Vocal Cues in Emotion Encoding and Decoding

Klaus R. Scherer\textsuperscript{2} and Rainer Banse
University of Geneva

Harald G. Wallbott and Thomas Goldbeck
University of Giessen

This research examines the correspondence between theoretical predictions on vocal expression patterns in naturally occurring emotions (as based on the component process theory of emotion; Scherer, 1986) and empirical data on the acoustic characteristics of actors' portrayals. Two male and two female professional radio actors portrayed anger, sadness, joy, fear, and disgust based on realistic scenarios of emotion-eliciting events. A series of judgment studies was conducted to assess the degree to which judges are able to recognize the intended emotion expressions. Disgust was relatively poorly recognized; average recognition accuracy for the other emotions attained 62.8\% across studies. A set of portrayals reaching a satisfactory level of recognition accuracy underwent digital acoustic analysis. The results for the acoustic parameters extracted from the speech signal show a number of significant differences between emotions, generally confirming the theoretical predictions.

Research on the vocal expression of emotion lags significantly behind the study of facial affect expression. The reasons for this relative neglect are manifold (see Scherer, 1982, 1986, for a detailed discussion of this problem). One of the most important factors is the difficulty of obtaining

\textsuperscript{1}This research was supported by a grant from the Deutsche Forschungsgemeinschaft (Sche 156/8-5). The authors acknowledge collaboration of Westdeutscher Rundfunk, Cologne, in producing professional versions of actor emotion portrayals, and thank Kurt Balser, Theodor Gehm, Christoph Gierschner, Judy Hall, Ursula Hess, Alice Isen, and one anonymous reviewer for helpful comments and contributions.

\textsuperscript{2}Address all correspondence, including requests for reprints, to Klaus R. Scherer, University of Geneva, FPSE, 24 rue General Dufour, 1211 Geneva 4, Switzerland.
real-life records of vocal expression of specific emotions. It is generally true that access to naturalistic affect expression in public settings is highly limited (see Scherer, 1986; Wallbott & Scherer, 1986b). However, recent technological advances in photography and video recording (using telephoto lenses, unobtrusively positioned cameras, etc.) permit researchers to obtain candid shots of facial and gestural emotion expression samples. In contrast, the limited pick-up range of microphones as well as the pervasive noise problem make it difficult to "eavesdrop" on naturally occurring vocal affect expressions.

Similarly, the induction of strong and highly differentiated emotional states in the laboratory, where high-quality recording of expressions could be obtained, is rendered difficult by both practical and ethical concerns (see Scherer, 1986; Wallbott & Scherer, 1986b). Furthermore, vocal expression, just like facial expression, is subject to a high degree of social control (in the form of display rules and masking attempts; see Ekman & Friesen, 1969; Ekman, Friesen, & O'Sullivan, 1988; Zuckerman, DePaulo, & Rosenthal, 1981).

As a consequence, researchers interested in studying the different patterns of vocal cues accompanying the expression of particular emotional states cannot dispense with the use of actor portrayals of emotion. While this type of stimulus production obviously does not allow one to generalize the results to naturalistic expression of emotion, the systematic study of how actors manipulate their voices in order to communicate specific emotions and the analysis of observers' ability to use the respective vocal cues to infer the states of the speakers provide important corollary information for the study of vocal affect expressions (see Wallbott & Scherer, 1986b).

Only very few studies have attempted to measure objectively the acoustic cues used in vocal emotion expression (Fairbanks & Hoaglin, 1941; Van Bezooijen, 1984; Wallbott & Scherer, 1986b; Williams & Stevens, 1969, 1972). Unfortunately, such studies are beset by a number of methodological shortcomings which render the evaluation of the results rather difficult and often prevent the accumulation of findings. Some of the most serious problems will be briefly outlined below.

Sometimes only a single actor or actress is used for the portrayal, which makes it difficult to evaluate the effect of individual differences or of the type of training the actor has had. Gender differences between actors and actresses are rarely investigated. The conditions under which the emotions were portrayed, in particular the instructions given, are often not specified in sufficient detail.

Frequently actors are just asked to read a passage in such a way as to make the rendering sound representative of a particular emotion. A