

Perceiving affect from the voice and the face

DOMINIC W. MASSARO and PETER B. EGAN
University of California, Santa Cruz, California

This experiment examines how emotion is perceived by using facial and vocal cues of a speaker. Three levels of facial affect were presented using a computer-generated face. Three levels of vocal affect were obtained by recording the voice of a male amateur actor who spoke a semantically neutral word in different simulated emotional states. These two independent variables were presented to subjects in all possible permutations—visual cues alone, vocal cues alone, and visual and vocal cues together—which gave a total set of 15 stimuli. The subjects were asked to judge the emotion of the stimuli in a two-alternative forced choice task (either HAPPY OR ANGRY). The results indicate that subjects evaluate and integrate information from both modalities to perceive emotion. The influence of one modality was greater to the extent that the other was ambiguous (neutral). The fuzzy logical model of perception (FLMP) fit the judgments significantly better than an additive model, which weakens theories based on an additive combination of modalities, categorical perception, and influence from only a single modality.

Research has shown that we use multiple sources of information when we comprehend speech (Massaro, 1987b, 1989; Massaro & Cohen, 1990). Visual information from a speaker's face, for example, can strongly influence speech perception, especially when the auditory information is degraded: in one study, recognition of auditory sentences in noisy environments improved from 23% to 65% when the perceivers could also see the speaker's face (Summerfield, 1979). We also use multiple sources of information when we perceive a speaker's emotion. These sources include a variety of paralinguistic signals, as well as the speech's verbal content. The emotion may be interpreted in different ways, depending on the voice quality, facial expression, and body language of the speaker. To study the degree to which paralinguistic sources of information are used, it is important that one first define these sources and then determine how they are evaluated and integrated. In the present study, in order to investigate the perception of a speaker's emotion, two sources of paralinguistic information were varied: facial expressions and vocal cues.

Facial expressions are an effective means of communicating emotion. Darwin (1872) argued that facial expressions originate in basic acts of self-preservation common to human beings and other animals, and that these expressions are related to the emotional states that they convey. Research by Meltzoff and Moore (1977) suggests that we are biologically prepared from birth to respond to facial expressions. They produced evidence which showed that

infants as young as 12 days old are able to imitate adult facial expressions. Although many possible facial movements are possible, only a few such combinations are used to communicate emotions (Ekman, 1984). In other research, Ekman (1993) concluded that the processing of affect is particularly well developed in humans, who appear to be able to recognize and characterize facial expressions of emotional affect in other humans with a high degree of accuracy and consistency. Studies of animals, children born deaf and blind, and preliterate and isolated societies have provided evidence that some facial expressions, such as rage, startle, fear, and pleasure, may be universal. But these findings are not conclusive, and other research (Fridlund, 1991, 1994) suggests that behavioral and ecological factors have been the main influence in the development of facial expressions. As an example, we might ask why smiling is usually interpreted as a friendly gesture. Smiling while speaking raises the fundamental frequency (F0) and formant frequencies of speech, which can be interpreted as coming from a smaller and less threatening organism (Ohala, 1984).

Tanaka and Farah (1993) have found that individual facial features are recognized more easily when displayed as part of a whole face than when displayed in isolation. They concluded that facial recognition is a holistic process. However, *holistic processing* is a loaded term that is difficult to quantify. Etcoff and Magee (1992) tested subjects on computer-averaged faces, which differed by constant increments along a dimension of emotional affect. Given the similarity of their results to previous findings of "categorical perception," they concluded that these facial expressions were perceived categorically. However, this type of result is not necessarily evidence for categorical perception (Massaro, 1987a). Recently Ellison and Massaro (1995) tested models of recognition of facial affect in an attempt to overcome the limitations of verbal theories. They used a set of stimuli that would be standardized and replicable,

This research was supported, in part, by grants from the the National Institute on Deafness and Other Communication Disorders, the National Institutes of Health (2 R01 DC00236-13A1), the National Science Foundation (BNS 8812728), and the University of California, Santa Cruz. The authors thank Michael M. Cohen for help at all stages of this research. Correspondence should be addressed to D. W. Massaro, Program in Experimental Psychology, Clark Kerr Hall, University of California, Santa Cruz, CA 95064 (e-mail: massaro@fuzzy.ucsc.edu).

