SPEECH COMMUNICATION:

THE HUMAN VOICE

Jody E. Kreiman
Division of Head and Neck Surgery, UCLA School of Medicine
Los Angeles, California 90095–1794

Patricia A. Keating
Department of Linguistics, University of California, Los Angeles
Los Angeles, California 90095–1543

David A. Berry
The Laryngeal Dynamics Laboratory, UCLA School of Medicine
Los Angeles, California 90095–1794

The Speech Communication Technical Committee (SCTC) includes about 950 scientists within the Acoustical Society of America (ASA) who share an interest in the production, transmission, and perception of spoken language. This interest comprises a broad range of research topics, including the acoustic, physiological, psychological, and linguistic phenomena related to human speech processes; the properties of speech transmission systems; machine processing of speech, including speech analysis, synthesis and automatic recognition; and the measurement and assessment of speech as to its intelligibility and its quality. Thus, members of the SCTC come from many different disciplines, including at least physics, speech and hearing science, experimental psychology, linguistics, electrical and mechanical engineering, music, communication disorders, and otolaryngology.

One area of recent cross-disciplinary emphasis within the SCTC is the study of the human voice. Researchers have long known that voice conveys significant amounts of information about speakers. Speakers may sound young, or tired, or elated, or distracted. They may sound as if they are drunk, or lying, or ill, or bearing secret, exciting news. By their voices, adult speakers usually reveal whether they are male or female, and in addition, they may signal that they come from Texas, or Wisconsin, or France. Over the telephone, we may recognize the speaker as someone we know, or we may form a distinct impression of the physical appearance of someone we have never seen. The impressions listeners gain from voices are not necessarily accurate; for example, everyone has known the surprise of meeting a telephone acquaintance who does not match the mental picture we have previously formed of them. Despite such occasional mismatches, voice quality is one of the primary means by which speakers project their identity—their "physical, psychological, and social characteristics" (Laver, 1980)—to the world.

Beyond these paralinguistic functions, it has also become increasingly apparent that voice quality plays a number of important linguistic roles. In the same way that voice quality can be controlled and varied by an individual to convey emotions and attitudes, it can be varied to give information about the structure of long utterances within a conversation or discourse. For example, "creaking" the voice at the end of a long sentence or paragraph-sized utterance can signal that the person speaking has finished, and that another speaker can now take a turn. This sort of use of voice quality is probably characteristic of most languages in the world. A very different use of voice quality occurs in some languages that make distinctions among words partly on the basis of the voice quality used. For example, Peter Ladefoged (recipient of the ASA Silver Medal in Speech Communication) reported that in the language Mazatec (as spoken in Jalapa, Mexico), the words for "buttocks" and "horse" have the same consonants and vowel–something like "nda"–but the first must be said with a creaky voice while the second must be said with a breathy voice (Kirk, Ladefoged, and Ladefoged, 1993). [To hear these and other examples of this sort from several languages, go to www.phonetics.ucla.edu/index/sounds.html and choose the categories "breathy voice" and "creaky voice".] Because languages that use voice quality to distinguish meanings of words in this way tend to be in some danger of disappearing, it is likely that opportunities for research on this kind of voice quality use will decrease in the future. [For information about endangered languages, see for example www.yourdictionary.com/elr/ or www.sil.org/sociolx/ndg-lg-home.html.] Finally, substantial evidence indicates that familiarity with a talker's voice facilitates deciphering the spoken message itself (e.g., Goldinger, Pisoni, and Logan, 1991; Nygaard and Pisoni, 1998).

Recent advances in the study of voice production and laryngeal biomechanics provide insights into the physiology of voice production as well. For example, through high-speed imaging of the medial surface of the vocal folds during phonation, it has been shown that most vibration patterns of the folds consist of at least two dominant modes of vibration. During modal phonation, such modes are entrained to a common fundamental frequency. However, during creaky voice or other nonmodal phonation types, the underlying modes of vibration may exhibit complex entrainment patterns or may not entrain at all, resulting in complex, nonlinear behavior. As understanding of perceptual and biomechanical processes increases, it may be possible to describe *how* listeners derive the information they do from voice signals, and why certain cues emerge as salient.

Links to many other sites describing these and other research foci of the Speech Communication Technical Committee can be found on the SCTC website at sal.shs.arizona.edu/~asaspeechcom/. **AT**

References and Further Reading

Goldinger, S.D., Pisoni, D.B., and 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.

Kirk, P.L., Ladefoged, J., and Ladefoged, P. (**1993**). "Quantifying acoustic properties of modal, breathy and creaky vowels in Jalapa Mazatec," in *American Indian Linguistics and Ethnography in Honor of Laurence C. Thompson*, edited by A. Mattina and T. Montler (University of Montana Press, Missoula, MT).

Laver, J. (**1980**). *The Phonetic Description of Voice Quality* (Cambridge, Cambridge University Press).

Nygaard, L.C., and Pisoni, D.B. (**1998**). "Talker-specific learning in speech perception," Perception and Psychophysics **60**, 355-376.