In P. K. Kuhl and L. A. Crum (Eds.) *Proceedings of the 16th International Congress on Acoustics and the 135th Meeting of the Acoustical Society of America*, Volume 4 (pp. 2063-2064). New York: Acoustical Society of America.

# On the Perception of Qualitative and Phonetic Similarities of Voices

Robert E. Remez*, Jennifer L. Van Dyk*, Jennifer M. Fellowes† & Philip E. Rubin‡

*Department of Psychology, Barnard College, 3009 Broadway, New York, New York 10027,
†College of Physicians & Surgeons, Columbia University, 630 West 168th Street, New York, New York 10032
‡Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511

**Abstract.** A perceiver who learns to recognize a talker becomes familiar with attributes of the talker's voice that are present in any utterance regardless of the linguistic message. Customary accounts of individual identification presume that such durable personal aspects of an individual's speech are independent of the acoustic properties that evoke segmental phonetic contrasts. Alternatively, some classic and recent studies alike suggest that familiarity includes attention to attributes of dialect and idiolect conveyed in the articulation of consonants and vowels. The present investigation sought direct evidence of attention to phonetic attributes of speech in identifying talkers. Natural samples and sinewave replicas of sentences were used in a perceptual similarity tournament establishing the resolution of phonetic attributes in the perception of talkers.

The identification of linguistic and personal attributes from a speech sample are customarily considered to occur independently (1, 2). Characterizations of linguistic analysis have typically included a process of abstraction by which a perceiver discovers a linguistic form within the acoustic variation of the signal, whether the variation is attributable to paralinguistic aspects of an utterance or to affective or anatomical properties of a talker. Analogously, the identification of a talker from a speech sample has been taken to warrant a perceptual process to ignore the specific acoustic details of a speech signal that promote the identification of segmental and lexical distinctions. Instead, attention focuses on acoustic attributes of a talker's voice common to all utterances, specific to none. Studies have implicated apparently basic auditory attributes, such as the pitch and compass of phonatory frequency or vocal timbre, as well as more elaborated attributes, such as vocal strength or melodiousness, in this regard.

Recent refinements of this disjunctive approach have included specification of the acoustic attributes unique to a female voice (3) independent of linguistic contrasts, and estimation of the effects on phoneme quality judgments of parametric variation in vowel spectra (4) independent of talker attributes. Neuropsychological evidence also shows that different cortical areas may be dedicated to each function (5), linguistic analysis and individual identification, tempting speculation that each process is fed by different sensory attributes: short-term elements pertain to symbolic contrasts, and long-term to distinctions among talkers.

Our recent observations (6) require a revision in this approach to individual and linguistic identification, and suggests that similar attributes of speech underlie the perception of words and talkers. Our project showed that talkers were well identified, both by strangers and by acquaintances, under listening conditions in which the acoustic correlates of voice quality were eliminated from the speech samples. Indeed, we also found that individual identification survived acoustic conditions preventing the perceptual determination of a talker's sex (7). The use of sinewave replicas of natural samples permitted such tests, in which a tone was set equal in frequency and amplitude to each of the three lowest oral resonance peaks. A three-tone replica of a natural utterance lacks vocal timbre, though it evidently preserves the spectrotemporal variation sufficient to elicit phonetic impressions. Our results were sensible on the hypothesis that individual identification exploits coarse grain attributes of the familiar kind—average fundamental frequency or glottal spectrum—and phonetic attributes as well. Because phonetic attributes differ among talkers, manifesting contrasts in dialect and idiolect, phonetic attributes in aggregate are available for distinguishing among talkers perceptually. Of course, phonetic segments are also useful in identifying spoken words.

Our claim that talkers were identifiable without recourse to acoustic correlates of qualitative attributes of the voice was based on direct albeit opportunistic evidence. Performance levels in our perceptual tests (6, 7) were high, especially in a condition in which listeners were personally acquainted with the talkers. By comparing the distribution of errors of identification across talkers, we were able to estimate the perceptual similarity within our set of talkers. The present project posed this question with a sharper point: Does the perception of a talker's identity depend on natural acoustic correlates of voice quality

## THE PRESENT TEST—A SIMILARITY TOURNAMENT

The acoustic test materials consisted of two sentences (The drowning man let out a yell; The scarves were made of shiny silk.) spoken by ten talkers (5 male, 5 female). There were two versions of every sentence, natural or sinewave. A listener in the similarity tournament heard ten talkers speak natural or sinewave versions of one of the sentences only. No listener heard both natural and sinewave samples. We asked 104 listeners to participate in a

perceptual similarity tournament; each was assigned to one of the four acoustic conditions. On each trial, two versions of the same sentence were presented, each spoken by a different talker, and a listener indicated the apparent similarity of the talkers on a 5 point scale ranging from VERY SIMILAR to NOT VERY SIMILAR.

How do listeners perceive similarity among a set of ten talkers? According to the classic account based upon fundamental frequency and glottal spectrum, we must predict that perceptual outcome would be quite different between a test using natural samples and a test using sinewave replicas of natural utterances, because the timbre of natural speech differs greatly from the timbre of three harmonically unrelated tones sounded concurrently. If a component of the identification of talkers includes phonetic attributes, perhaps as an impression of dialect or idiolect, or a global articulatory parameter (8), we would predict far less difference in the test outcome under these drastically different acoustic conditions. The results are shown in a multidimensional scaling analysis for each of the acoustic signal types, natural and sinewave, in Figure 1. This analysis produced a graphic representation of the relative similarities that we derived by collapsing over the two sentences or each of acoustic type. Fifty-two listeners contributed to each panel of Figure 1.

The outcome of the test shows, first, that the male talkers were judged to be highly similar to each other, and females to be less similar to each other, regardless of acoustic conditions. Although there are some small differences in the scatter of values across the plot frame, the overall impression of similarity despite the acoustic differences in the samples is borne out by a test of the ranked similarities in a hierarchical clustering analysis. The ranks were highly correlated, consistent with the conjecture that listeners based their similarity judgments on properties of the speech that were present in both acoustic vehicles, the natural and the sinewave. This finding extends the conclusion of our prior studies (6, 7) that qualitative impressions of a talker's voice are based no less on the phonetic attributes of speech than on the acoustic properties held conventionally to elicit an experience of vocal timbre.

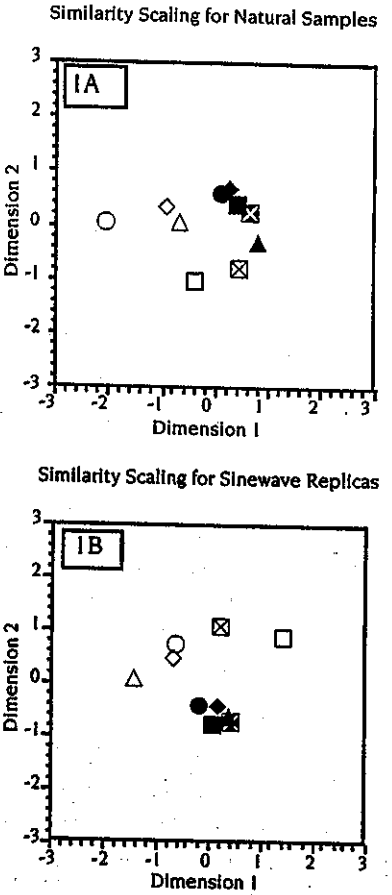


**Figure 1.** Multidimensional scaling of perceived talker similarities: 1A) natural; 1B) sinewave.

## ACKNOWLEDGMENTS

The authors thank Dalia Shoretz for assistance in analyzing the findings of this project. This research was supported by grants from the NIH (DC00308 to Barnard College and HD01994 to Haskins Laboratories).

## REFERENCES

1. Halle, M. 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 (1985).
2. Bricker, P. D., & Pruzansky, S. Speaker recognition. In N. J. Lass (Ed.), *Contemporary Issues in Experimental Phonetics* (pp. 295-326). New York: Academic Press (1976).
3. Klatt, D. H., & Klatt, L. C. Analysis, synthesis and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America* **87**, 820-857 (1990).
4. Frieda, E., Walley, A., Flege, J., & Sloane, M. Adults perception of native and nonnative vowels: Implications for the perceptual magnet effect. *Perception & Psychophysics* (in press).
5. Van Lancker, D., Cummings, J. L., Kreiman, J., & Dobkin, B. H. Phonagnosia: A dissociation between familiar and unfamiliar voices. *Cortex* **24**, 195-209 (1988).
6. Remez, R. E., Fellowes, J. M., & Rubin, P. E. Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance* **23**, 651-666 (1997).
7. Fellowes, J. M., Remez, R. E., & Rubin, P. E. Perceiving the sex and identity of a talker without natural vocal timbre, *Perception & Psychophysics*, **59**, 839-849 (1997).
8. Laver, J. *The Phonetic Description of Voice Quality.* Cambridge, England: Cambridge University Press (1980).