 1994).

In essence, a sinewave sentence is intelligible though it does not sound like it is spoken by a natural voice. Consequently, if a listener is able to identify a familiar talker from a sinewave replica, then we can conclude that phonetic attributes are useful perceptually for identifying talkers independent of impressions of voice quality. This outcome is consistent with a perceptual mechanism that finds phonetic attributes in the course of lexical access and talker recognition alike and would encourage a hypothesis that phonetic analysis in these two perceptual functions underlies the apparent contingency of lexical identification on instance-specific attributes of utterances.

Although the sinewave sentences that we used in this study did not sound like speech, we otherwise meant to approximate the circumstance in which a listener recognizes a familiar talker. Consistent with this objective, we composed a set of speech samples for our tests from individuals who knew each other and from whose colleagues we could recruit volunteers for our listening tests. Every test used utterances produced by this set of talkers or sinewave replicas modeled from them. This method of assembling acoustic test materials satisfied the constraint of personal relevance (Van Lancker, 1991); the listeners knew the talkers through collegial interactions occurring over many years and were not trained in our brief test session to acquire familiarity with a collection of voices. Control procedures to impose uniformity in the variation of age, dialect, memorability, and distinctiveness of talkers in a set were relinquished[2] in order to ensure that we were studying the perceptual effects of naturally developed personal familiarity among talker and listener.

Three experiments spanned five tests. In the first experiment listeners were not familiar with the talkers; here, we estimated the apparent similarity of the natural speech sample of a particular talker and its sinewave replica in a preliminary assessment of the preservation of utterance-specific characteristics in the tone patterns. Listeners who were unfamiliar with the talkers accurately matched the sinewave and natural signals; this finding revealed that the transformation from a natural sample to a sinewave analog does preserve some utterance-specific characteristics. In a second experiment, we modified this task to prevent a superficial comparison of natural and sinewave tokens and to see whether listeners were able to resolve the differences among the talkers when the task required the listeners to select the natural and sinewave sentences produced by the same individual. Here, the lexical and syntactic constituents of the sentences were the same, but the natural samples had not been used as models for the sinewave items. To identify common properties of natural and sinewave signals, the listener was forced to rely on less superficial attributes than the auditory form of the signals. Again, listeners performed well in this test, which suggests that a close physical similarity between natural and sinewave signals was not required to allow listeners to identify the natural-sinewave correspondences that stemmed from the productive characteristics of each talker. In a third experiment listeners were familiar with the talkers; here, we estimated the ability of listeners to identify the source of a sinewave sentence by relying on well established familiarity with a talker. These listeners were able to identify talkers from the sinewave patterns, which revealed that phonetic attributes alone can be sufficient for recognizing a familiar voice.

Experiment 1

Method

The talkers. A set of natural sentences was compiled from the utterances of five male and five female talkers. All were members of the staff of Haskins Laboratories. The talkers were familiar with each others' voices, and their voices were familiar to the listeners of Experiment 3, as a result of formal and informal interactions occurring over many years. Instructions to talkers requested a fluent reading that was neither normative nor vernacular, and other aspects of speech production were left to each talker's habit. The talkers were not informed about the purpose of the experiment.

Acoustic test materials. Speech samples were obtained in a sound-attenuating chamber from each talker, who read a list of sentences aloud twice. The sentence "The drowning man let out a yell" appeared in the list and was used in acoustic analyses and as the model for sinewave synthesis in this project. The natural utterances were recorded on audiotape with a high-quality voice microphone and then converted to digital records by filtering

[2] The regions represented in the speech of the talkers included Great Britain (Received Pronunciation) and the Northeast and Midwest of the United States.

(4.5-kHz low-pass, −40 dB/octave rolloff) and sampling (at 10 kHz), using a pulse code modulation system implemented on a VAXstation II/GPX. We analyzed speech samples in two ways to estimate the center frequency and amplitude of the three lowest frequency formants throughout each utterance: (a) the peak-picking method of linear prediction and (b) the spectral analysis method of discrete Fourier transforms. We derived formant frequencies and amplitudes at 5-ms intervals and captured them interactively to compose a table of sinusoidal synthesis specifications for each utterance. A sinewave synthesizer (Rubin, 1980) generated the waveforms according to the synthesis parameters, with a temporal resolution of 10 kHz. These waveforms were stored on the VAX as digital records.

Test sequences composed of synthetic sinusoidal patterns and the natural utterances were converted from digital records to analog signals, recorded on half-track 0.25 in audiotape, and presented to listeners through tape playback. Listeners sat in carrels in a sound-shielded room, and signals were presented binaurally at an approximate level of 65 dB (SPL) over matched and calibrated headsets.

Procedure. On every trial of the test used in this experiment, a natural sentence was followed by two sinewave sentences. One of the pair of sinewave patterns always had been derived from the natural utterance presented on that trial. The other sinewave pattern had been derived from a natural utterance produced by one of the other nine talkers. A listener was asked to report which of the two sinewave sentences was based on the natural utterance presented on each trial. Each of the 10 natural sentences was presented with nine sinewave foils and the true replica.

With 10 different talkers, there were nine comparisons of each sinewave sentence with every other, making 90 trials; counterbalancing for order of presentation of the alternatives resulted in a test of 180 trials. On each trial, the natural sentence and the first sinewave sentence were separated by 750 ms of silence, and the first and second sinewave sentences were also separated by 750 ms of silence. Between each trial, there were 3 s of silence, with the exception of every 10th trial, after which there were 6 s of silence.

Listeners. Thirteen students at Barnard College were tested in groups of 6 or fewer. All were native speakers of English and reported no history of disorders of speech or hearing; none had participated in any other experiment that used sinusoidal signals. The listeners were drawn from introductory psychology classes and received course credit for their participation. They were briefly instructed that natural and sinewave speech was to be presented over headphones, and they were asked to decide on each trial whether the natural sentence was more similar to the first or to the second sinewave sentence.

Results and Discussion

Our test required listeners to select from a pair of sinewave patterns the one that matched the natural sample presented on each trial. Accordingly, guessing would have produced results approaching 50% correct. We performed a one-way repeated measures analysis of variance on the factor of talker to determine (a) whether listeners were able to match natural models and sinewave replicas equally well for all 10 talkers in the set and (b) whether matching performance differed from guessing. The analysis showed that performance differed across the set of talkers, $F(9, 108) = 11.8$, $p < .001$. We used a Tukey post hoc means test to estimate the likelihood that performance differed from guessing for each talker. This test revealed that listeners matched 8 of the 10 natural models to sinewave replicas better than chance. These data are shown in Figure 1, a histogram in which the height of each bar corresponds to the mean performance of the 13 listeners in identifying the sinewave replica of each of the 10 natural speech samples.

In this experiment, listeners were able to identify the sinewave version of a natural sentence despite the fact that every voice said the same sequence of words, as our listeners readily acknowledged. This result allows for two alternative interpretations. One explanation for the finding is that sinewave replicas contain information about the talker as well as the message, despite the absence of the acoustic correlates of voice quality as typically conceptualized. This information, whether specific to the age, dialect, style, or idiolect[3] of the talker, exists in the phonetic form of the utterances and was exploited by a listener making a correct match. However appealing this explanation is, the results are also consistent with the explanation that listeners based their performance on more superficial auditory attributes of specific tokens composing the test materials, attributes that are irrelevant to the characteristics of particular talkers. Although the physical variation across the set of utterances may ultimately derive from phonetic differences among the talkers, the speech samples also differ in meter, the rate of the frequency changes of the tonal components, the average frequency of each of the tones, and the overall duration. Listeners who focused on similarities of meter, spectrotemporal tempo, pitch, or duration between the natural sentences and the sinewave replicas could have made matches without actually registering phonetic differences among talkers. Because sinewave utterance replicas lack a fundamental frequency, and because the three components of the pattern have no consistent harmonic relationship, any superficial acoustic property responsible for the effect here must nonetheless lie outside the set considered by Church and Schacter (1994).

In Experiment 2 we used a test to distinguish performance based on the phonetic comparison of utterances from performance based on a superficial comparison of the tokens. We accomplished this within the same basic trial format of the first experiment by exchanging the natural tokens that were used as models for sinewave replication for a different utterance of the test sentence produced by each talker at the original recording session. Presumably, these utterances exhibited the characteristic phonetic properties of each talker, though they necessarily differed in the fine acoustic grain from the specific utterances that the sinewave replicas imitated. In this test we asked listeners on each trial to identify which of two sinewave sentences had been spoken by the talker who produced the natural sample. Clearly, listeners who chose correctly in this test would be those who were able to disregard dissimilarities in the fine structure of the auditory form of the tokens in favor of more abstract, phonetic similarities.

[3] The term *idiolect* refers to the manifestations of speech sounds that are unique to an individual talker.

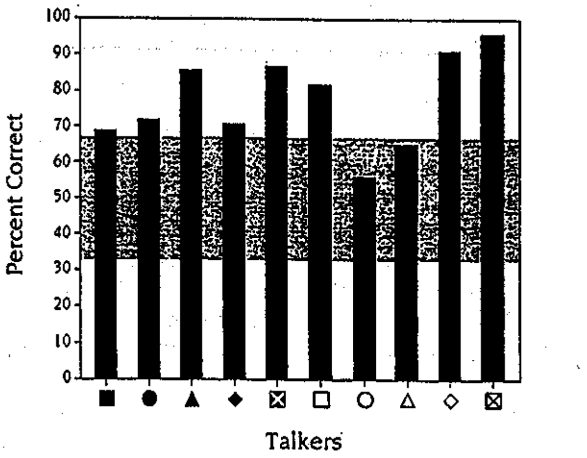

Figure 1. Results of Experiment 1, a test of similarity between natural utterances and their sinewave replicas. Each symbol on the abscissa corresponds to 1 of the 10 talkers (filled symbols = men; open symbols = women). Bars lying within the gray region do not differ from 50% correct, or chance, as shown by a Tukey post hoc means test.

Experiment 2

Method

Acoustic test materials. Each of the 10 talkers in the test set was represented by two signals in Experiment 2, one natural and one sinusoidal, though in contrast to the situation in Experiment 1, the natural signals that were used in this test differed from the utterances that were used as the models for the sinewave replicas. The natural samples used here had been obtained from the 10 talkers in our set at the original recording session. Again, the synthetic sinusoidal patterns and natural utterances were of the sentence "The drowning man let out a yell." The average absolute difference in duration between the natural utterances in Experiments 1 and 2 was 199.2 ms.

Procedure. Three signals were presented on each trial of the test, a natural sentence followed by two sinewave sentences. One of the pair of sinewave patterns always had been derived from a natural utterance produced by the same talker who had spoken the natural sample presented on that trial. The other sinewave pattern was derived from a natural utterance produced by one of the other nine talkers. A listener was asked to report which of the two sinewave sentences was produced by the same person who spoke the natural utterance on each trial.

There were nine comparisons of each sinewave sentence with every other, which made 90 trials, each of which occurred in two orders, for a test of 180 trials. Each trial had the same format: The natural sentence and the first sinewave sentence were separated by 750 ms of silence, and the first and second sinewave sentences were separated by 750 ms of silence; after each trial, there were 3 s of silence, with the exception of every 10th trial, after which there were 6 s of silence.

Listeners. Eighteen students at Barnard College were tested in groups of 6 or fewer. All were native speakers of English and reported no history of disorders of speech or hearing; none had participated in any other experiment that used sinusoidal signals. The listeners were drawn from introductory psychology classes and received course credit for participating.

Results and Discussion

Our test required listeners to choose between pairs of sinewave patterns, only one of which derived from the same talker whose natural sample had occurred on that trial. As in the first experiment, guessing would have produced results approaching 50% correct. We performed a one-way repeated measures analysis of variance on the factor of talker to determine (a) whether listeners were able to match natural and sinewave signals equally well for all 10 talkers in the set and (b) whether matching performance differed from guessing. The analysis showed that performance differed across the set of talkers, $F(9, 153) = 13.4$, $p < .001$. The likelihood that performance differed from guessing for each talker was estimated with a Tukey post hoc means test. This test revealed that listeners matched 8 of the 10 natural models to sinewave replicas at a rate better than chance. These data are shown in Figure 2, a histogram in which the height of each bar corresponds to the mean performance of the 18 listeners in identifying the sinewave based on the speech of the talkers whose natural samples were provided.

It is likely that the success of listeners in this task reflects an ability to register the phonetic properties of natural and sinewave signals alike and to compare them without recourse to the auditory correlates of the natural acoustic products of vocalization. This conclusion is warranted because in our test procedure, each talker's sinewave sentence was based on a natural model that differed from the natural sample used in the test. In Experiment 1, a listener could have performed the task by attending to subtle acoustic similarities—for instance, the temporal pattern of the natural and sinewave signals—because on each trial one of the sinewave complexes derived from the natural utterance that the listener heard. Likewise, the listener may have found the tempo of rise and fall of the energy envelope or the specific incidence of silences useful for assessing the physical sim-

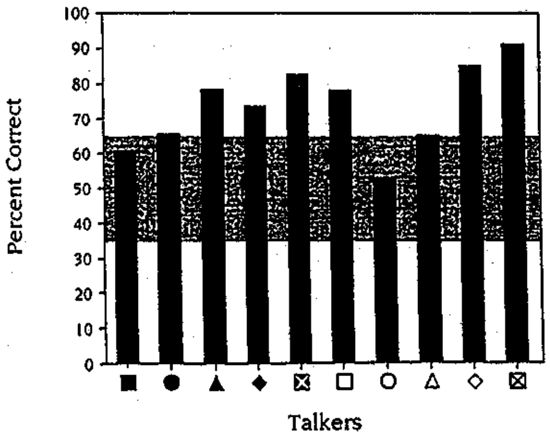

Figure 2. Results of Experiment 2, a test in which listeners matched the talkers of a natural and a sinewave sentence. Each symbol on the abscissa corresponds to one of the 10 talkers (filled symbols = men; open symbols = women). Bars lying within the gray region do not differ from 50% correct, or chance, as shown by a Tukey post hoc means test.

ilarity of natural and sinewave signals without necessarily considering the phonetic form of the tokens. In Experiment 2, however, the use of a natural sentence differing from the models of the sinewave patterns meant that no sinewave and natural token ever coincided in acoustic fine structure.

With respect to the conjecture that launched these studies—precisely, that the perceptual recognition of individuals can be sustained by the phonetic properties in a speech sample—a proof of its possibility still requires more stringent a test than we made in Experiment 2. In fact, were we to accept the argument that listeners compared phonetic properties as opposed to nonphonetic auditory attributes in matching natural and sinewave signals here, a test of the premise about talker recognition would nonetheless depend on evidence that the comparison is governed by a listener's implicit tolerance for phonetic variety specific to individuals. We sought this evidence in Experiment 3, which also allowed us to examine the data set of Experiment 2, through hindsight, for evidence of talker-scaled perceptual standards.

In Experiment 3 we tested the hypothetical phonetic basis for recognizing a familiar talker under test conditions that did not require a listener to compare natural and sinewave signals in succession. We recruited listeners who were familiar with the natural voices of the 10 talkers in our sample set. Familiarity had been established in an ordinary way, over the course of many years of collegial interaction. In the crucial test, we asked listeners to recognize the talkers from sinewave signals without offering them a successive comparison of natural and sinewave signals. The listeners relied here on long-term familiarity with a talker's vocal characteristics, which ensured that performance reflected perceptual sensitivity to phonetic variation at the scale of the individual talker and eliminated the possibility that performance in this study reflected successive comparison of natural and sinewave tokens.

Experiment 3

Method

Acoustic test materials. In the first two tests of Experiment 3 we used the sinewave sentences from Experiment 1. For the third test we used the natural sentences on which the sinewave sentences were modeled. However, the procedures differed across the three tests of Experiment 3.

Procedure. The experiment consisted of three tests, each with the same listeners. The first test was a lower uncertainty test of identification that required a listener to distinguish between sinewave voices to identify a designated talker. Without relying on a natural sample, the listener had to base this identification on long-term familiarity with the characteristics of the voices in the talker set. For the second test we used a 10-alternative forced-choice identification test to determine absolute identifiability of sinewave voices in conditions of relatively high uncertainty. The third test also used a 10-alternative forced-choice procedure but with natural utterances; its purpose was to verify that each of the 10 talkers was identifiable from the raw acoustic signal.

In the first test, each trial had the same structure: The listener read a colleague's name, heard two sinewave sentences in succession, and identified which of the two sinewave sentences was produced by the named talker. There were 10 target talkers and 10 different sinewave replicas, one for each of the talkers. This test had 180 trials, with each of the 10 sinewave sentences compared with every other in two orders of presentation. On each trial, the first and second sinewave sentences were separated by 750 ms of silence. Between each trial, there were 3 s of silence, with the exception of every 10th trial, after which there were 6 s of silence.

In the second test, one of the 10 sinewave replicas was presented on each trial, and the listener was asked to identify the colleague from whom the pattern derived. Each of the 10 sinewave patterns was presented six times in random order, for a total of 60 trials. Each trial was separated from the next by 3 s of silence, with the exception of every 10th trial, after which there were 6 s of silence.

The third test used the 10 natural sentences from which the sinewave replicas were derived. Participants were asked to listen to each natural sentence in turn and to identify the talker who produced it. As in the second test of Experiment 3, each sentence was presented six times in random order, for a test of 60 trials. Each trial was separated from the next by 3 s of silence, with the exception of every 10th trial, after which there were 6 s of silence.

Listeners were told that natural or sinewave signals were to be presented over the headphones and were instructed separately for each of the three tests in this experiment. Because of the crowded schedules of listeners, who could not sacrifice so much time to our project all on a single day, at least 1 week intervened between the first and second test sessions. The second and third tests were presented during the same experimental session.

Listeners. Nineteen coworkers at Haskins Laboratories ranging in age from 35 to 74 years served as listeners. Eight of the listeners had contributed speech samples that were used to compose the listening tests. Two who had contributed speech samples did not participate in the listening tests. Four were unable to complete all of the test conditions of this experiment and were excluded from the data set. One listener used a hearing aid. All were volunteers and were not told the purpose of the experiment until the testing session began. They were tested in groups of 3 or fewer in visual isolation. Each reported familiarity with the voices of the talkers in the test.

Results and Discussion

In this experiment we tested whether the consonant and vowel allophones conveyed in a sinewave replica provide useful information about a talker. Overall, listeners reported that the sentences were readily understood in all conditions. In the first test, a listener read the name of a talker and then heard two sinewave replicas, one of which was based on the natural speech of that talker; guessing would have produced results approaching 50% correct. We performed a one-way repeated measures analysis of variance on the factor of talker to determine (a) whether the sinewave replicas were identified equally well for all talkers and (b) whether performance differed from guessing. The analysis showed that talkers were not equally identifiable, $F(9, 126) = 5.24, p < .001$. To estimate the likelihood that performance differed from guessing for each of the talkers, we used a Tukey post hoc means test. This test revealed that performance for all 10 talkers significantly differed from chance. These data are shown in Figure 3.

In the second test administered to these listeners, we presented a single sinewave replica of an utterance on each trial for listeners to identify. This was a 10-alternative

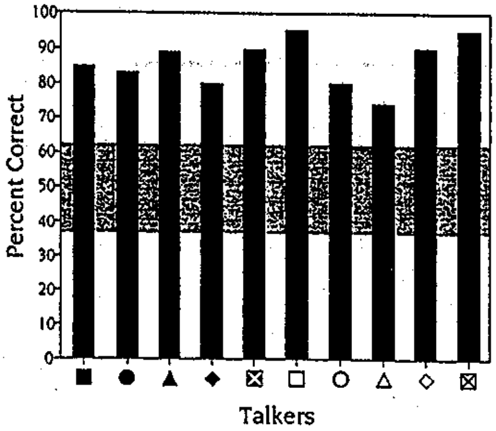

Figure 3. Results of the first part of Experiment 3, a lower uncertainty test of the identification of familiar talkers from sinewave replicas. Each symbol on the abscissa corresponds to one of the 10 talkers (filled symbols = men; open symbols = women). The gray region does not differ from 50% correct, or chance, as shown by a Tukey post hoc means test.

forced-choice test; guessing would have produced results approximating 10% correct. We performed a one-way repeated measures analysis of variance on the factor of talker to determine whether sinewave signals were identified equally well for all talkers and, again, whether performance differed from guessing. The analysis showed that talkers were not equally identifiable, $F(9, 126) = 11.1, p < .001$. We used a Tukey post hoc means test to estimate the likelihood that performance differed from guessing for each of the 10 talkers. This test revealed that identification performance for 6 of the 10 talkers differed significantly from chance performance. These data are shown in Figure 4.

We administered a final test to these listeners. The natural speech versions of each of the 10 sentences were presented in an identification test with 10 alternatives on each trial. The average of all of the scores was 97% correct, with chance being 10% correct; there was insufficient variance to perform a statistical analysis.

In this experiment, listeners were able to identify many of the talkers from whose utterances the sinewave replicas were derived. These listeners had not heard the natural samples at this juncture and must have drawn on long-term knowledge of a talker's voice in performing the test. This result reveals that information about a talker remains available in a sinewave replica despite the elimination of intonation and natural vocal timbre. It also allows for the possibility that listeners in Experiments 1 and 2 used the same grain of phonetically based information about a talker to match sinewave replicas to natural samples; we return to this point in the General Discussion. Overall, this research shows that listeners are able to identify voices from signals that lack natural intonation and voice quality, which implies that the phonetic properties of utterances can convey both lexical and personal information.

General Discussion

How Did Perceivers Identify Sinewave Talkers?

In interpreting the outcomes of these three experiments, we submit that dissimilar groups of listeners performing in three rather different tasks identified talkers in much the same way, that is, by virtue of their sensitivity to phonetic attributes in the speech of individuals. More pointedly, we suggest that a listener in any of our three experiments registered the phonetic attributes of sinewave or natural signals in a manner specifically scaled to the segmental phonetic varieties produced by individual talkers. There are two obvious contrasting hypotheses that appeal to a narrower, token-based evaluation of the superficial properties of test sentences. To explain the outcome of Experiment 1, one of these alternatives appeals to the ability of listeners to compare auditory forms. To explain Experiment 2, the other appeals to the ability of listeners to compare phonetic form in piecemeal fashion, one segment at a time. Neither of these alternatives is likely. To see that this is so, recall the evidence of the third experiment, in which listeners could not have detected superficial acoustic or phonetic similarities between sinewave replicas and natural samples because no natural models were available for comparison. No matter how a perceiver represents a familiar talker, the long-term traces of the voices of colleagues surely do not include a natural instance of the sentence "The drowning man let out a yell." Therefore, we infer that the performance of listeners in Experiment 3 did not rely on an exact comparison of present perception and remembered perception. This rather clear outcome in Experiment 3 provides a standard for evaluating Experiments 1 and 2.

Without an occasion to hear a natural sample of each talker during the test, a listener in Experiment 3 nonetheless

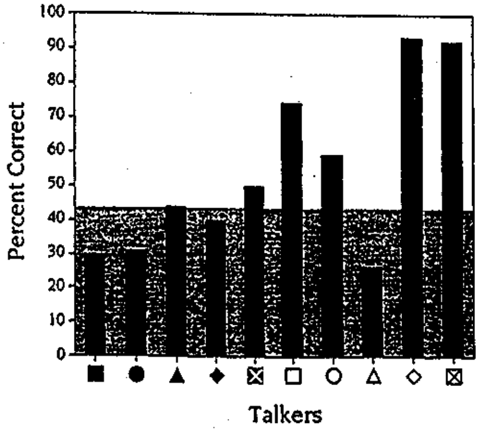

Figure 4. Results of a test of absolute identification of the talker producing the model for a sinewave replica; a 10-alternative forced-choice procedure was used. Each symbol on the abscissa corresponds to 1 of the 10 talkers (filled symbols = men; open symbols = women). Bars lying within the gray region do not differ from 10% correct, or chance, as shown by a Tukey post hoc means test.

identified colleagues from the phonetic attributes preserved in sinewave sentences. In contrast, the listeners in Experiments 1 and 2 had a good opportunity on each trial to appraise the physical similarity of sinewave and natural sentences, though apparently they did not do so. This inference rests on a series of comparisons we made of the performance in the three experiments. First we estimated the perceived similarity of the talkers to each other from the misidentifications in Experiment 3 in order to create an index of the perceptual inclination to treat the recognition of talkers as a matter of phonetic form in aggregate. Then we scaled the errors for the results of Experiments 1 and 2 to represent the perceived similarities in those two studies. Comparison of the three experiments revealed that the pattern of perceived similarity observed in the three groups of listeners was roughly the same. It is implausible that this outcome would have occurred if listeners in Experiments 1 and 2 had used very different perceptual criteria than did the listeners in Experiment 3.

Evaluating the data sets in order to make these comparisons was straightforward and conventional. We tabulated the confusion errors in Experiments 1 and 2 and the parallel test in Experiment 3 and applied a scaling solution (Kruskal, 1964) and a hierarchical clustering analysis (Johnson, 1967). These analyses are shown graphically in the panels of Figure 5. With one panel for each experiment, a talker is represented as a point in a plane, and the distance between any pair of talkers reflects the likelihood that our listeners misidentified one for the other; curves enclose each successive nesting of similarity given in the hierarchical clustering analysis. The visual impression that the similarity of the talkers, each to each, exhibited the same general pattern in Experiments 1, 2, and 3 was confirmed by a test of correlation of the nesting ranks produced by the clustering analyses. For Experiments 1 and 3, Spearman's $r_s = .62$, $p < .05$; for Experiments 2 and 3, Spearman's $r_s = .88$, $p < .05$; both tests were corrected for tied ranks by the method of Siegel and Castellan (1988). One conclusion warranted by this outcome is that the attributes of sinewave speech that were available to the listeners who identified colleagues in Experiment 3 were also available to listeners who compared natural and sinewave signals in Experiments 1 and 2. The slightly higher correlation between Experiments 2 and 3 than between Experiments 1 and 3 may reflect the fact that exact comparisons were specifically disadvantageous in Experiment 2 and that listeners may have responded to this circumstance by attending more consistently to phonetic attributes, in the manner of the listeners in Experiment 3.

By themselves, coincident estimates of perceived similarities constitute only equivocal confirmation of the hypothesis that listeners in the first two experiments recognized talkers. The scaling analyses show only that listeners in each of the groups applied similar assays. Were there no other evidence, we might conclude that this congruence reflected the action of an auditory analysis in all three cases, as opposed to a similar perceptual focus on segmental phonetic properties in aggregate. Even as a first approximation, though, a hypothetical reduction of speech perception to auditory sensitivity is generally inconsistent with evidence

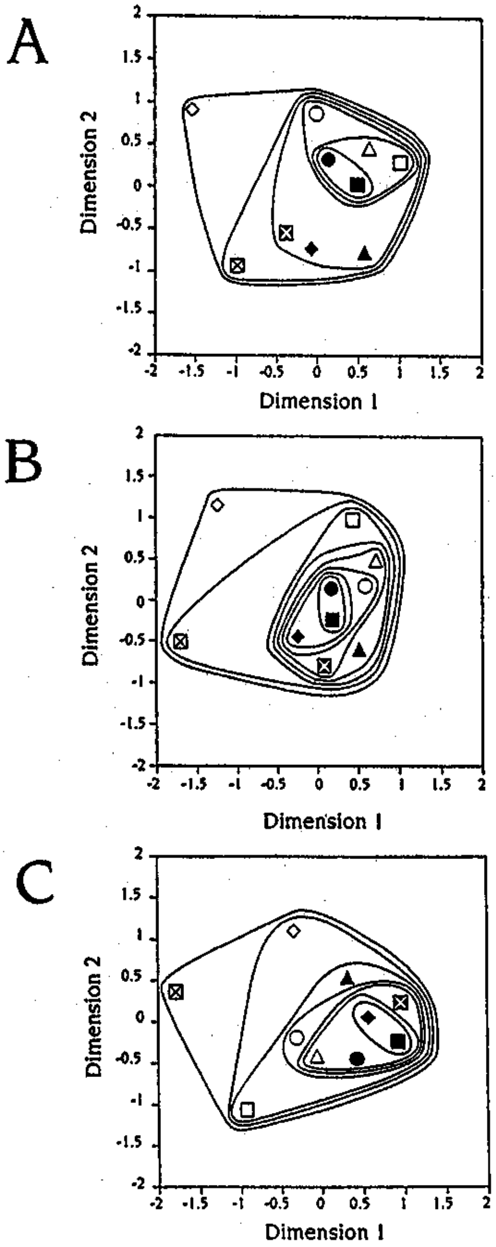

Figure 5. Multidimensional scaling and clustering analyses of the perceived similarity of 10 talkers, based on the error data of (A) Experiment 1, in which listeners matched the natural models; (B) Experiment 2, in which listeners matched natural and sinewave talkers; and (C) Experiment 3, in which listeners identified a familiar talker. Each bullet corresponds to 1 of the 10 talkers (filled symbols = men; open symbols = women). On each plot, the position of the points is given by the scaling solution; a curve encloses successive levels of similarity.

(summarized in Remez, 1994) that perceivers do not form distinct auditory impressions of formant patterns—nor are they even able to resolve the fine acoustic grain of a speech spectrum without considerable determination and training (e.g., Carney, Widin, & Viemeister, 1977)—though it is possible that a sinewave presentation of phonetic information skirts this constraint. Certainly, our own studies have shown that sinewave sentences evoke impressions concurrently of the phonetic and the auditory forms of the tones (Remez, Pardo, & Rubin, 1992; Remez & Rubin, 1984, 1993). This means that the frequency of the first formant, usually taken to be a correlate of the overall length of the vocal tract and therefore a potential source of personal information, was undoubtedly available to our listeners as an impression of time-varying pitch. The second-formant analog in a sinewave replica is likewise a potential source of pitch impressions, though its role in marking personal attributes is less certain.

In a converging analysis intended to gauge the likelihood that the perceived similarity of the talkers in our tests was driven by auditory attributes subordinate to phonetic properties, we compared the similarity of individuals within the talker set considered acoustically with the estimates of perceptual similarity produced by the scaling analyses. Our expectation was that an acoustic analysis would echo the pattern of perceptual similarity observed in the scaling analyses if the dimensions of the plots in Figure 5 were approximately acoustic and physical (e.g., formant frequencies) as opposed to phonetic (e.g., a more complex dimension reflecting whether vowels are raised or lowered, whether /t/ is held and released or tapped, and whether consonant clusters are more or less assimilated). We began our acoustic analyses by defining each talker as a central tendency in formant space. The graph shown in Figure 6 plots the mean values of the 10 speech samples along an abscissa of the first formant and an ordinate of the second.

Apart from revealing that the frequency of the first formant provides a poor way to distinguish the talkers from one another in our samples, the pattern of acoustic similarity differs starkly from the data of Figure 5. On this acoustic criterion, there is noticeable ordering of the talker means by sex. None of the similarity scaling solutions shown in A, B, or C of Figure 5 reflects this sex-correlated segregation. Although this perceptual outcome is unusual for a study of talker recognition, it is not surprising in the present case. The acoustic correlates of laryngeal pulsing and, therefore, the differences in voice pitch and spectrum that typically distinguish men and women are eliminated in sinewave sentences. Listeners must use other attributes exclusively to identify talkers, if they can.

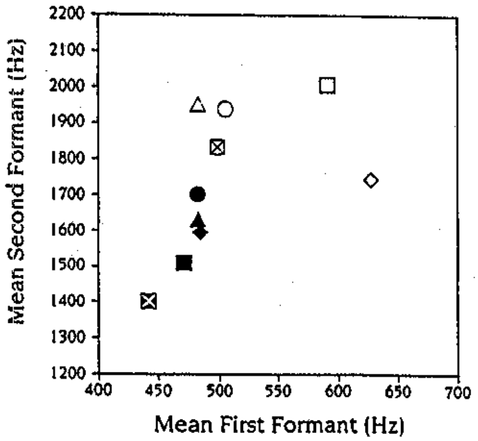

Figure 6. An acoustic analysis of the 10 speech samples used in the tests of talker identification. Each symbol corresponds to 1 of the 10 talkers (filled symbols = men; open symbols = women). The position of each symbol in the plane is determined by the mean values of the first and second formants of the talker. Compare to the three panels of Figure 5.

Because a talker's central spectral tendency is correlated with the talker's vowel space (Nearey, 1978; Peterson & Barney, 1952) it should also be evident that perceptual similarity here did not reflect the characteristic range of formant frequencies across the talker set. Of course, specific vowel contrasts can be manifest quite differently under phonetic control within the same frequency band, for which reason we also examined the acoustic properties of individual vowels across the talker set but to no avail. Evidently, the discrepancy between the acoustic and perceptual plots shows that the transformation of a natural signal to a sinewave replica draws a listener's attention away from superficial characteristics of a voice, such as its fundamental pitch or spectral pitch, in favor of more abstract linguistic attributes that appear to index the talker just as well.

We have discussed average formant frequency here in part because such acoustic properties historically have been both the first and the last resorts in characterizing variation across a set of talkers, though our conclusion is no less true of a comparison of the perceptual scaling with the other acoustic variables that we considered: (a) formant range, (b) absolute formant frequency variation, (c) the standard deviation of formant frequency normal to the mean formant frequency of a sample, and (d) the formant frequency of the nucleus of the vowel /æ/ in the word *man*. In each of these estimates, we were searching for an acoustic correlate of a talker's anatomical scale or an acoustic property reflecting a characteristic vocal posture (Laver, 1980). In contrast to the work of Nolan (1983), who found that changes in formant ratios were adequate acoustic indices of contrastive voice qualities, we found that both the obvious and the obscure acoustic correlates poorly matched the scaling solutions.

Could a momentary or average spectrum envelope of a sinewave replica have provided a basis for identifying talkers on acoustic grounds? This aspect of a natural utterance is replicated schematically in a sinewave complex, which presents the formant center frequencies at the estimated amplitudes of the original natural spectrum. This acoustic property is typically correlated with impressions of timbre and potentially allows a listener to resolve similarities between natural and sinewave signals on a basis other than the phonetic grain. However, several considerations militate against this possibility, among them the fact that this spectral property has proven to be far from ideal for characterizing personal quality of natural speech samples (Matsumoto et al., 1973), to say nothing of the problematic status of impressions of timbre associated with nonharmonic components.

No less important is the finding that listeners simply do not rate sinewave replicas of speech to be particularly natural in quality (Remez et al., 1981; Remez & Rubin, in press), which means that determining the likeness of sinewave and natural sentences requires something other than a straightforward comparison of impressions of timbre. Last, a caveat offered by Klatt (1985) suggests that the spectrum envelope must play a minor role if it contributes at all. Considering the design characteristics that lend robustness to speech, Klatt speculated that the amplitude characteristics of a speech signal are easily corrupted by the vagaries of transmission, including ambient noise and changes in the proximity or direction of a talker's mouth relative to the listener's ear. In contrast, the potential for frequency distortion of the speech spectrum is far narrower, and on these grounds Klatt reasoned that the listener relies less on the precise shape or slope of the spectrum envelope than on the frequencies of its resonance peaks. Together, these arguments dissuade us from supposing that the spectrum envelope of a sinewave replica evokes an impression of timbre similar to that of its natural model or from supposing that perceivers are especially adept at registering a property of acoustic signals that is unlikely to provide much information about a talker's identity or message.

Overall, this review of the evidence opposes a conclusion that the listener's performance in our experiments is based on the perception of typical short- or long-term auditory manifestations of a talker's speech. At the very least, it is possible to say with conviction that the perceptual scaling of these particular sinewave signals differs unmistakably from a representation of their physical similarity. It may be possible to use empirical tests to confirm the conclusion that we recommend: that the perceptual criteria of each group of listeners inherently promoted the salience of segmental phonetic properties of the sentences, by virtue of which the perceivers identified both the words and who said them. Because Experiment 3 also precluded a piecemeal phonetic comparison of sinewave signals and correlated natural samples, the perception of talkers in that study surely reflected attention to more global aspects of phonetic variety—and by analogy, so did the perception of talkers in Experiments 1 and 2. We turn next to consider personal information in idiolect, style, and dialect.

Idiolect, Style, and Dialect

Although we describe the basis for sinewave talker identification as phonetic in nature, we have no way to tell presently whether the attributes pertinent to a listener's performance were registered as phonetic segments, as a concurrent impression of a talker's identity, or even as an impression of style or dialect unbound to particular phonetic ingredients. In debriefing sessions, many listeners in Experiment 3 reported that the familiar personal quality of one or another talker was uncannily apparent despite the distracting timbre of the tones. On the evidence of perceptual reports such as these, it is tempting to speculate that a characterization of the variety of allophonic manifestations of an individual's speech is available in long-term traces of a familiar voice. If this faculty was exploited by listeners in Experiment 3, then perhaps listeners in Experiments 1 and 2 performed so similarly because they were learning the talker set by inducing similar compilations of phonetic characteristics.

Technically, this kind of linguistic designation falls under the rubric of idiolect, dialect, or style, all three superordinate to a segmental grain. However, the differential distribution of phonetic segments in the speech of a single talker or group of talkers arguably underlies distinctions at these superordinate levels. Of particular relevance to the recognition of individuals without listener access to voice quality are reports by Amerman and Daniloff (1977) and Bladon and Al-Bamerni (1976) that there are characteristic talker-specific differences in coarticulatory assimilation of consonants. This evidence encourages the hypothesis that sensitivity to consonant allophones promotes the identification of individuals. At a coarser grain, though, even the case of rhythmic style is more profitably pursued with a segmental account than an explicitly suprasegmental one, as can be seen in, for example, the irreducibility of the speech rhythm of individuals to distributions of acoustic duration (Doherty & Hollien, 1978; Markel, Oshika, & Gray, 1977). An instructive parallel case can be seen in Granström's recent efforts to improve the liveliness and presence of synthetic speech by varying the speaking style of an automatic text-to-speech system; these improvements used segmental phonetic alternations to produce impressions of style alternations (Granström, 1992; Granström & Nord, 1991). Of course, characteristic dialectal distributions of segmental properties are well known (Byrd, 1994; Labov, 1972; Ladefoged, 1967; McDonough, Ladefoged, & George, 1993; Nolan, 1983; Trudgill, 1974; Wells, 1982). Therefore, it is not implausible that listeners in Experiment 3 perceived characteristic differences in the idiolect, dialect, or style of their colleagues by virtue of the listeners' sensitivity to the phonetic attributes conserved in the sinewave signals.

Because biographical and social factors influence the attainment of familiarity with a talker through ordinary means, one needs a large measure of good luck in constructing talker sets in order to pursue this hypothesis empirically. Chiefly, one would need to control the dimensions of variation within a set of talkers more than we aimed to here, in order to evaluate the use of phonetic information in the perceptual encoding of voices. An ideal group of talkers in this regard would be composed of a set of adults who vary in idiolect alone but share anatomical dimensions, dialect, formal and vernacular forms, the habits of changing speech registers, and chronological age. It may be possible to assemble an approximation to this ideal talker set.

In the present experiments these aspects of variation went unconstrained so that we could observe the prescription of Van Lancker (1991) to use voices that were personally familiar to the listeners. In her experiments, and often more widely, Van Lancker observed that measures of familiarity in tests of talker recognition appear to vary systematically as a function of the method by which familiarity has formed. Familiarity that develops through specific arbitrary training in an experimental setting is fragile and distinct from familiarity that accrues in ordinary contexts through ordinary means. This distinction was useful when Van Lancker observed phonagnosic patients, clinical listeners who sustained no impressions of familiarity with voices that were in fact familiar yet who were able to distinguish the voices responsible for producing different speech samples. This ability to perceive the unifying properties of a person's speech appears early in childhood (Jusczyk, Hohne, Jusczyk, & Redanz, 1993; Mann, Diamond, & Carey, 1979), and, as our findings imply, linguistic criteria may accompany attributes of vocal timbre. Although the similarity in perceptual dispositions of our two groups of listeners reveals that it is possible to study some aspects of talker recognition with naive listeners in arbitrary procedures, we did observe greater proficiency when listeners knew the talkers in ordinary circumstances. The nature of this performance difference remains to be determined.

Contingency and Independence of Lexical Access and Talker Identification

Although our specific technical goal in these experiments was to calibrate the phonetic contribution to the identification of a familiar talker, our motive in this project was to understand how talker- or utterance-specific properties of speech can affect the identification of words. From a conventional perspective, a contingency of lexical processes on talker identification is troubling, because it appears to require the commingling of sensory attributes and perceptual analyses that have formerly seemed different in kind. The presumed dissociation of talker identification and word identification that pervades the literature precludes such a contingency or at least offers no way to accommodate it coherently.

In our experiments we pursued a solution for the problem by identifying a common form of description in the perception of words and talkers. To play such a role, a representation of speech must hold the potential to distinguish spoken words and, at the same time, reflect the variability inherent in specific utterances. A representation of this hypothetical kind could moderate both lexical and individual identification yet present an appearance of contingency between lexical and personal identification that would be consistent with the traditional perspective that segregates the two functions. What kind of analysis would serve both lexical and individual perception?

It seemed self-evidently unreasonable to suspect an auditory attribute code for this purpose. Although one recent model of recognition, LAFS (lexical access from spectra; Klatt, 1980), adopted an auditory–acoustic basis for recognition of words, the technique was admittedly an unrealistic characterization of perception (Klatt, 1989). At the same time that research has identified the fundamental frequency and various components of the speech spectrum as critical variables in the perception of voices, other studies have shown that the primary perceptual representation of speech is unlikely to consist of such auditory forms captured directly from the speech signal (Hadding-Koch & Studdert-Kennedy, 1964; Lieberman, 1965; Remez et al., 1994; Silverman, 1985). It is easy to conceive of a memory process that uses derived auditory representations of vocal timbre, such as strength, melodiousness, or forcefulness (Gelfer, 1988), but an addressing scheme for lexical access based upon that attribute set would hardly be possible. Words do not differ from each other distinctively in such terms.

A phonetic level of description seemed likelier, in principle, as a common code for words and talker-specific characteristics. Segmental phonetic descriptions characterize the consonant and vowel allophones of a specific utterance in a form that is directly related to the abstract phonemic constituents that distinguish lexical items from one another. Some accounts of lexical access already describe the perceptual resolution of phonemic representations based on phonetic considerations (Luce et al., 1990), and the evidence is good that phonetic descriptions characterize words and specific spoken utterances. However, there is no evidence in the archival literature that talkers are identifiable solely from allophonic variation. Our tests here show that talker identification persists despite the absence of much of the first-order spectrotemporal properties that have figured prominently in the classic literature on the identification of talkers. The outcomes encourage a phonetic hypothesis, though we recognize the accumulated evidence for the functional independence of word and talker recognition and the research showing that each function depends on different sensory ingredients.

In retrospect, the arguments and evidence favoring functional independence in the identification of words and talkers are weaker than its proponents allow. Although the dissociation of functions observed in the neuropsychological literature cannot be casually denied, the psychological evidence for the independence of linguistic and individual recognition is equivocal. In light of our findings, the evidence contributed by studies of filtered and reversed speech may not justify a claim of independence of these two faculties. Although a band-limited or temporally reversed speech signal can prevent a listener from understanding what the talker said, it does not prevent phonetic analysis. Such listening conditions may preserve portions of vowels, fricatives, nasals, and laterals, because these segments are associated with acoustic attributes that are conserved over reversal. (Eliminated in the transformation are the acoustic patterns appropriate for stop-consonants, of course, which are associated with a critical sequence of acoustic events, [Bastian, Eimas, & Liberman, 1961; Fitch, Halwes, Erickson, & Liberman, 1980; Halle, Hughes, & Radley, 1957]. Nonetheless, the spectra of stop releases among other constituents of the pattern are available in a reversed signal, though it remains to be shown whether any of these constituents is specifically useful for identifying talkers.) Therefore, a test that uses reversed speech assesses voice recognition without lexical recognition but does not arrest phonetic analysis. Identification of a talker's voice under such circumstances cannot confirm the exclusive importance of paralinguistic aspects of speech. Likewise, filtering a speech signal leaves a residue of acoustic structure that is recognizable as speech and may permit perceptual analysis of phonetic attributes relevant to the identification of talkers even when lexical access is equivocal.

A phonetic account of the kind we propose can also offer a parsimonious description of Church and Schacter's (1994) recent tests of implicit memory[4] of words. In their report, explicit memory effects were brought about by procedures that ensured the encoding of selected aspects of the utterances that listeners studied: They asked the listeners to complete a spoken fragment of a word, to rate the clarity of a talker's enunciation of each word in a spoken list, or to report the degree of relative polysemy of each studied word. They produced implicit effects by changing the acoustic properties of some of the test items between the study phase and the test phase of a session, and they accordingly described the ensuing deterioration in performance as *voice-specific*. The mechanism that Church and Schacter entertained to accommodate this diversity of elaboration of spoken words linked the familiar phonological representations to a perceptual representation system devoted to acoustic properties of voices, chiefly the phonatory fundamental frequency. This echoes the customary stipulation of different sensory ingredients underlying lexical and personal identification, but it is instructive to review the incidence of effects with phonetic attributes in mind.

Not all of the acoustic differences that Church and Schacter used affected implicit memory. A difference in the gross signal amplitude of a word between the study and test phase had no effect. Of course, this acoustic manipulation would have been arbitrary with respect to articulation and, therefore, would have been phonetically neutral (also, see Nygaard, Sommers, & Pisoni, 1995; Sheffert & Fowler, 1995). In contrast, the phonetic properties of spoken words would have been altered in each condition in which an allegedly acoustic manipulation produced an implicit memory effect. Consider the conditions: (a) When a word was produced by different talkers in the study and test phases, allophonic differences would have arisen that were due to idiolectal contrast, even between talkers of the same regional dialect. (b) When a word in the study and test phases differed by the pattern of fundamental frequency variation, manipulated through speech synthesis, vowel quality would have been altered at the very least (House & Fairbanks, 1953; Silverman, 1985). (c) When a word exhibited the contrasting emotional attitudes of the talker in the study and test phases, phonetic attributes of vowels would have also been likely to differ because of changes in vocal resonances associated with different postures of facial expression (Tartter, 1980; Tartter & Braun, 1994). If intonation was also affected by a difference in attitude of the talker, this would surely have altered vowel quality.

Altogether, phonetic attributes can resolve the manipulations of spoken words that are effective in tests of implicit memory. Moreover, a phonetic premise precisely excludes the acoustic manipulation that elicited no implicit effect, though it occasioned large physical differences in the acoustic power of a word between study and test phases. This is evidence of a phonetic contingency in implicit memory and discourages the hypothesis that listeners always recognize words and talkers by analyzing different acoustic constituents in different perceptual processes. It seems more likely that phonetic attributes of speech serve both functions and that implicit memory effects are attributable to the same phonetic variations with which the lexical system contends. Although we have demonstrated the feasibility of an alternative to the account of Church and Schacter (1994), we hesitate to dismiss their speculations on the basis of our findings alone. We simply note that when accounts of memory are contingent on accounts of perception, a false assumption about the encoding can be a great hazard.

Nonetheless, the usefulness of phonetic descriptions is not evidence that a single perceptual function identifies words and talkers alike, and we must abjure any proposal to combine lexical and individual identification in a single act of retrieval (see Goldinger et al., 1991). Though there are only shaky grounds for rejecting this gambit in principle, the neuropsychological literature warrants no less (Van Lancker et al., 1988; Van Lancker & Kreiman, 1985, 1987). In phonagnosia, conscious access to the identity of a talker is impaired, although the allophonic variants specific to the speech of an individual are registered and accommodated in phonetic perception and lexical access. It remains for additional investigations to show whether this intriguing deficit reflects an inability to analyze the talker-relevant constituents of a speech signal or occurs because of an incapacity to convert tacit perceptual analysis of voices to overt perceptual experience (cf. Bauer, 1984; De Haan, Young, & Newcombe, 1991).

Conclusion

In Experiments 1 and 2, listeners were able to identify a talker from a sinewave replica of a speech signal with accuracy. In Experiment 1, listeners compared sinewave ensembles to natural samples. Their performance could have derived from acoustic or paralinguistic similarities between the sinewave and natural sentences, although this proved unlikely. In Experiment 2, the natural signals presented on each trial were different utterances than those that had been used as models for the sinewave sentences. Listeners who reported that the same talker had produced a natural sentence and a sinewave sentence could not have relied on an assessment of physical similarity or on piecemeal phonetic similarity of natural and sinewave samples. In Experiment 3 we conducted a test using listeners who knew the talkers so we could determine whether it was possible to recognize a familiar voice from a sinewave replica. In this experiment, we found that listeners identified the talker of a sinewave sentence without access to a natural sample to facilitate performance. These listeners must have been using phonetic information preserved in the sinewave signals to distinguish and to identify talkers whose vocal characteristics were already known. Taken together, the findings lead us to conclude that the information promoting

[4] *Implicit memory* is observed when experience facilitates performance without a conscious or deliberate attempt to remember; *explicit memory* is observed when performance requires a recollection of experience (Graf & Schacter, 1985; Schacter, 1987).

talker and word identification is harder to separate than traditionally proposed. Without acoustic information about voice quality, our listeners recognized talkers by using only information about the linguistically governed articulation.

Our studies suggest a solution to the problem that a phonemic addressing scheme has in admitting talker-specific effects in word recognition. Sinewave replicas do not preserve the acoustic correlates of voice quality, and the fact that listeners can identify talkers from the phonetic properties preserved in tonal analogs of speech exposes a formerly unexamined means of recognizing talkers and calls into question the traditional account. Because phonetic information is useful in talker recognition and in word recognition alike, perhaps this common code is ordinarily exploited toward both ends.

If phonetic segmental properties provide a common basis for voice and word recognition, it is likely that allophonic variation is represented at a lexical level. A corollary of this premise is that talker-specific facilitation of allophonic sensitivity should occur. We need to know, as well, whether the differential identifiability of talkers from sinewave replicas stems from variability in allophonic similarity among a talker set or from intrinsic differences in the ability of sinewave replication to preserve critical variables differentiating talkers. Last, it will be useful to determine whether allophonic properties are used for talker identification when acoustic correlates of voice quality are present in an intact signal. In conclusion, our experiments demonstrate the plausibility of a common phonetic basis for the recognition of words and the identification of talkers, in contrast to the parallel recognition of linguistic and personal attributes of utterances, and point the way to empirical tests that offer a clearly resolved portrait of this component of speech perception.

References

Amerman, J. D., & Daniloff, R. G. (1977). Aspects of lingual coarticulation. *Journal of Phonetics, 5,* 107–113.

Atal, B. S. (1972). Automatic speaker recognition based on pitch contours. *Journal of the Acoustical Society of America, 52,* 1687–1697.

Atal, B. S. (1974). Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification. *Journal of the Acoustical Society of America, 55,* 1304–1312.

Bastian, J., Eimas, P. D., & Liberman, A. M. (1961). Identification and discrimination of a phonemic contrast induced by silent interval. *Journal of the Acoustical Society of America, 33,* 842.

Bauer, R. M. (1984). Autonomic recognition of names and faces in prosopagnosia: A neuropsychological application of the guilty knowledge test. *Neuropsychologia, 22,* 457–469.

Bladon, R. A. W., & Al-Bamerni, A. (1976). Coarticulation resistance in English /1/. *Journal of Phonetics, 4,* 137–150.

Bricker, P. D., & Pruzansky, S. (1966). Effects of stimulus content and duration on talker identification. *Journal of the Acoustical Society of America, 40,* 1441–1449.

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.

Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication, 15,* 39–54.

Carney, A. E., Widin, G. E., & Viemeister, N. F. (1977). Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America, 62,* 961–970.

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

Catford, J. C. (1988). *A practical introduction to phonetics.* New York: Oxford University Press.

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

Clarke, F. R., Becker, R. W., & Nixon, J. C. (1966). *Characteristics that determine speaker recognition* (Report ESD-TR-66-638). Hanscom Field, MA: Electronic Systems Division, Air Force Systems Command.

Coleman, R. O. (1973). Speaker identification in the absence of inter-subject differences in glottal source characteristics. *Journal of the Acoustical Society of America, 53,* 1741–1743.

Compton, A. J. (1963). Effects of filtering and vocal duration upon the identification of speakers, aurally. *Journal of the Acoustical Society of America, 35,* 1748–1752.

Cutler, A. (1989). Auditory lexical access: Where do we start? In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 342–356). Cambridge, MA: MIT Press.

Cutler, A., & Norris, D. G. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 113–121.

De Haan, E. F., Young, A. W., & Newcombe, F. (1991). Covert and overt recognition in prosopagnosia. *Brain, 114,* 2575–2591.

Doherty, E. T., & Hollien, H. (1978). Multiple-factor speaker identification of normal and distorted speech. *Journal of Phonetics, 6,* 1–8.

Dommelen, W. A. van. (1987). The contribution of speech rhythm and pitch to speaker recognition. *Language and Speech, 30,* 325–338.

Dommelen, W. A. van. (1990). Acoustic parameters in human speaker recognition. *Language and Speech, 33,* 259–272.

Endres, W., Bambach, W., & Flösser, G. (1971). Voice spectrograms as a function of age, voice disguise and voice imitation. *Journal of the Acoustical Society of America, 49,* 1842–1848.

Fant, G. (1966). *A note on vocal tract size factors and nonuniform F-pattern scaling* (Speech Transmission Laboratory, Quarterly Progress and Status Report, 4/66). Stockholm, Sweden: Royal Institute of Technology. (Reprinted in *Speech sounds and features,* pp. 84–93, by C. G. M. Fant, 1973, Cambridge, MA: MIT Press)

Fitch, H. L., Halwes, T., Erickson, D. M., & Liberman, A. M. (1980). Perceptual equivalence of two acoustic cues for stop-consonant manner. *Perception & Psychophysics, 27,* 343–350.

Forster, K. I. (1976). Accessing the mental lexicon. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms* (pp. 257–287). Amsterdam: North-Holland.

Fowler, C. A., Rubin, P. E., Remez, R. E., & Turvey, M. T. (1980). Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Language production: Vol. 1. Speech and talk* (pp. 373–420). New York: Academic Press.

Furui, S. (1978). Effects of long-term spectral variability on speaker recognition. *Journal of the Acoustical Society of America, 64,* S183.

Gelfer, M. P. (1988). Perceptual attributes of voice: Development and use of rating scales. *Journal of Voice, 2,* 320–326.

Glenn, J. W., & Kleiner, N. (1968). Speaker identification based on nasal phonation. *Journal of the Acoustical Society of America, 43,* 368–372.

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

Goldstein, U. (1976). Speaker-identifying features based on formant tracks. *Journal of the Acoustical Society of America, 59,* 176–182.

Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal subjects and amnesic patients. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 501–518.

Granström, B. (1992). The use of speech synthesis in exploring different speaking styles. *Speech Communication, 11,* 347–355.

Granström, B., & Nord, L. (1991). Ways of exploring speaker characteristics and speaking styles. *Actes du XIIIième Congrès International des Sciences Phonetique* (Vol. 4, pp. 278–281). Aix-en-Provence, France: International Congress of Phonetic Sciences.

Hadding-Koch, K., & Studdert-Kennedy, M. (1964). An experimental study of some intonation contours. *Phonetica, 11,* 175–185.

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.

Halle, M., Hughes, G. W., & Radley, J.-P. A. (1957). Acoustic properties of stop consonants. *Journal of the Acoustical Society of America, 29,* 107–116.

Hecker, M. (1971). Speaker recognition: An interpretive survey of the literature. *ASHA Monographs, 16.*

Hollien, H., & Klepper, B. (1984). The speaker identification problem. In *Advances in forensic psychology and psychiatry* (pp. 87–111). Norwood, NJ: Ablex.

Hollien, H., Majewski, W., & Doherty, E. T. (1982). Perceptual identification of voices under normal, stress and disguise speaking conditions. *Journal of Phonetics, 10,* 139–148.

House, A. S., & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America, 25,* 105–113.

Jassem, W. (1971). Pitch and compass of the speaking voice. *Journal of the International Phonetic Association, 1,* 59–68.

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika, 32,* 241–254.

Jusczyk, P. W., Hohne, E. A., Jusczyk, A. M., & Redanz, N. J. (1993). Do infants remember voices? *Journal of the Acoustical Society of America, 94,* 2373.

Klatt, D. H. (1980). Speech perception: A model of acoustic-phonetic analysis and lexical access. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 243–288). Hillsdale, NJ: Erlbaum.

Klatt, D. H. (1985). A shift in formant frequencies is not the same as a shift in the center of gravity of a multiformant energy concentration. *Journal of the Acoustical Society of America, 77,* S7.

Klatt, D. H. (1989). Review of selected models of speech perception. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 169–226). Cambridge, MA: MIT Press.

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika, 29,* 1–27.

Labov, W. (1972). *Sociolinguistic patterns.* Philadelphia: University of Pennsylvania Press.

Labov, W. (1986). Sources of inherent variation in the speech process. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 402–425). Hillsdale, NJ: Erlbaum.

Ladefoged, P. (1967). *Three areas of experimental phonetics.* London: Oxford University Press.

Ladefoged, P., & Ladefoged, J. (1980). The ability of listeners to identify voices. *UCLA Working Papers in Phonetics, 49,* 43–51.

Laver, J. (1980). *The phonetic description of voice quality.* Cambridge, England: Cambridge University Press.

Lieberman, P. (1965). On the acoustic basis of the perception of intonation by linguists. *Word, 21,* 40–54.

Lisker, L. (1978). Rabid vs. rapid: A catalog of acoustic features that may cue the distinction (Status Report on Speech Research SR-54, pp. 127–132). New Haven, CT: Haskins Laboratories.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20,* 384–422.

Lively, S. E., Pisoni, D. B., & Goldinger, S. D. (1994). Spoken word recognition: Research and theory. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 265–301). New York: Academic Press.

Luce, P. A., Pisoni, D. B., & Goldinger, S. D. (1990). Similarity neighborhoods of spoken words. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 122–147). Cambridge, MA: MIT Press.

Lummis, R. C. (1973). Speaker verification by computer using speech intensity of temporal registration. *IEEE Transactions on Audio and Electroacoustics, 21,* 50–59.

Mann, V. A., Diamond, R., & Carey, S. (1979). Development of voice recognition: Parallels with face recognition. *Journal of Experimental Child Psychology, 27,* 153–165.

Markel, J. D., Oshika, B. T., & Gray, A. H. (1977). Long term feature averaging for speaker recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing, 25,* 330–337.

Marslen-Wilson, W. (1984). Function and process in spoken word recognition—A tutorial review. In H. Bouma & D. Bouwhuis (Eds.), *Attention and performance X* (pp. 125–150). Hillsdale, NJ: Erlbaum.

Matsumoto, H., Hiki, S., Sone, T., & Nimura, T. (1973). Multidimensional representation of personal quality and its acoustic correlates. *IEEE Transactions on Audio and Electroacoustics, 21,* 428–436.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception. Part 1: An account of basic findings. *Psychological Review, 88,* 375–407.

McDonough, J., Ladefoged, P., & George, H. (1993). Navajo vowels and phonetic universal tendencies. *UCLA Working Papers in Phonetics, 84,* 143–150.

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.

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

Nearey, T. M. (1978). *Phonetic feature systems for vowels.* Bloomington, IN: Indiana University Linguistics Club.

Nolan, F. (1983). *The phonetic bases of speaker recognition.* Cambridge, England: Cambridge University Press.

Nygaard, L. C., & Kalish, M. L. (1994). Modeling the effect of learning voices on the perception of speech. *Journal of the Acoustical Society of America, 95,* 2873.

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

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1995). Effects of stimulus variability on perception and representation of spoken words in memory. *Perception & Psychophysics, 57,* 989–1001.

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.

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

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research, 29,* 434–446.

Pollack, I., Pickett, J. M., & Sumby, W. H. (1954). On the identification of speakers by voice. *Journal of the Acoustical Society of America, 26,* 403–406.

Remez, R. E. (1994). A guide to research on the perception of speech. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 145–172). New York: Academic Press.

Remez, R. E., Pardo, J. S., & Rubin, P. E. (1992, November 13). *On the independence of phonetic and auditory perception.* Paper presented at the 33rd annual meeting of the Psychonomic Society, St. Louis, Missouri.

Remez, R. E., & Rubin, P. E. (1984). Perception of intonation in sinusoidal sentences. *Perception & Psychophysics, 35,* 429–440.

Remez, R. E., & Rubin, P. E. (1993). On the intonation of sinusoidal sentences: Contour and pitch height. *Journal of the Acoustical Society of America, 94,* 1983–1988.

Remez, R. E., & Rubin, P. E. (in press). Acoustic shards, perceptual glue. In J. Charles-Luce, P. A. Luce, & J. R. Sawusch (Eds.), *Theories in spoken language: Perception, production, and development.* Norwood, NJ: Ablex.

Remez, R. E., Rubin, P. E., Berns, S. M., Pardo, J. S., & Lang, J. M. (1994). On the perceptual organization of speech. *Psychological Review, 101,* 129–156.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science, 212,* 947–950.

Rubin, P. E. (1980). Sinewave synthesis [Internal memorandum]. New Haven, CT: Haskins Laboratories.

Sambur, M. R. (1975). Selection of acoustic features for speaker identification. *IEEE Transactions on Acoustics, Speech and Signal Processing, 23,* 176–182.

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. A. (1992). Auditory priming: Implicit and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 915–930.

Segui, J. (1984). The syllable: A basic perceptual unit in speech processing? In H. Bouma & D. Bouwhuis (Eds.), *Attention and performance X* (pp. 165–181). Hillsdale, NJ: Erlbaum.

Sheffert, S. M., & Fowler, C. A. (1995). The effect of voice and visible speaker change on memory for spoken words. *Journal of Memory and Language, 34,* 665–685.

Siegel, S., & Castellan, N. J., Jr. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill.

Silverman, K. (1985). Vowel intrinsic pitch influences the perception of intonational prominence. *Journal of the Acoustical Society of America, 77,* S38.

Su, L.-S., Li, K.-P., & Fu, K. S. (1974). Identification of speakers by use of nasal coarticulation. *Journal of the Acoustical Society of America, 56,* 1876–1882.

Tartter, V. C. (1980). Happy talk: Perceptual and acoustic effects of smiling on speech. *Perception & Psychophysics, 27,* 24–27.

Tartter, V. C., & Braun, D. (1994). Hearing smiles and frowns in normal and whisper registers. *Journal of the Acoustical Society of America, 96,* 2101–2107.

Trudgill, P. (1974). *Sociolinguistics: An introduction.* Harmondsworth, Middlessex, England: Penguin.

Van Lancker, D. (1991). Personal relevance and the human right hemisphere. *Brain and Cognition, 17,* 64–92.

Van Lancker, D. R., Cummings, J. L., Kreiman, J., & Dobkin, B. H. (1988). Phonagnosia: A dissociation between familiar and unfamiliar voices. *Cortex, 24,* 195–209.

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

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

Wells, J. C. (1982). *Accent of English 3: Beyond the British Isles.* Cambridge, England: Cambridge University Press.

Williams, C. E. (1964). *The effects of selected factors on the aural identification of speakers* (Section III, Report ESD-TDR-65-153). Hanscom Field, MA: Electronic Systems Division, Air Force Systems Command.

Zue, V. W., & Schwartz, R. M. (1980). Acoustic processing and phonetic analysis. In W. A. Lea (Ed.), *Trends in speech recognition* (pp. 101–124). Englewood Cliffs, NJ: Prentice Hall.

Received February 1, 1995
Revision received October 3, 1995
Accepted January 18, 1996 ■