Voice and Emotion

6. Voice and emotion

ARVID KAPPAS, URSULA HESS, AND KLAUS R. SCHERER

. . . with many kinds of animals, man included, the vocal organs are efficient in the highest degree as a means of expression.
– Darwin 1872/1965, p. 83

Approaches to the study of vocal affect communication

In a cross-cultural survey on emotional reactions, with 27 countries on all five continents, changes in the voice were reported for all the emotions studied (Wallbott & Scherer, 1986b). This finding should come as no surprise, given Darwin's early claim concerning the importance of vocal indicators of affective state for the communication of emotion. Yet, it took decades for the systematic study of human nonverbal vocal communication to emerge, unlike the study of facial expression, which was already pursued quite intensively at the beginning of this century (see Ekman, Friesen, & Ellsworth, 1982). Impeding early research on vocal phenomena was the very nature of vocalization: It is difficult to capture or to store. Although it has been possible to freeze both facial expressions and posture in paintings and sculptures or by using early photographic techniques, only the invention of phonographic records and magnetic tape, as well as the development of electroacoustic analysis facilities (Hollien, 1981), has made research on vocal cues possible at all. In recent years the development of digital storage and analysis methods have been made feasible through advanced computer technology (Scherer, 1982; Wolf, 1981).

Apart from the problem of storing the vocal signal, the early research was hampered by the lack of a concise and reliable terminology to specify the phenomenon. The terms used in everyday language to describe voice and speech characteristics are frequently confounded with de-

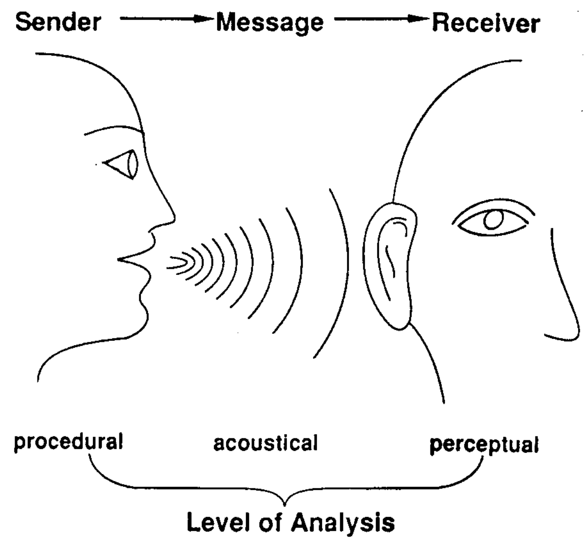

Figure 6.1. The relationship of different levels of analysis of vocal communication to Shannon and Weaver's model of communication.

scriptors of speaker state or are too general to apply to a specific vocal feature. Frequently, the categories used to denote vocal features (e.g., "harsh") refer to a gestalt that cannot be easily characterized in terms of objective acoustic parameters, such as mean fundamental frequency or spectral energy distribution ratios. Although there will be significant differences for some of these parameters between a vocalization that is perceived as harsh and one that is not perceived as such, these might not be solely responsible for the subjective experience of harshness. In outlining the possibilities for the analysis of vocal communication, it is helpful to distinguish among three levels or domains of description (see, e.g., Titze, 1974):

1. The production (procedural) level, referring to the state of the human vocal apparatus at the time an utterance was produced (e.g., subglottal pressure, vocal cord length, vocalis contraction, soft palate position, vocal tract shape).
2. The acoustical level, referring to an objective representation of the voice signal as it travels through air (e.g., sound pressure level, fundamental frequency, glottal pulse shape, formant amplitude ratios, formant frequency ratios).
3. The perceptual level, referring to the descriptions provided by everyday language for the categorization of voices (e.g., loudness, pitch, vibrato, register, ring).

Although these three levels can easily be related to elements of a simple communication model à la Shannon and Weaver (1949; Figure 6.1),

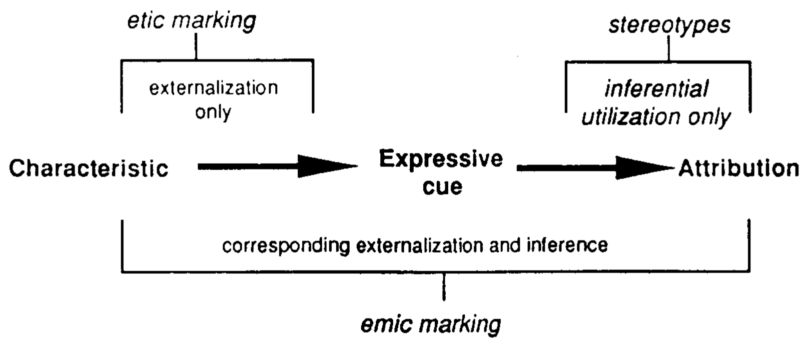

Figure 6.2. Etic and emic markers.

it is more appropriate to regard the sender → message → receiver distinction and the level of analysis as closely related but not identical. For example, it is possible both to perform acoustical analyses on emotional utterances and to relate particular acoustical features to the attribution processes of a receiver, thus remaining within the acoustical level of description yet analyzing sender as well as receiver processes. To clarify the distinction between the two processes and strategies of investigation, the concept of etic and emic markers is helpful (Giles, Scherer & Taylor, 1979; and Figure 6.2). *Emic markers* are behaviors that are both correlated with a particular state and recognized by conspecifics as being correlated with that state. *Etic markers*, on the other hand, while being correlated with an internal state, are not recognized by the interaction partner and therefore are not used in the attribution process. (Obviously, attributions may be based not only on actual markers but also on stereotypes, superstition, misperception, or other factors.)

To highlight the usefulness of this distinction, we shall consider two examples: If we are interested in how we can tell whether our conversation partner on the other end of the telephone is smiling (as we actually can; see Tartter, 1980), we do not need to investigate acoustic features not transmitted over the telephone line (such as energy components above 5 kHz). If, on the other hand, we are interested in using the vocal indicators of the effects of a psychoactive drug treatment as a diagnostic tool, we may not necessarily want to know whether anyone but the computer can tell the difference. Scherer suggested (e.g., 1978, 1982) a modified Brunswikian lens model (Brunswik, 1956) as helpful for separating these two very different classes of studies (Figure 6.3). In its simplest form the lens indicates an encoding process (externalization of a

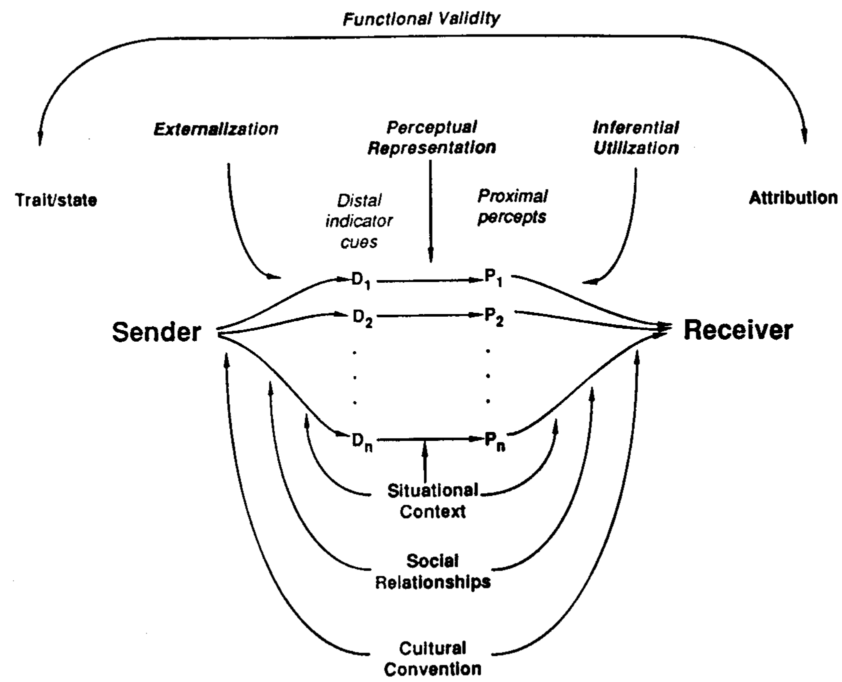

Figure 6.3. Modified version of the Brunswikian lens model.

speaker's state) and a decoding process (attribution of a receiver). Information may be transmitted through more than one channel, sequentially and in parallel, with different levels of redundancy. Indeed, communication is not context free but is embedded in a framework of cultural or social rules and information, shared by the participating individuals. Information transmission may be disturbed at any point of the process through noise. The model can simply be applied as an aid to distinguish conceptually among the processes in (nonverbal) communication and be tested empirically using path analytic methods (Scherer, 1978).

Depending on the level of analysis or description used, researchers have concentrated on particular questions regarding the communication of affective state. The fact that each domain requires a different expertise has discouraged many researchers from investigating the communication process at all three levels of description or at both the sender (encoding) and the receiver (decoding) side. Consequently, a fascinating accumulation of knowledge regarding specific phenomena emerged, but these bits and pieces have not yet been successfully integrated into a theory of the role of vocal parameters in the communication of affect, thus disappointing all those who had hoped for easy answers concerning the relationship of affective state and voice. Harrison and Wiemann described the study of communication in the nonverbal domain as follows: "It appears that the easy findings in nonverbal communication have been skimmed off by early explorers. What lies ahead is a difficult and often tedious task" (1983, pp. 279–280).

In this chapter we shall try to take stock of where we are in the endeavor to add vocal analysis to nonverbal communication research. Our research group, first at the University of Giessen and then at the University of Geneva, has specialized in this research domain over the past 15 years, and this chapter will outline our group's theoretical and methodological development. In addition, we shall review some of the major empirical results obtained in our own and other laboratories. In line with our predominant research concerns, much of our discussion will focus on the vocal communication of emotion. A brief review of the major production mechanisms of vocalization will lead into a summary of the work on vocal affect expression in animals, highlighting the phylogenetic continuity of this mode of communication. We shall then outline the major parameters of vocal analysis and describe the pattern of findings on the encoding or externalization of affect in the voice. Finally, we shall turn to the question of decoding, the ability of naive judges to infer emotion from the voice, with and without additional information from other channels. We shall conclude with an outlook that attempts to specify some of the promising perspectives for future research.

Effects of emotion on the voice

There is considerable evidence that emotional states have a direct impact on the production of both speech and nonspeech vocalization. Scherer (1979a) attempted to show the effects of affect-induced muscle tension changes on vocalization. Similarly, Ohala (1981) mentioned three bodily changes that have an impact on the sounds produced during vocalization:

1. Dryness in the mouth or larynx.
2. Accelerated breathing rate.
3. Muscle tremor.

There is not, as yet, a clear and exhaustive mapping of the bodily changes evoked by emotional states and the resulting vocal changes. Its

absence is mainly due to theoretical problems in defining what exactly emotions are and discriminating on a physiological basis among the states that we understand as basic or primary emotions. Although there is a growing consensus that there are physiological differentiations among some affective states (e.g., Campos, 1988; Davidson, 1988; Ekman, 1988; Levenson, 1988), it might not be possible to find autonomic signatures of emotional states as distinct as the facial expressive indicators of affective state. In this chapter, we shall review the evidence available today as well as advance some hypothetical predictions for further research.

To understand the effects of both affective states and voluntary control on vocalizations in humans, it is helpful to understand the nature of the vocalization process. We shall therefore first outline the mechanisms of voice production in humans and subhuman animals.

Mechanisms of voice and speech production

Both speech and non-speech vocalizations are composed of two separate mechanical processes: phonation and articulation. The flow of air through the vocal tract is the basis of all human vocalizations, and *phonation* refers to the production of sound waves, which are then modified or filtered by articulatory processes.

Next we shall describe the relevant mechanical concepts in speech production. A thorough discussion of the central pathways relevant to the control of phonation and articulation would be beyond the scope of this chapter, but interested readers are referred to Ploog (e.g., 1986, 1988) for more detailed information.

Phonation. The larynx functions as a valve regulating the flow of air in and out of the lungs and preventing food and drink from entering the lungs and in addition, it is specially adapted to act as a vibrator "chopping up" the airstream issued from the lungs. These two functions are accomplished by a complex arrangement of cartilages, muscles, and other tissues (Figure 6.4). The larynx is similar in all primates (Ploog, 1988).

The vibrating elements in the larynx are the vocal cords, folds along the lateral walls of the larynx that can be stretched and positioned by several specific muscles within the larynx itself *(intrinsic laryngeal muscles).* Each vocal fold extends between the thyroid cartilage and one of the arytenoid cartilages. Contraction of the posterior cricoarytenoid muscles

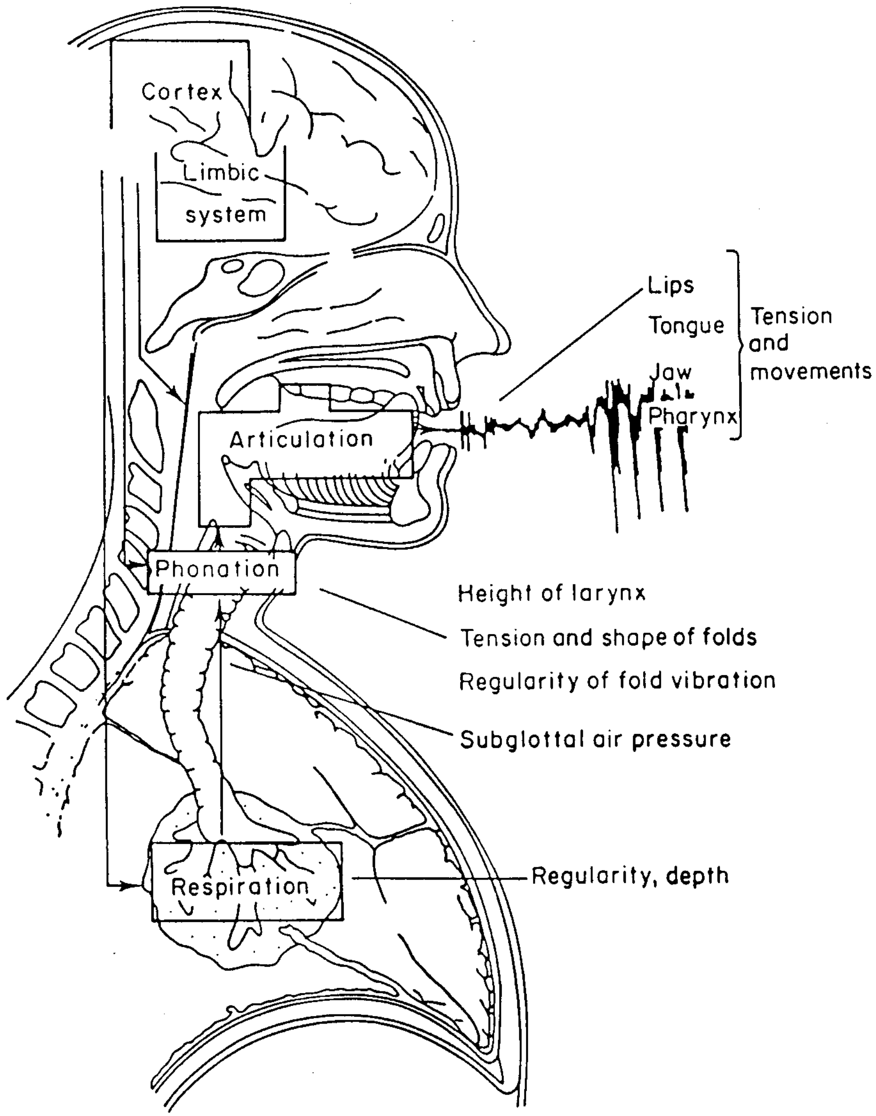

Figure 6.4. Overview of the voice production system and its major determinants (Scherer, 1989).

pulls the arytenoid cartilages away from the thyroid cartilage, thus stretching the vocal folds. The transverse arytenoid muscle pulls the arytenoid cartilages together, approximating the two vocal folds so that they can vibrate in a stream of expired air. Conversely, contraction of the lateral cricoarytenoid muscles pulls the arytenoid cartilages forward and apart to allow normal respiration. The thyroarytenoid muscles are made of groups of small muscle fibers, each controlled separately by different nerve fibers. The muscle fibers adjacent to the wall of the larynx and other individual portions also contract independently of one another.

These contractions control the shape of the vocal folds during the different types of phonation. The folds may be thick or thin or may have sharp or blunt edges. The different shapes and distances between the folds result in various acoustic changes. The use of computer models has helped considerably in estimating the acoustical changes caused by the shape of the vocal folds (e.g., Titze & Talkin, 1979). The farther apart the folds are while vibrating, the more air is allowed to rush through, increasing the amount of noise in the speech signal. The resulting phonation type is described as "whispery" or, in extreme cases, "breathy" (Laver, 1980). If the folds are tensely contracted, the effect is typically perceived as "harsh." If both folds are very close together and only a small part is allowed to vibrate while the other parts are firmly closed, a falsetto phonation results. The shape of the vocal folds during phonation can be studied with a laryngoscope – which in its simplest form is a 45° mirror with a light source attached to it – that can be inserted into the pharynx to allow visual inspection of the larynx.

The rapid opening and closing action of the vocal folds creates a vibratory pattern whose frequency constitutes the fundamental frequency (F_0) of speech. The fundamental frequency is an important aspect of the perception of the pitch of voice and can be changed in at least two very different ways. First, it can be changed by stretching or relaxing the vocal folds either by the mechanism of the intrinsic muscles of the larynx, as described, or by the muscles attached to the external surface of the larynx: These muscles can pull against the cartilages and thereby help stretch or relax the folds. The action of the external laryngeal muscles is usually combined with an upward or downward movement of the larynx to control pitch of voice. The second means for changing fundamental frequency is to change the shape and mass of the vocal fold edges. When very high frequency sounds are emitted, different parts of the thyroarytenoid muscles contract so that the edges of the vocal fold are sharpened and thinned, whereas when lower-frequency sounds are emitted, the thyroarytenoid muscles contract in a different pattern so that broad edges with a large mass are approximated. There has been some debate about the importance of subglottal pressure for determining the fundamental frequency. It seems that although subglottal pressure changes are proportional to fundamental frequency changes, the frequency changes are too small to account for subglottal pressure as a major determinant of pitch variation (Ohala, 1978).

Articulation. The sound impulses emanating from the glottal folds are filtered by the vocal tract which, according to its shape, amplifies certain frequencies (resonance frequencies) and dampens other frequencies. The major organs of articulation are the lips, the tongue, and the soft palate. The resonators include the mouth, the nose and associated nasal sinuses, the pharynx, and the chest cavity itself (Figure 6.4). By changing the position of the articulators, the shape of the vocal tract as a whole changes so that the sound waves, depending on their frequencies, are selectively amplified or attenuated to produce different vowels or are obstructed to produce stops, fricatives, and plosives. It is in the size, shape, and details of the articulators that humans differ from all other primates. In articulation, as in phonation, it is frequently possible to achieve the same acoustical result by employing different means. And it is therefore not commonly possible to decide on the basis of a single acoustical parameter the state of articulators and laryngeal settings.

Vocal expression of affective state in animals

The study of vocalizations in animals is of particular interest for the understanding of vocal nonverbal communication, because it is likely that many of the processes involved are phylogenetically continuous (Darwin, 1872; Ekman, 1973; Ploog, 1986; Scherer, 1979a, 1985). Ample evidence has been collected indicating that most animal vocalizations express the motivational and emotional state of the individual (Dittus, 1988; Eisenberg, 1974, 1976; Green, 1975; Jürgens, 1979, 1982, 1988; Jürgens & Ploog, 1976; Marler, 1984; Marler & Tenaza, 1977; Morton, 1977; Ploog, 1974; Seyfarth & Cheney, 1982; Tembrock, 1971, 1975; Todt, 1988). In a neuroethological approach, vocal behavior patterns are regarded as fixed-action patterns used as signals in social communication (Ploog, 1986). One possible gain from studying vocalizations of nonhuman species is to find cross-species universality and phylogenetic

continuity that allow insights into basic biological patterns of affect vocalization. Three notions are of specific interest in this context:

1. Structural continuity in the brain structures involved in vocalizations and affect.
2. Cross-species similarity in acoustic features of vocalizations because of physiological influences on the phonation and articulation mechanisms, given functionally similar antecedent events and consequentially similar behavioral patterns (such as preparatory actions for flight or fight when confronted with a dangerous aggressor).
3. Cross-species similarity of acoustic structures of vocalizations that might have developed as a function of specific ecological "demands" of the natural environment.

Development of control structures. Anatomical studies and studies employing electrical brain stimulation and brain lesions have found evidence for a progression from mostly automatic call productions in lower vertebrates to more voluntary vocal productions in higher vertebrates, particularly subhuman primates and humans (Ploog, 1986, 1988). Although the ability to speak seems to be unique to humans, many relevant structures are homologous in humans and monkeys, except for the direct motor pathway from the laryngeal representation in the primary motor cortex to the laryngeal motor neurons in the medulla, which does not exist in monkeys. On the basis of the neurological evidence Jürgens (e.g., 1988) concluded that monkeys seem to lack the capacity to communicate information unrelated to the momentary motivational state. (There is, however, evidence for representational vocalizations in infrahuman species, such as the alarm calls of vervet monkeys [Marler, 1984; Seyfarth & Cheney 1982; see also Marler & Mitani, 1988]). The final pathway from the motor nuclei to the phonatory muscles is homologous in humans and monkeys (Ploog, 1988), but differences in the shape of the supralaryngeal vocal tract prevent nonhuman species from articulating most of the sounds that are necessary for speech. A summary of the neurobiology and pathology of subhuman vocal communication and human speech may be found in Ploog (1988).

Development of acoustical structures. Although each of many hundred primate species has its own vocal repertoire (Ploog, 1986), physiological factors influencing phonation and articulation under affective arousal might still be similarly structured across different species. Tembrock (1971, 1975) and Morton (1977) studied this notion, and Tembrock (1975) found, based on acoustic analyses of calls across different species, repeated short sounds with low frequencies for calls in the contact range, seemingly concurrent with states of relaxation and contentment (e.g., play calls). Low frequencies were characteristic of threat calls and dominance calls in agonistic encounters, and short calls with a high-amplitude onset and a broad-frequency spectrum were characteristic of defense calls. High frequencies, repeated frequency shifts, and a tendency toward temporal prolongation were found for submission calls. High frequencies and temporal prolongation were characteristic of attraction calls over long distances. Morton (1977) conceptualized the relationship between acoustic structure and affective state with a "motivation-structural rule." According to this rule, low frequency and harsh sounds are used by birds and mammals in hostile contexts, whereas more pure and tonelike sounds are utilized when animals are frightened, appeasing, or approaching in a friendly manner. Backed by a body of empirical data on bird and mammal sounds, Morton proposed a selective value in using these calls, namely, that low-frequency utterances might feign a large body size in combat situations, whereas high-frequency vocalizations might appear more like those of an infant, owing to its smaller resonance body.

Based on this argument and spectrographic studies of acoustical changes as a function of retracting the corners of the mouth, Ohala (1980) hypothesized an acoustical origin of the smile in humans. Ohala's evidence suggests that smiling produces upward shifts in the acoustical structure of vocalizations that are similar to utterances produced with a shorter vocal tract, such as is typical for infants. Ohala's proposal is an example of the creative use of ethological empirical evidence in the framework of human affective displays.

One problem with most studies of vocalization patterns in subhuman animals is the use of various nonstandardized descriptors of call types. The lack of consensual classification systems of acoustic characteristics on one hand and the difficulty to assess the sender's motivational state on the other hand encumber quantitatively sound comparisons across species and frequently within species. Clearly, there is a need for more experimental work both in the field and in the laboratory (see Scherer & Kappas, 1988).

Ecological demands. Regarding environmental influences on the structure of nonhuman vocalizations, there seem to be a variety of instances in which the acoustical pattern of a call is particularly suited to serve a specific function in a specific habitat (e.g., Brown & Waser, 1988). One example is the structure of the loud vocalizations of howling monkeys that seem to regulate intergroup spacing. The barklike structure of these calls is not only particularly well suited to propagate over long distances but also provides very precise information about the distance to the caller, based on reverberation phenomena (Whitehead, 1987) that would not be easy to detect with slow-onset, temporally prolongated vocalizations. It would be beyond the scope of this chapter to discuss this matter in detail, but we shall note that there might be an evolutionary continuity with regard to the use of specific communication channels for particular communicative functions that is easier to understand by studying habitat-related communicative behaviors. For a functional analysis of the use of particular channels, or signs within channels in humans, the study of environmental fits between habitat and sign is of particular interest. It is amusing to note, in concluding this section, that Charles Dickens was convinced that the Americans had such loud voices because of the large distances their voices had to cover on the huge farms.

Acoustical characteristics of affective vocalizations

Parameters of acoustical analysis. A comprehensive analysis of the findings regarding the acoustical signatures of affective vocalizations is difficult because many of the researchers used (trained or untrained) judges and differing sets of scales to assess the vocal parameters in their studies (e.g., Constanzo, Markel, & Constanzo, 1969; Davitz, 1964a,b; Eldred & Price, 1958; Green & Cliff, 1975; Markel, Bein, & Phillis, 1973; Scherer, Koivumaki, & Rosenthal, 1972). On the other hand, studies using objective analyses of fundamental frequency, intonation, loudness, and rate measures frequently have utilized only a subset of measures, of emotions to be compared, or of both. Ideally, the objective acoustical measurement of affective vocalizations is the most promising level of description. Parameters are well defined and can be easily extracted. However, there is some cost involved, in money and time, as well as in acquiring the necessary acoustic–phonetic expertise.

Next we shall describe the acoustic parameters that have been employed and then give an overview of the results accumulated so far. Extensive reviews of studies relating to the acoustic characteristics of discrete emotional states can be found in Scherer (1981a, 1986a) and Bergmann, Goldbeck, and Scherer (1988). For a more detailed description of the parameters typically employed, see Scherer (1982, 1989).

TIME PARAMETERS. The temporal aspects of speech that are usually assessed are speech rate (number of syllables per time unit, including pauses), articulation rate (number of syllables per time unit, excluding pauses), and the frequency and duration of pauses. Another set of measures commonly employed pertains to disturbances of continuity (e.g., Speech Disturbance Ratio, ah-ratio, non-ah-ratio, repeats; see, e.g., Mahl & Schulze, 1964). It would be beyond the scope of this chapter to discuss these in detail, but see Siegman and Feldstein (1987) for a detailed review.

AMPLITUDE PARAMETERS. Amplitude parameters describe in the context of human speech the intensity of the sound pressure of waves as they propagate through air, and the intensity of these pressure waves helps determine our perception of loudness. Although intensity is technically easy to assess, it is very difficult to interpret, as it is dependent not only on vocal effort but also on the distance and orientation of the microphone and speaker. Recordings must be controlled with the utmost care concerning differences in the microphone placement and amplifier (tape recorder) settings between conditions.

FREQUENCY PARAMETERS. The parameter that has probably been used most often is the fundamental frequency of the vocalization (F_0). The *fundamental frequency* is the frequency of glottal vibration and is closely related to the perceived pitch of an utterance. Perceived pitch, however, is also dependent on other factors such as voice quality. Standke, Kappas, and Scherer (1984) found a correlation of .89 between estimated (relative) pitch of voice and mean fundamental frequency. Indeed, changes in fundamental frequency can be so fast that they are not perceived as changes in pitch, but the voice appears to have a harsh character, a phenomenon called *jitter*. Slower variations in F_0 are usually perceived as pitch changes, and variations are referred to as *vibrato*. Over longer periods of time – for example, within a sentence – there will be extensive changes in F_0 that are typically referred to as *intonation contours*. Apart from mean F_0, range and variance have been used as parameters of fundamental frequency.

SPECTRAL PARAMETERS (PARAMETERS OF FREQUENCY AND AMPLITUDE). As we outlined in our description of the articulation process, certain frequencies are amplified, and others are attenuated according to the shape of the vocal tract. For a small number of frequencies, given a specific vocal tract shape, there is resonation. The locations of these resonances are called *formants*. They are described by their frequency, their amplitude at the peak, and their bandwidth, indicating whether the resonance is very much localized on a particular frequency or whether neighboring frequencies are amplified as well. The location of the formants determines the perception of the vocalization's sound quality. Vowels are characterized by a specific pattern of formant locations. Although there are interindividual variations in their formant locations, listeners generally have no difficulty identifying a particular vowel. Formants are commonly extracted from a frequency-by-amplitude display of an acoustic waveform, the *spectrum*. Over shorter periods of time the spectrum is dependent on the particular phonemes uttered. Over a longer period of time (several sentences and/or minutes) the variations due to language average out, and the resulting *long-term spectrum* indicates the speaker's habitual settings. If long-term spectra are compared within a person in differing affective states, the differences in the spectra will reflect the impact of those states on either the habitual articulatory settings or the precision with which those settings are achieved. These comparisons should always be made as within-subject designs, because most of the variation is caused by the particular shape of the individual's vocal tract, and large interindividual differences exist.

Empirical data. As we mentioned earlier (see Scherer, 1979a, 1986a) most of the studies in this area lack theoretical and methodological rigor, and some of the most serious shortcomings are the following: using actor-portrayed emotional utterances, as opposed to naturally occurring emotional vocalizations; not systematically controlling important variables such as the number of speakers, the type of emotions studied, the instructions for portrayal, and the verbal material used; and assessing a very limited number of parameters, often without using the latest technology. Yet the findings often converge. In this section we shall review discrete emotions, emotional disorders (particularly depression), and stress (research summaries are based on a more extensive review in Scherer, 1989).

VOCAL CUES OF DISCRETE EMOTIONS. As shown elsewhere (Scherer, 1986a), it is necessary to distinguish the mild or passive forms of a particular emotion category from the strong and active forms of the same type (e.g., cold anger or irritation vs. hot anger or flaring rage). Because this distinction is rarely made in the literature, the following review in some cases had to be based on educated guesses regarding the particular type of the emotion category studied.

Boredom–indifference. The results generally indicate a decrease in mean F_0 (Davitz, 1964a; Fairbanks & Pronovost, 1939) and mean intensity (Bortz, 1966; Davitz, 1964a; Müller, 1960). Most likely, actors simulate reduced sympathetic arousal.

Displeasure–disgust. Three of the studies (Plaikner, 1970; Scherer, 1979a; Scherer, Wallbott, Tolkmitt, & Bergmann, 1985) found an increase in mean F_0, and two others (Kaiser, 1962; Van Bezooijen, 1984) found a decrease. The discrepancy could be due to differences in the respective induction procedures (viewing of an unpleasant film vs. an actor's portrayal of disgust). It might be necessary to differentiate the nature of the antecedent situation more clearly.

Irritation–cold anger. The relevant findings indicate an increase in mean F_0 (Eldred & Price, 1958; Roessler & Lester, 1976), mean intensity (Constanzo, Markel, & Constanzo, 1969; Eldred & Price, 1958), and high-frequency energy (Kaiser, 1962; Roessler & Lester, 1976) as well as a tendency toward downward-directed intonation contours (Höffe, 1960; Kaiser, 1962).

Rage–hot anger. Similar to the pattern for cold anger, increases in mean F_0 (Davitz, 1964a; Fairbanks & Pronovost, 1939; Fonagy, 1978; Höffe, 1960; Wallbott & Scherer, 1986a; Williams & Stevens, 1969, 1972) and mean intensity (Bortz, 1966; Davitz, 1964a; Höffe, 1960; Kotlyar & Morosov, 1976; Müller, 1960; Williams & Stevens, 1969; Van Bezooijen, 1984) have been reported. The specific cues, probably related to very high sympathetic arousal, are increases in F_0 variability and F_0 range (Fairbanks & Pronovost, 1939; Havrdova & Moravek, 1979; Höffe, 1960; Williams & Stevens, 1969).

Sadness–dejection. Possibly because of a high degree of similarity between eliciting situations (see Scherer, Wallbott, & Summerfield, 1986), the results for this affective state are very consistent: decrease in mean F_0 (Coleman & Williams, 1979; Davitz, 1964a; Eldred & Price, 1958; Fairbanks & Pronovost, 1939; Fonagy, 1978; Kaiser, 1962; Sedlacek & Sychra, 1963; Wallbott & Scherer, 1986a; Williams & Stevens, 1969), and F_0 range (Fairbanks & Pronovost, 1939; Fonagy, 1978; Sedlacek & Sychra, 1963; Van Bezooijen, 1984; Williams & Stevens, 1969; Zuberbier, 1957; Zwirner, 1930) as well as downward-directed F_0 contours (Fairbanks & Pronovost, 1939; Kaiser, 1962; Sedlacek & Sychra, 1963; Zwirner, 1930). In addition, mean intensity decreases (Davitz, 1964a; Eldred & Price, 1958; Hargreaves, Starkweather, & Blacker, 1965; Huttar, 1968; Kaiser, 1962; Müller, 1960; Skinner, 1935; Zuberbier, 1957) as does high-frequency energy (Hargreaves, Starkweather, & Blacker, 1965; Kaiser, 1962; Skinner, 1935) and precision of articulation (Van Bezooijen, 1984; Zuberbier, 1957).

Grief–desperation. We believe that one should distinguish between a passive reaction to loss, that is, sadness and dejection, and a more aroused state of desperation or violent grief, because theoretically different acoustic correlates for grief–desperation in relation to sadness–dejection are to be expected. Unfortunately, we have been unable to find any studies of these states.

Worry–anxiety. Although there is a large body of literature on verbal and temporal correlates of anxiety (Siegman & Feldstein, 1987) the vocal parameters we described have been studied only rarely. Several studies (Bonner, 1943; Hicks, 1979; Höffe, 1960; Plaikner, 1970) reported a tendency toward a mean F_0 increase.

Fear–terror. Studies consistently report an increase in mean F_0 (Coleman & Williams, 1979; Duncan, Laver, & Jack, 1983; Fairbanks & Pronovost, 1939; Fonagy, 1978; Höffe, 1960; Kuroda, Fujiwara, Okamura, & Utsuki, 1976; Niwa, 1971; Roessler & Lester, 1976; Sulc, 1977; Utsuki & Okamura, 1976; Williams & Stevens, 1969), F_0 range (Fairbanks & Pronovost, 1939; Utsuki & Okamura, 1976; Williams & Stevens, 1969), F_0 variability (Fairbanks & Pronovost, 1939; Williams & Stevens, 1969), and high-frequency energy (Roessler & Lester, 1976, 1979; Simonov & Frolov, 1973).

Enjoyment–happiness. Because the positive emotions have been generally studied in the form of active joy or elation, we have not been able to find empirical evidence for acoustic cues accompanying this state of peaceful enjoyment.

Joy/elation. There are consistent reports of increases in mean F_0 (Coleman & Williams, 1979; Davitz, 1964a; Fonagy, 1978; Havrdova & Moravek, 1979; Höffe, 1960; Huttar, 1968; Kaiser, 1962; Sedláček & Sychra, 1963; Skinner, 1935; Van Bezooijen, 1984), F_0 range (Fonagy, 1978; Havrdova & Moravek, 1979; Höffe, 1960; Huttar, 1968; Sedlacek & Sychra, 1963; Skinner, 1935), F_0 variability (Havrdova & Moravek, 1979; Sedlacek & Sychra, 1963; Skinner, 1935), and mean intensity (Davitz, 1964a; Höffe, 1960; Huttar, 1968; Kaiser, 1962; Kotlyar & Morosov, 1976; Müller, 1960; Skinner, 1935; Van Bezooijen, 1984).

As shown earlier (Scherer, 1981a, 1986a) the frequent reports of increases in F_0 and intensity, both mean and variability, may indicate a general sympathetic response syndrome. We believe, however, that it would be wrong to assume that only sympathetic arousal is reflected in the voice and that it is unlikely that differential emotional states can be distinguished via vocal parameters. The major argument for this point is that many different affective states can in fact be recognized by naive judges on the basis of the vocal cues alone. Thus, one must assume the existence of differences in the vocal parameters on the basis of which this differential recognition is possible. We are quite certain that greater vocal differentiation will be found as soon as a more complete range of the aforementioned acoustic parameters are used in the relevant studies. The parameters most closely linked to sympathetic arousal happen to be the ones that are most conveniently measured. In addition, the arousal dimension is probably the most dominant one in terms of overall effect.

VOCAL CUES OF AFFECT DISTURBANCE. We shall restrict our review to studies of depressive patients. As always in studies with clinical populations, a major problem is the difficulty of finding patient groups with homogeneous diagnostic criteria. Indeed, the lack of replication of findings in this area can often be traced back to this very problem. For example, one could scarcely expect similar vocal cues for endogenously depressed patients and biphasic, manic-depressive patients, particularly depending on the phase during which the latter might be studied. This difficulty is even more pronounced for the psychoses and, in particular, schizophrenia.

The literature on vocal cues of depression is quite encouraging, however, showing fairly consistent differences between normal and depressed speech and indices for change after therapy. Depressive patients speak with relatively low intensity (Eldred & Price, 1958; Moses, 1954; Whitman & Flicker, 1966; Zuberbier, 1957) and with reduced dynamic range (Zuberbier, 1957). Conversely, the intensity tends to increase after therapy (Hargreaves & Starkweather, 1964, 1965; Ostwald, 1961, 1963; Tolkmitt, Helfrich, Standke, & Scherer, 1982).

Some studies have reported low mean F_0 for depressives (Bannister, 1972; Eldred & Price, 1958; Moses, 1954; Roessler & Lester, 1976). Yet, F_0 seems to decrease after therapy (Klos, Ellgring, & Scherer, 1987; Tolkmitt et al., 1982), a finding that has been explained by assuming a decrease in general tension in the striated musculature (Scherer, 1979a). It might be that the F_0 level is linked to the severity of the depression (Whitman & Flicker, 1966).

This apparent discrepancy might also be due to a lack of homogeneity in the nosologic criteria or to the difference between objective measures of F_0 and subjective ratings of pitch (pitch judgment being influenced by other acoustic variables, such as energy distribution in the spectrum and variability of pitch). As several authors have discovered a narrow range and restricted variability for F_0/pitch in depressive speech (Bannister, 1972; Hargreaves et al., 1965; Newman & Mather, 1938; Ostwald, 1964; Zuberbier, 1957) it may be that the judges are affected by reduced range and other acoustic factors when rating pitch.

Furthermore, several studies (Moses, 1954; Newman & Mather, 1938; Zwirner, 1930) have shown that depressives tend to employ repeated, downward-directed intonation contours (which are not necessarily flat), which give an impression of monotonousness. There is also evidence that depressives use rather lax, imprecise articulation (as reflected in formant precision) and that the precision of articulation improves after therapy (Tolkmitt et al., 1982; Zuberbier, 1957).

Much of this is still very preliminary evidence and urgently requires replication. And moreover, research in this area should become less atheoretical. A number of concrete hypotheses concerning the vocal changes to be expected for different types of emotional disorders were suggested by Scherer (1987).

VOCAL CUES OF STRESS. "Voice lie detectors" have been highly publicized, and quite a number of expensive units have been sold based on dubious claims concerning the function of certain features (e.g., *microtremor*) as markers of deception. Apart from the obvious ethical problems posed by the surreptitious "assessment" of lying for criminal defendants or employees under suspicion, there is no hard evidence that lying can be detected through simple electroacoustic devices (Hollien, 1981; Hollien, Geison, & Hicks, 1987; Scherer, 1981b). One also needs to distinguish between the detection of lying and the detection of stress. Lying does not necessarily produce stress, and there is no reason that such nonstressful lies should produce specific voice changes, given that presumably no physiological reaction takes place.

The strong sympathetic arousal accompanying stress should clearly influence vocal parameters, and in particular one would expect increases in fundamental frequency and F_0 variability. This theoretical expectation has been confirmed in a large number of studies in which stress was experimentally induced (cf. Ekman, Friesen, & Scherer, 1976; Scherer, 1981b; Streeter, Krauss, Olson, Geller, & Apple, 1977; Williams & Stevens, 1969), although individual differences are often very pronounced (as shown in a high degree of response specificity).

To account for the individual differences, we ran a number of experimental studies using personality variables in the form of coping styles. We found that for both male and female subjects, only those with relatively elevated anxiety scores (regardless of whether they repressed or admitted their anxiety) showed an F_0 increase. Furthermore, our data indicated that a consistent effect for female anxiety-deniers (repressers) is more precise articulation with increasing cognitive stress and decreasing precision with mounting emotional stress (Scherer, Wallbott, Tolkmitt, & Bergmann, 1985). Thus, personality factors, coping style, and type of stress should be systematically varied or at least controlled in further studies. To provide a more extensive theoretical basis, Scherer (1986b) proposed that stress be analyzed within the framework of a general theory of emotion, assuming that stress occurs in cases in which a problem cannot be solved through normal emotional responding (with return to baseline within a standard time frame).

The recognition of affective state from vocal cues

There is ample evidence that the voice carries a wealth of information about the speaker. From nonverbal cues, listeners can determine the speaker's gender (Smith, 1979, 1980), age (Helfrich, 1979), socioeconomic background (Robinson, 1979), ethnic group (Giles, 1979), personality (Scherer & Scherer, 1981), and status in small groups (Scherer, 1979b). From vocal cues, listeners can also determine affective states when presented with only a few sentences or, even more clearly, shouts or cries. Many studies have assessed the ability to identify affective states from voices, typically using recordings. The samples ranged from single words or nonsense syllables (spoken in isolation with different inflections) to paragraphs of texts or whole conversations. Scherer (1981a) showed in a review of studies of the perception of affective states that the recognition of emotions from voice and speech samples is as good or better than the recognition of affect from facial expressions (on average 56% accuracy, compared with 12% expected by chance in 28 studies).

The interpretation of the results of many judgment studies is problematic, owing to the stimulus material's lack of ecological validity. Studies using numbers, letters, nonsense syllables, or isolated words as stimuli may introduce artifacts by inviting speakers to produce stereotypical samples that may lead to overestimating the accuracy of the judgment (Bortz, 1966; Davitz, 1964a; Green & Cliff, 1975; Levin & Lord, 1975; Skinner, 1935). The use of poems or similar material might prompt speakers to use voice settings habitually used for reciting such material (Sedlacek & Sychra, 1963; Zuberbier, 1957). The technique most commonly employed is to inject standard sentences into texts of varying emotional content (e.g., Constanzo, Markel, & Constanzo, 1969; Davitz, 1964a; Fairbanks & Pronovost, 1939; Fairbanks & Hoaglin, 1941). Unfortunately, there is frequently an emotional bias in interpreting these sentences (see Kramer, 1963). Given these problems, it is advisable to pretest the stimulus material extensively and to use more than one stimulus type. In one study Sincoff and Rosenthal (1985) demonstrated using the same encoders, differences in recognition of six emotions, depending on whether the neutral sentence technique or letters of the alphabet were used as the stimulus material.

Another problem pertains to the encoder of the stimulus material. On one hand, students, typically employed as subjects in such studies, may frequently not be able to stage believable vocal affective displays, but the use of professional actors (Scherer, Koivumaki, & Rosenthal, 1972; Wallbott & Scherer, 1986a; Williams & Stevens, 1972), on the other hand, does not guarantee vocal samples that are similar to real emotional vocal expressions. Therefore, more than one speaker should encode the stimulus material, as there is considerable evidence for interindividual differences in the ability to portray emotions vocally (see Wallbott & Scherer, 1986a).

A major difficulty is defining the target emotion to be displayed. Frequently, it is not defined satisfactorily if an encoder (speaker) is to portray, for example, suppressed anger or exploding anger. A better approach is to use an explicit scenario so that the (imaginary) antecedent conditions are similar or the same for different actors (Wallbott & Scherer, 1986a).

A problem that has not been addressed relates to the difference of hearing an encoder for the first time (and for only a few utterances) or judging the effective state of a well-known person. It seems that the variability in encoding might present a baseline problem for the judgment of affect. Cues that might be interpreted as indicators of arousal (e.g., tenseness or a high-pitched voice) might be the habitual vocal setting of a specific speaker. The resulting biases in the attribution of emotions are comparable to biases introduced by the facial furrows of elderly people (Malatesta, Fiore, & Messina, 1987). Familiarity with the encoder is a context variable in the framework of attribution of affective states and should be taken into account as such, as it is otherwise difficult to quantify the impact of the phenomena studied in context-free paradigms on everyday interaction (see also Buck, 1983).

In order to address questions regarding the universality of vocal expressive patterns of affective states, Scherer and colleagues are currently undertaking a multinational research project on emotion recognition from vocal cues using the Vocal Emotion Recognition Test (Scherer & Wallbott, in preparation). The test consists of a set of 30 emotions (joy–happiness, sadness, fear, anger, disgust) portrayed by professional actors. These episodes have been selected from a pool of 80 stimuli used in a pilot study. The episodes were produced using a scenario approach, in which the scenarios were selected from a pool of situation descriptions collected in a large-scale cross-cultural study on antecedents of emotion (Scherer, Wallbott, & Summerfield, 1986). The pseudo-sentences were constructed by a phonetician to represent at least one typical syllable from each of six languages (German, English, French, Italian, Spanish, and Danish). The first data from several countries point toward better-than-chance recognition independent of language or culture.

Summarizing the evidence on recognition of emotions from vocal samples, we conclude that there is ample evidence of the subjects' ability to correctly identify affective states from vocal stimuli. This is especially noteworthy, as an identification of the relevant acoustical parameters has so far been successful only for the dimension of arousal. However, if listeners are able to decode the sender's emotional state from vocal cues, the voice must be carrying that information in its acoustical parameters. Systematic analyses of the acoustical parameters on which listeners base their judgments may guide their exploration. The systematic and rule-based variation of voice parameters using synthetic or resynthetized vocalizations can achieve this end.

Recognition of synthetically produced affective vocalizations

Scherer and colleagues used a Moog synthesizer to manipulate systematically a variety of acoustic cues subjects used to determine a speaker's affective state to investigate the cues (Scherer, 1974; Scherer & Oshinsky, 1977). The researchers synthesized tone sequences modeled after the intonation contour of a short sentence, with differences in pitch, duration, and contour. Tempo and pitch variations were found to have the strongest influence on subjects' ratings. Moderate pitch variation led to an impression of sadness, fear, disgust, and boredom, and extreme pitch variations and a rising contour produced attributions of happiness, interest, and fear.

Although the use of fully synthetic stimuli allows complete control over every acoustic parameter, there are drawbacks to this method: The stimuli may appear artificial and thus elicit judgments with low ecological validity. The subjects' attention may be directed to unrealistic acoustic variations, because the linguistically determined vocal structure of an utterance is not present, giving undue importance to the acoustic variation performed on the stimuli.

If intonation carries emotional information, then the relevant information generally must be laid on top of the linguistically required changes in intonation. For instance, a rising intonation contour is a linguistic feature that signals a question. Knapp (1978) offered five examples of how vocal emphasis modifies the meaning of a message:

1. *He's* giving this money to Herbie. (HE is the one giving the money; nobody else.)
2. He's *giving* this money to Herbie. (He is GIVING, not lending, the money.)
3. He's giving *this* money to Herbie. (The money being exchanged is not from another source; it is THIS particular money.)
4. He's giving this *money* to Herbie. (CASH is being exchanged, not a check.)
5. He's giving this money to *Herbie*. (The recipient is HERBIE, not Eric or Rod.)

Add a final rise of intonation, and you get

6. He's giving this money to *Herbie?* (Why is he giving the money to HERBIE and not Mark?)

Actually, examples 1 through 4 can also be converted into questions by changing the final accent to that of a question.

Scherer and colleagues constructed two models to explain how voice parameters associated with affective states might interact with linguistic parameters (Ladd, Silverman, Tolkmitt, Bergmann, & Scherer, 1985; Scherer, Ladd, & Silverman, 1984). The *covariance model* states independence between the two communicative systems. Information regarding the speaker's internal state is encoded independently of the linguistic meaning. Relevant acoustic parameters vary with the strength of a particular state. The *configuration model*, on the other hand, states that verbal and nonverbal cues exhibit categorical linguistic structures, and different speaker states are conveyed by means of different configurations of categorical variables. The assumption is that speaker states involving more complex cognitive processes are better understood in terms of the configuration model, whereas simpler affective states are better understood in terms of the covariation model.

To evaluate the two models, Scherer and associates performed a series of experiments (Scherer, Ladd, & Silverman, 1984). First, recordings or transcripts of utterances were evaluated on several scales using a between-subjects design. The results showed high interrater agreement for the audio and low interrater agreement for the transcript group. Scherer and colleagues concluded that the nonverbal aspects of the statements, and not the content itself, carried the emotional meaning. For the second experiment the 24 utterances yielding the clearest emotional meaning were submitted to several degradation or masking procedures to identify the acoustical features leading subjects to their judgments. (1) The high-frequency components of the utterances were attenuated, completely masking verbal content, but the intonation contour was left intact (low-pass filtering); (2) the utterances were randomly cut and mixed, removing the temporal organization of the utterances and destroying both content and intonation completely, but the spectral cues to voice quality were not changed (random splicing, Scherer, 1971); and (3) the utterances were played backward to mask the content, but the voice quality was left intact. Together with the original full audio stimuli there were 84 stimuli, which were presented to groups of subjects who evaluated them on a set of scales. The results showed that the evaluations were largely determined by voice quality cues, independent of the presence or the absence of content (text) or gross distortions of F_0 contours. These results are seen as evidence that the affective force of the utterance is encoded independently of the verbal content, as posed by the covariance model.

To test the configuration model a different approach had to be used, as masking techniques do not address the assumption that intonational cues signal affect in conjunction with text. To investigate this assumption, the utterances were classified according to contour type (rise vs. fall), and question type (why vs. yes–no), as well as fundamental frequency range and standard deviation. The subjects' ratings on the scales were analyzed using multiple regression analyses. The results demonstrate interactions between contour type and text in communicating aspects of speaker affect.

To control experimentally the intonational parameters relevant to the configurational model, digital speech resynthesis was employed (Goldbeck, Tolkmitt, & Scherer, 1988; Ladd et al., 1985). Resynthesis allows the modification of systematically varied acoustic features of natural speech, leaving all other cues unchanged. Although there is some loss of tone quality, the subjects usually do not have the impression of artificiality generally associated with synthetic speech. The findings suggest that features of cognitive attitude can be transmitted by linguistically coded aspects of sentence structure, whereas gradations of emotions are conveyed by speech activity as a whole (e.g., fundamental frequency range). The same pattern of results was found across speakers and was independent of verbal content.

Kappas (1986) could show that the intonation contour as a whole may

capture the affective valence of ambiguous statements. In this study a set of utterances, judged as neutral in transcript, were spoken by an actor expressing either anger or happiness. The resulting samples were judged on a set of emotion and attitude scales. Samples that clearly elicited positive or negative ratings were submitted to digital analysis. The intonation contours as a whole were then transferred onto neutral samples of the same sentences using digital resynthesis. Rate, loudness, and spectral composition were exactly the same; only intonation was varied. A clear valence difference for the resulting sets of sentences was revealed.

These studies emphasize the necessity of integrating the investigation of affective vocal cues with linguistic approaches to understand the role of specific acoustic features in the attribution of emotional states. The linguistic context thus should be taken into account when interpreting the attribution process regarding the use of vocal cues to affective states.

The recognition of affective state from vocal cues in conjunction with other nonverbal channels

Evidence concerning the recognition of affective state from vocal cues indicates that vocal parameters are a prime candidate for diagnosing affective state in interaction. This knowledge is particularly valuable in situations in which no other communication channels are available, such as communication via radio or telephone. But if information from other communication channels is available, the picture will not be as clear. Generally, if both facial and vocal cues are available, subjects tend to base their attributions regarding the speakers' affective state predominantly on facial cues (e.g., Berman, Shulman, & Marwitt, 1976; Ekman & Friesen, 1969; Graham, Ricci Bitti, & Argyle, 1975; Hess, Kappas, & Scherer, 1988; Levitt, 1964; Mehrabian & Ferris, 1967; Wallbott & Scherer, 1986a; Zaidel & Mehrabian, 1969; for an overview, also see Noller, 1985).

Research on the integration of vocal and other communication channels has been hampered by technical as well as methodological problems that make it difficult to determine the processes of attributing affective state (and attitudes). For instance, the vocal and facial stimuli used could not be produced, and thus controlled independently, leading to confounds. The length of the stimuli and the method of presentation were not adequate for a judgment task because of technological limitations. To address some of these problems, Hess, Kappas, and Scherer (1988) used an actor to display specific facial actions that were previously defined using Ekman and Friesen's (1978) Facial Action Coding System (FACS). Vocal parameters were produced independently and synthesized using digital signal–processing algorithms. The synchronized dubbing of facial and vocal stimuli with information regarding the social context provided a systematic and independent variation of the parameters involved, thereby avoiding the confounds of some earlier studies. The increasing availability of synthetic signal production and modification facilities for audio and video material provides the basis for future studies that may systematically investigate the attribution processes that integrate facial and vocal parameters in a more realistic context, while retaining full experimental control and thus providing a better understanding of humans as "regulated multisensory stations in a transmission system" (Birdwhistell, 1968, quoted in Wallbott, 1979).

Hess, Kappas, and Scherer (1988) suggested an integration model that applies varying weights to the importance of channels according to social context (deception, "normal" social interaction). It seems that it is not sufficient to estimate the relative "merits" of vocal parameters in a contextual void. Rather, one must take into account contextual parameters determining the use of specific vocal (or other nonverbal) cues for the attribution process of emotional state. It is likely that we base our attributions in different circumstances on different cues. There is evidence that the focus on facial cues is reduced if deception is involved or expected (e.g., Noller, 1985). Buck (1983) introduced the concept of decoding rules (analogous to the display rules suggested by Ekman [1972], a phenomenon already described by Wundt, 1903). He defined decoding rules as heuristics guiding attention to specific nonverbal cue complexes, depending on the situational context. This notion is appealing, particularly given the functional advantages of a cue selection heuristic with a view to the sheer volume of information provided in a multichannel interaction situation. The results of the synthetic multichannel study support the context dependency of relative channel importance as a function of channel discrepancy (Hess, Kappas, & Scherer, 1988), as well as a function of intimacy of context (Hess & Kappas, 1985).

Integrating approaches to vocal affect communication in the context of social psychology

Many vocal phenomena fulfill both linguistic and nonlinguistic functions, and frequently the concurrence of both phenomena discourages scientists interested in communication processes. It seems, however,

that the multifunctionality of vocal signals indicates their pragmatic significance. In fact, as Karl Bühler (1933) pointed out in his *organon* model of language, most signals serve several functions simultaneously. Scherer (1980) suggested using the semiotic approach (Morris, 1946; Peirce, 1931–1935) to differentiate between semantic functions (i.e., nonverbal signs replacing, amplifying, contradicting, or modifying verbal signs), pragmatic functions (expression of sender state, reactions, and intentions), dialogic functions (relationship between sender and receiver, regulation of interaction), and, in addition, syntactic functions, related to the ordering of signs in their sequential and hierarchical organization. One cannot understand the communicative function of messages, be it vocal, facial, or otherwise without acknowledging these levels of functions. Similarly, the knowledge of the context in which the communication occurs determines the translation of the received message. For instance, Wallbott (1988) showed the importance of context information for the attribution of emotions from facial stimuli. One may therefore assume that context information plays a similar role in the interpretation of vocal signals.

Scherer (1977) and Goffman (1979) suggest that nonspeech vocal interjections such as "oh," "oops," or "ouch" may be used strategically to achieve an impression of genuineness. The actual interpretation of such interjections depends on whether a listener is able to recognize a specific affect (e.g., surprise), but this is only a necessary, not a sufficient, precondition. If the listener knows the speaker well (i.e., if the listener has a baseline), he or she might interpret an occasional "oh" quite differently, depending on the speaker's known habits. It is here that our knowledge regarding the marker function of voices gains importance. There is empirical evidence that information about a speaker's gender, age, socioeconomic status, and may be personality is transmitted through paralinguistic features of speech. Even if the interaction partners do not know each other and the interaction is based solely on audible cues (e.g., during a phone call), these markers provide a context for the attribution of emotional states. It is therefore clear that it does not help to try to discover a set of vocal parameters predicting an attribution of affect, personality, or attitudes if one does not take into account the context information regarding the interaction, the interaction history, the interplay of different communication channels, and the linguistic context.

Scherer repeatedly argued (e.g., 1986a) that an explicit theory is required to integrate the results regarding the impact of affective state on vocalizations. He contended that vocalizations are determined by push

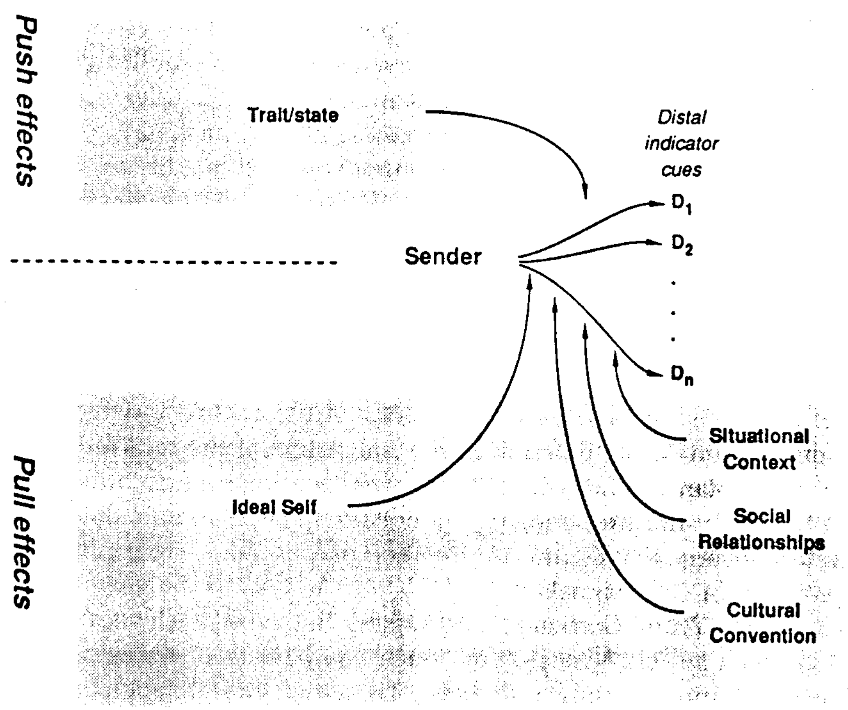

Figure 6.5. Push and pull effects.

and pull factors (Scherer, 1985; Scherer, Helfrich, & Scherer, 1980). Push factors are physically and physiologically linked to internal states such as muscular tension, and pull factors are linked to psychological processes such as display rules (Figure 6.5).

This dichotomy relates to an involuntary–voluntary distinction important to understanding the actual physiological implementation of the production of speech. Knowledge of these factors enables the prediction of specific acoustical changes, given models of speech production. In order to predict push and pull factors, a model that links antecedents to emotional states and the accompanying changes in bodily systems is required. More specifically, we need to explore the physiological mechanisms that underlie the observed correlation between emotional change and change in vocal production, because with the current technology available, we are scarcely able to assess directly the psychophysiological processes accompanying emotional arousal and their effect on voice production. Unfortunately, even on a theoretical level there have been

few attempts to predict specific physiological changes in the voice-producing organs as a result of specific emotions. In part, the difficulty stems from the nature of most emotion theories, which have either been too global, for example, in assuming "packaged" neural programs for discrete emotion, or too unspecific, for example, in placing too much emphasis on nonspecific arousal, or too social, for example, by neglecting the psychophysiological underpinnings of the emotional states.

Scherer (1984) proposed a "component process" theory according to which emotional states are produced by the outcomes of a series of five stimulus evaluation checks and that predicts the adaptive effect of the results of these checks on the major subsystems of the organism, including the autonomic nervous system (ANS) and the somatic nervous system (SNS). The componential nature of this approach enables detailed predictions for the emotional impact on voice production. The specific predictions for the acoustic changes as a function of the various emotions just mentioned are shown in Table 6.1. (The procedure used to derive these predictions is outlined in detail in Scherer, 1986a.) We shall give only a brief overview of the theoretical argument.

The subsystem changes (i.e., psychophysiological responses, changes in motor expression, feeling states, or motivational changes) expected as a result of each type of stimulus evaluation check outcome are based on functional considerations with a strong phylogenetic bias. Starting from these assumptions – concerning the need for specific adaptive responses – detailed patterns of predications for voice production are derived from the existing literature on the relationships between ANS and SNS and their effect on the vocal organs (e.g., in relation to vocal cord functioning, tension of the intra- and extralaryngeal musculature, vocal tract wall tension, degree of salivation, articulatory settings). The hypotheses shown in Table 6.1 are derived using standard phonetic assumptions about the effect of such production variables on acoustic parameters.

Although this approach is somewhat speculative, it offers the possibility to link theories of emotion, psychophysiological research, and the acoustic measurement of vocal behavior, particularly insofar as the push effects are concerned. What is still missing and what will constitute the next phase of the theoretical work in our group is a set of predictions concerning the acoustic changes likely to result from pull effects with respect to the emotions (such as sociocultural norms regarding regulation and the control of affect vocalizations, strategic use of vocal emotion display, and accommodation of vocal affect responses). Should this

Table 6.1. *Changes predicted for selected acoustic parameters as a function of emotional state*

	Parameters											
Voice type	ENJ/ HAP	ELA/ JOY	DISP/ DISG	CON/ SCO	SAD/ DEJ	GRI/ DES	ANX/ WOR	FEAR/ TER	IRR/ COA	RAGE/ HOA	BOR/ IND	SHA/ GUI
F_0												
Perturbation	<=	>			>	>		≥		>		
Mean	<	≥	>	<>	<>	>	>	≥≥	<>	<>	≤	>
Range	<=	≥			≤	>		≥≥	<	≥≥		
Variability	<	≥			≤	>		≥≥	<	≥≥		
Contour	<	>			≤	>	>	>>	<	=		>
Shift regularity	=	<						<		<	>	
F_1 mean	<	<	>	>	>	>	>	>	>	>	>	>
F_2 mean			<	<	<	<	<	<	<	<	<	<
F_1 bandwidth	>	<>	<<	<	<>	<<	<	<<	<<	<<	<	<
Formant precision		>	>	>	≤	>	>	>	>	>		>
Intensity												
Mean	<	≥	>	>>	≤≤	>		>	≥	≥≥	<>	
Range	<=	>			<			>	>	>		
Variability	<	<			<			>		>		
Frequency range	>	>	>	>>	>	>>		>>	>	>	>	
High-frequency energy	<	<>	>	>	<>	>>	>	≥≥	>>	>>	<>	>
Spectral noise					>							
Speech rate	<	≥			≤	>		≥≥		≥		
Transition time	>	<			>	<		<		<		

Abbreviations: ANX/WOR, anxiety/worry; BOR/IND, boredom/indifference; CON/SCO, contempt/scorn; DISP/DISG, displeasure/disgust; ELA/JOY, elation/joy; ENJ/HAP, enjoyment/happiness; FEAR/TER, fear/terror; GRI/DES, grief/desperation; IRR/COA, irritation/cold anger; RAGE/HOA, rage/hot anger; SAD/DEJ, sadness/dejection; SHA/GUI, shame/guilt. F_0, fundamental frequency; F_1, first formant; F_2, second formant.
Symbols: >, increase; <, decrease; =, same. Double symbols indicate increased predicted strength of the change. Two symbols pointing in opposite directions refer to cases in which antecedent voice types exert opposing influences.
Source: Scherer, 1986a.

prove feasible, it would allow us to replace the atheoretical correlational approach that currently prevails in this research domain, as well as to provide links between biological, more specifically psychophysiological, and social–psychological research traditions in the study of communication.

Outlook

The objective analysis of emotional vocalizations might be facilitated by current developments in powerful microcomputers capable of performing complex acoustic analyses at a cost many laboratories can afford. As important as the progress in computer hardware is the development in user-interface technology, making it possible to conduct sophisticated analyses without going through painful programming efforts or long training times. The development of high-density storage devices makes it feasible to store entire discussions, allowing instant access to analyses of any point in the interaction, with pinpoint accuracy and high resolution.

Regarding research centered on the decoding of vocal affective cues, we can expect a simplification of synthesis and resynthesis procedures, allowing the modification and synthesis of speech sequences by rule and permitting the explicit testing of hypotheses without sacrificing the appeal of genuine voice samples. Similarly, the development in the analysis and synthesis of visual material, interactive digital video, and other techniques will allow us to study the interplay of vocal features with facial expression, posture, and gaze.

These technological advances alone, however, will not provide the means for a better understanding of nonverbal vocal communication. We might have to give up the search for a handful of vocal parameters that serve as a "window to our soul" and instead accept the idea that vocal cues to emotion, like the cues that other modalities offer, are deeply embedded in the psychological attribution context. We need powerful attribution models to define the conditions under which parameters have what effect on which decoder.

Our understanding of acoustical characteristics of affective vocalizations can be enhanced only if the necessary theoretical bases are covered. More work on the evolutionary development of affective vocalizations is needed, as is work on human emotion itself. If we do not understand physiological changes as a function of psychological reactions to emotional stimuli (or situations), we will not be able to predict vocal changes, which are, after all, driven by the emotional arousal ("push effects"). At the same time, it is necessary to model the relevant psychological context (e.g., "pull effects," such as vocal display rules). Keys to understanding the psychological context of voice production are studies in the area of development and pathology. Technology has been a decisive factor in the research on voice and emotion. It will remain important, but now, because technology enables the researchers to do what he or she wants, it is the theoretical development that has to follow suit.

References

Bannister, M. L. (1972). *An instrumental and judgmental analysis of voice samples from psychiatrically hospitalized and nonhospitalized adolescents.* Unpublished doctoral dissertion, University of Kansas, Lawrence.

Bergmann, G., Geldbeck, T., & Scherer, K. R. (1988). Emotionale Eindruckswirkung von prosodischen Sprechmerkmalen. *Zeitschrift für experimentelle und angewandte Psychologie, 35,* 167–200.

Berman, H. J., Shulman, A. D., & Marwit, S. J. (1976). Comparison of multidimensional decoding of affect from audio, video, and audiovideo recordings. *Sociometry, 39,* 83–89.

Bonner, M. R. (1943). Changes in the speech pattern under emotional tension. *American Journal of Psychology, 56,* 262–273.

Bortz, J. (1966). *Physikalisch–akustische Korrelate der vokalen Kommunikation.* Arbeiten aus dem physikalischen Institut der Universität Hamburg, 9.

Brown, C. H., & Waser, P. M. (1988). Environmental influences on the structure of primate vocalizations. In D. Todt, P. Goedeking, & D. Symmes (Eds.), *Primate vocal communication* (pp. 51–66). Berlin: Springer-Verlag.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments.* Berkeley and Los Angeles: University of California Press.

Buck, R. (1983). Nonverbal receiving ability. In J. M. Wiemann & R. P. Harrison (Eds.), *Nonverbal interaction* (pp. 209–242). Beverly Hills, CA: Sage.

Bühler, K. (1933). *Ausdruckstheorie.* Jena: Fischer.

Campos, J. J. (1988). Theoretical issues in the psychophysiology of emotion (structured panel discussion). *Psychophysiology, 25,* 422 (abstract).

Coleman, R. F., & Williams, R. (1979). Identification of emotional states using perceptual and acoustic analyses. In V. Lawrence & B. Weinberg (Eds.), *Transcript of the eighth symposium: Care of the professional voice* (Pt. I). New York: Voice Foundation.

Constanzo, F. S., Markel, N. N., & Constanzo, P. R. (1969). Voice quality profile and perceived emotion. *Journal of Counseling Psychology, 16,* 267–270.

Darwin, C. (1872). *The expression of emotions in man and animals.* London: John Murray.

Davidson, R. J. (1988). Theoretical issues in the psychophysiology of emotion (structured panel discussion). *Psychophysiology, 25,* 422 (abstract).

Davitz, J. R. (1964a). Auditory correlates of vocal expressions of emotional meanings. In *The communication of emotional meaning* (pp. 101–112). New York: McGraw-Hill.

(1964b). Personality, perceptual, and cognitive correlates of emotional sensitivity. In *The communication of emotional meaning* (pp. 57–68). New York: McGraw-Hill.

Dittus, W. (1988). An analysis of Toque Macaque cohesion calls from an ecological perspective. In D. Todt, P. Goedeking, & D. Symmes (Eds.), *Primate vocal communication* (pp. 31–50). Berlin: Springer-Verlag.

Duncan, G., Laver, J., & Jack, M. A. (1983). A psycho-acoustic interpretation of variations in divers' voice fundamental frequency in a pressured helium–oxygen environment. *Work in Progress, 16,* 9–16. Edinburgh: University of Edinburgh, Department of Linguistics.

Eisenberg, J. F. (1974). The function and motivational basis of hystericomorph vocalizations. *Symposium of the Zoological Society of London, 34,* 211–247.

(1976). Communication mechanisms and social integration in the black spider monkey *(Ateles fusciceps robustus)* and related species. *Smithsonian Contributions to Zoology, 213,* 1–108.

Ekman, P. (1972). Universals and cultural differences in facial expressions of emotion. In J. R. Cole (Ed.), *Nebraska Symposium on Motivation* (pp. 207–283). Lincoln: University of Nebraska Press.

(1973). *Darwin and facial expression: A century of research in review.* New York: Academic Press.

(1988). Theoretical issues in the psychophysiology of emotion (structured panel discussion). *Psychophysiology, 25,* 422 (abstract).

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica, 1,* 49–98.

(1978). *Facial Action Coding System (FACS): A technique for the measurement of facial action.* Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., Friesen, W. V., & Ellsworth, P. (1982). Conceptual ambiguities. In P. Ekman (Ed.), *Emotion in the human face* (2nd ed., pp. 7–21). Cambridge: Cambridge University Press.

Ekman, P., Friesen, W. V., & Scherer, K. R. (1976). Body movement and voice pitch in deceptive interaction. *Semiotica, 16,* 23–27.

Eldred, S. H., & Price, D. B. (1958). A linguistic evaluation of feeling states in psychotherapy. *Psychiatry, 21,* 11–121.

Fairbanks, G., & Hoaglin, L. W. (1941). An experimental study of the durational characteristics of the voice during the expression of emotion. *Speech Monographs, 8,* 85–90.

Fairbanks, G., & Pronovost, W. (1939). An experimental study of the pitch characteristics of the voice during the expression of emotion. *Speech Monographs, 6,* 87–104.

Fonagy, I. (1978). A new method of investigating the perception of prosodic features. *Language and Speech, 21,* 34–49.

Giles, H. (1979). Ethnicity markers in speech. In K. R. Scherer & H. Giles (Eds.), *Social markers in speech* (pp. 251–289). Cambridge: Cambridge University Press.

Giles, H., Scherer, K. R., & Taylor, D. M. (1979). Speech markers in social interaction. In K. R. Scherer & H. Giles (Eds.), *Social markers in speech* (pp. 343–381). Cambridge: Cambridge University Press.

Goffman, E. (1979). Response cries. In M. von Cranach, E. Foppa, W. Lepenies, & D. Ploog (Eds.), *Human ethology* (pp. 203–249). Cambridge: Cambridge University Press.

Goldbeck, T., Tolkmitt, F., & Scherer, K. R. (1988). Experimental studies on

vocal affect communication. In K. R. Scherer (Ed.), *Facets of emotion: Recent research* (pp. 119–137). Hillsdale, NJ: Earlbaum.
Graham, J. A., Ricci Bitti, P., & Argyle, M. (1975). A cross-cultural study of the communication of emotion by facial and gestural cues. *International Journal of Psychology, 10,* 57–67.
Green, S. (1975). Variation of vocal pattern with social situation in the Japanese monkey *(Macaca fuscata):* A field study. In L. A. Rosenblum (Ed.), *Primate behavior: Developments in field and laboratory research* (Vol. 4, pp. 2–102). New York: Academic Press.
Green, R. S., & Cliff, N. (1975). Multidimensional comparisons of structures of vocally and facially expressed emotion. *Perception and Psychophysics, 17,* 429–438.
Hargreaves, W. A., & Starkweather, J. A. (1964). Voice quality changes in depression. *Language and Speech, 7* 84–88.
(1965). *Vocal and verbal indicators of depression.* Unpublished manuscript, San Francisco.
Hargreaves, W. A., Starkweather, J. A., & Blacker, K. H. (1965). Voice quality in depression. *Journal of Abnormal Psychology, 70,* 218–220.
Harrison, R. P., & Wiemann, J. M. (1983). The nonverbal domain: Implications for theory, research, and practice. In J. M. Wiemann & R. P. Harrison (Eds.), *Nonverbal interaction* (pp. 271–285). Beverly Hills, CA: Sage.
Havrdova, Z., & Moravek, M. (1979). Changes of the voice expression during suggestively influenced states of experiencing. *Activitas Nervosa Superior, 21,* 33–35.
Helfrich, H. (1979). Age markers in speech. In K. R. Scherer & H. Giles (Eds.), *Social markers in speech* (pp. 63–107). Cambridge: Cambridge University Press.
Hess, U., & Kappas, A. (1985). *Decoding discrepant information in facial and vocal communication channels, with regard to contextual effects.* Paper presented at the Second European Conference on Facial Measurement, Saarbrücken, Germany.
Hess, U., Kappas, A., & Scherer, K. R. (1988). Multichannel communication of emotion: Synthetic signal production. In K. R. Scherer (Ed.), *Facets of emotion: Recent research* (pp. 161–182). Hillsdale, NJ: Erlbaum.
Hicks, J. W. (1979). An acoustical/temporal analysis of emotional stress in speech. *Dissertation Abstracts International, 41,* 417.
Höffe, W. L. (1960). Über Beziehungen von Sprachmelodie und Lautstärke. *Phonetica, 5,* 129–159.
Hollien, H. (1981). Analog instrumentation for acoustic speech analysis. In J. Darby (Ed.), *Speech evaluation in psychiatry* (pp. 79–103). New York: Grune & Stratton.
Hollien, H., Geison, L., & Hicks, J. W., Jr. (1987). Voice stress evaluators and lie detection. *Journal of Forensic Sciences, 32,* 405–418.
Huttar, G. L. (1968). Relations between prosodic variables and emotions in normal American English utterances. *Journal of Speech and Hearing Research, 11,* 481–487.
Jürgens, U. (1979). Vocalization as an emotional indicator. A neuroethological study in the squirrel monkey. *Behaviour, 69,* 88–117.
(1982). A neuroethological approach to the classification of vocalization in the squirrel monkey. In C. T. Snowdon, C. H. Brown, & M. R. Petersen (Eds.), *Primate communication* (pp. 50–62). Cambridge: Cambridge University Press.

(1988). Central control of monkey calls. In D. Todt, P. Goedeking, & D. Symmes (Eds.), *Primate vocal communication* (pp. 162–167). Berlin: Springer-Verlag.
Jürgens, U., & Ploog, D. (1976). Zur Evolution der Stimme. *Archiv für Psychiatrice und Nervenkrankheiten, 222,* 117–137.
Kaiser, L. (1962). Communication of affects by single vowels. *Synthese, 14,* 300–319.
Kappas, A. (1986). Der Einfluβ von Intonation und Stimmqualität auf die Attribution von Emotionen und Einstellungen. Unpublished master's thesis, University of Giessen, Department of Psychology.
Klos, T. Ellgring, H., & Scherer, K. R. (1987). *Vocal changes in depression.* Manuscript submitted for publication.
Knapp, M. L. (1978). *Nonverbal communication in human interaction* (2nd ed.). New York: Holt, Rinehart and Winston.
Kotlyar, G. M., & Morozov, V. P. (1976). Acoustical correlates of the emotional content of vocalized speech. *Journal of Acoustics of the Academy of Sciences of the USSR, 22,* 208–211.
Kramer, E. (1963). Judgement of personal characteristics and emotions from nonverbal properties of speech. *Psychological Bulletin, 60,* 408–420.
Kuroda, I., Fujiwara, O., Okamura, N., & Utsuki, N. (1976). Method for determining pilot stress through analysis of voice communication. *Aviation, Space, and Environmental Medicine, 47,* 528–533.
Ladd, D. R., Silverman, K. E. A., Tolkmitt, F., Bergmann, G., & Scherer, K. R. (1985). Evidence for the independent function of intonation contour type, voice quality, and F_0 range in signalling speaker affect. *Journal of the Acoustical Society of America, 78,* 435–444.
Laver, J. (1980). *The phonetic description of voice quality.* Cambridge: Cambridge University Press.
Levenson, R. (1988). Theoretical issues in the psychophysiology of emotion (structured panel discussion). *Psychophysiology, 25,* 422–423 (abstract).
Levin, H., & Lord, W. (1975). Speech pitch frequency as an emotional state indicator. *IEEE Transactions on Systems, Man, and Cybernetics, 5,* 259–273.
Levitt, E. A. (1964). The relationships between abilities to express emotional meaning vocally and facially. In J. R. Davitz (Ed.), *The communication of emotional meaning* (pp. 43–55). New York: McGraw-Hill.
Mahl, G. F., & Schulze, G. (1964). Psychological research in the extralinguistic area. In T. A. Sebeok, A. S. Hayes, & M. C. Bateson (Eds.), *Approaches to semiotics* (pp. 51–61). The Hague: Mouton.
Malatesta, C. Z., Fiore, M. J., & Messina, J. J. (1987). Affect, personality, and facial expressive characteristics of older people. *Psychology and Aging, 2,* 64–69.
Markel, N. N., Bein, M. F., & Phillis, J. A. (1973). The relationship between words and tone-of-voice. *Language and Speech, 16,* 15–21.
Marler, P. (1984). Animal communication: Affect or cognition? In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 345–368). Hillsdale, NJ: Erlbaum.
Marler, P., & Mitani, J. (1988). Vocal communication in primates and birds: Parallels and contrasts. In D. Todt, P. Goedeking, & D. Symmes (Eds.), *Primate vocal communication* (pp. 3–14). Berlin: Springer-Verlag.
Marler, P., & Tenaza, R. (1977). Signalling behavior of apes with special reference to vocalization. In T. A. Seboek (Ed.), *How animals communicate* (pp. 965–1033). Bloomington: Indiana University Press.

Mehrabian, A., & Ferris, S. (1967). Inference of attitudes from nonverbal communication in two channels. *Journal of Consulting Psychology, 31*, 248–252.
Morris, C. W. (1946). *Signs, language, and behavior*. Englewood Cliffs, NJ: Prentice-Hall.
Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *American Nature, 111*, 855–869.
Moses, P. J. (1954). *The voice of neurosis*. New York: Grune & Stratton.
Müller, A. L. (1960). *Experimentelle Untersuchungen zur stimmlichen Darstellung von Gefühlen*. Unpublished doctoral dissertation, University of Göttingen.
Newman, S. S., & Mather, V. G. (1938). Analysis of spoken language of patients with affective disorders. *American Journal of Psychiatry, 94*, 913–942.
Niwa, S. (1971). Changes of voice characteristics in urgent situations (2). *Reports of the Aeromedical Laboratory. Japan Air Self Defense Force, 11*, 246–251.
Noller, P. (1985). Video primacy – A further look. *Journal of Nonverbal Behavior, 9*, 28–47.
Ohala, J. J. (1978). Production of tone. In V. A. Fromkin (Ed.), *Tone: A linguistic survey* (pp. 5–39). New York: Academic Press.
(1980). The acoustic origin of the smile. *Journal of the Acoustical Society of America, 72*, 66 (abstract).
(1981). The nonlinguistic components of speech. In J. Darby (Ed.), *Speech evaluation in psychiatry* (pp. 39–49). New York: Grune & Stratton.
Ostwald, P. F. (1961). The sound of emotional disturbance. *Archives of General Psychiatry, 5*, 587–592.
(1963). *Soundmaking: The acoustic communication of emotion*. Springfield, IL: Thomas.
(1964). Acoustic manifestations of emotional disturbance. *Disorders of Communication – Research Publications, 42*, 450–465.
Peirce, C. S. (1931–1935). Collected papers of Charles Sanders Peirce, 8 vols., C. Hartshorne & P. Weiss (Eds.). Cambridge, MA: Harvard University Press.
Plaikner, D. (1970). *Die Veränderungen der menschlichen Stimme unter dem Einfluß psychischer Belastung*. Unpublished doctoral dissertation, University of Innsbruck.
Ploog, D. (1974). *Die Sprache der Affen und ihre Bedeutung für die Verständigungsweisen des Menschen*. Munich: Kindler.
(1986). Vocal expressions of emotion. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, research, and experience: Vol. 3. Biological foundations of emotion* (pp. 173–197). Orlando, FL: Academic Press.
(1988). Neurobiology and pathology of subhuman vocal communication and human speech. In D. Todt, P. Goedeking, & D. Symmes (Eds.), *Primate vocal communication* (pp. 195–212). Berlin: Springer-Verlag.
Robinson, W. P. (1979). Speech markers and social class. In K. R. Scherer & H. Giles (Eds.), *Social markers in speech* (pp. 211–249). Cambridge: Cambridge University Press.
Roessler, R., & Lester, J. W. (1976). Voice predicts affect during psychotherapy. *Journal of Nervous and Mental Disease, 163*, 166–176.
(1979). Vocal patterns in anxiety. In W. E. Fann, A. D. Pokorny, I. Koracau, & R. L. Williams (Eds.), *Phenomenology and treatment of anxiety* (pp. 225–235). New York: Spectrum.
Scherer, K. R. (1971). Randomized splicing: A note on a simple technique for masking speech content. *Journal of Experimental Research in Personality, 5*, 155–159.
(1974). Acoustic concomitants of emotional dimensions: Judging affects from synthesized tone sequences. In S. Weitz (Ed.), *Nonverbal communication* (pp. 105–111). New York: Oxford University Press.
(1977). Affektlaute und vokale Embleme. In R. Posner and H. P. Reinecke (Eds.), *Zeichenprozesse – Semiotische Forschung in den Einzelwissenschaften* (pp. 199–214). Wiesbaden: Athenaion.
(1978). Personality inference from voice quality: The loud voice of extraversion. *European Journal of Social Psychology, 8*, 467–487.
(1979a). Nonlinguistic vocal indicators of emotion and psychopathology. In C. E. Izard (Ed.), *Emotions in personality and psychopathology* (pp. 493–529). New York: Plenum.
(1979b). Voice and speech correlates of perceived social influence. In H. Giles & R. St. Clair (Eds.), *The social psychology of language*. London: Blackwell Publisher.
(1980). Personality, emotion, psychopathology and speech. In H. Giles, W. P. Robinson, & P. M. Smith (Eds.), *Language: Social psychological perspectives* (pp. 233–235). Oxford: Pergamon.
(1981a). Speech and emotional states. In J. Darby (Ed.), *Speech evaluation in psychiatry* (pp. 189–220). New York: Grune & Stratton.
(1981b). Vocal indicators of stress. In J. Darby (Ed.), *Speech evaluation in psychiatry* (pp. 171–187). New York: Grune & Stratton.
(1982). Methods of research on vocal communication: Paradigms and parameters. In K. R. Scherer & P. Ekman (Eds.), *Handbook of methods in nonverbal behavior research* (pp. 136–198). Cambridge: Cambridge University Press.
(1984). On the nature and function of emotion: A component process approach. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 293–318). Hillsdale, NJ: Erlbaum.
(1985). Vocal affect signalling: A comparative approach. In J. Rosenblatt, C. Beer, M.-C. Busnel, & P. J. B. Slater (Eds.), *Advances in the study of behavior* (Vol. 15, pp. 189–244). New York: Academic Press.
(1986a). Vocal affect expression: A review and a model for future research. *Psychological Bulletin, 99*, 143–165.
(1986b). Voice, stress, and emotion. In M. H. Appley & R. Trumbull (Eds.), *Dynamics of stress* (pp. 159–181). New York: Plenum.
(1987). Vocal assessment of affective disorders. In D. Kupfer & J. D. Maser (Eds.), *Depression and expressive behavior*. Hillsdale, NJ: Erlbaum.
(1988). Criteria for emotion-antecedent appraisal: A review. In V. Hamilton, G. H. Bower, & N. H. Frijda (Eds.), *Cognitive perspectives on emotion and motivation* (pp. 89–126). Dordrecht: Kluwer Academic Publishers.
(1989). Vocal correlates of emotion. In H. Wagner & A. Manstead (Eds.), *Handbook of psychophysiology: Emotion and social behavior* (pp. 165–197). London: Wiley.
Scherer, U., Helfrich, H., & Scherer, K. R. (1980). Paralinguistic behaviour: Internal push or external pull. In H. Giles, W. P. Robinson, & P. M. Smith (Eds.), *Language: Social psychological perspectives* (pp. 279–282). Oxford: Pergamon.
Scherer, K. R., & Kappas, A. (1988). Primate vocal expression of affective state. In D. Todt, P. Goedeking, & D. Symmes (Eds.), *Primate vocal communication* (pp. 171–194). Berlin: Springer-Verlag.
Scherer, K. R., Koivumaki, J., & Rosenthal, R. (1972). Minimal cues in the vocal communication of affect: Judging emotions from content-masked speech. *Journal of Psycholinguistic Research, 1*, 269–285.

Scherer, K. R., Ladd, D. R., & Silverman, K. E. A. (1984). Vocal cues to speaker affect: Testing two models. *Journal of the Acoustical Society of America, 76*, 1346–1356.
Scherer, K. R., & Oshinsky, J. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion, 1*, 331–346.
Scherer, K. R., & Scherer, U. (1981). Speech behavior and personality. In J. Darby (Ed.), *Speech evaluation in psychiatry* (pp. 115–135). New York: Grune & Stratton.
Scherer, K. R., & Wallbott, H. G. (in preparation). *Development of a standardized vocal emotion recognition test.*
Scherer, K. R., Wallbott, H. G., & Summerfield, A. B. (Eds.). (1986). *Experiencing emotion: A cross-cultural study.* Cambridge: Cambridge University Press.
Scherer, K. R., Wallbott, H. G., Tolkmitt, F., & Bergmann, G. (1985). *Die Streß-reaktion: Physiologie und Verhalten.* Göttingen: Hogrefe.
Sedlacek, K., & Sychra, A. (1963). Die Melodie als Faktor des emotionellen Ausdrucks. *Folia Phoniatrica, 15*, 89–98.
Seyfarth, R. M., & Cheney, D. L. (1982). How monkeys see the world: A review of recent research on East African vervet monkeys. In C. T. Snowdon, C. H. Brown, & M. R. Petersen (Eds.), *Primate communication* (pp. 239–252). Cambridge: Cambridge University Press.
Shannon, C., & Weaver, W. (1949). *The mathematical theory of communication.* Champaign: University of Illinois Press.
Siegman, A. W., & Feldstein, S. (1987). *Nonverbal behavior and communication* (2nd ed.). Hillsdale, NJ: Erlbaum.
Simonov, P. V., & Frolov, M. V. (1973). Utilization of human voice for estimation of man's emotional stress and state attention. *Aerospace Medicine, 44*, 256–258.
Sincoff, J. B., & Rosenthal, R. (1985). Content-masking methods as determinants of results of studies of nonverbal communication. *Journal of Nonverbal Behavior, 9*, 121–129.
Skinner, R. E. (1935). A calibrated recording and analysis of the pitch, force, and quality of vocal tones expressing happiness and sadness. And a determination of the pitch and force of the subjective concepts of ordinary, soft, and loud tones. *Speech Monographs, 2*, 81–137.
Smith, P. M. (1979). Sex markers in speech. In K. R. Scherer & H. Giles (Eds.), *Social markers in speech* (pp. 109–146). Cambridge: Cambridge University Press.
(1980). Judging masculine and feminine social identities from content-controlled speech. In H. Giles, W. P. Robinson, & P. M. Smith (Eds.), *Language: Social psychological perspectives* (pp. 121–126). Oxford: Pergamon.
Standke, R., Kappas, A., & Scherer, K. R. (1984). Die Attribution von Stimm und Sprechermerkmalen in Abhängigkeit von Stimmqualität und Kontinuität des Sprechflusses. In A. Schick & K. P. Walcher (Eds.), *Beiträge zur Bedeutungslehre des Schalls* (pp. 389–402). Bern: Lang.
Streeter, L. A., Krauss, R. M., Geller, V., Olson, C., & Apple, W. (1977). Pitch changes during attempted deception. *Journal of Personality and Social Psychology, 35*, 345–350.
Sulc, J. (1977). To the problem of emotional changes in the human voice. *Activitas Nervosa Superior, 19*, 215–216.
Tartter, V. C. (1980). Happy talk: Perceptual and acoustic effects of smiling on speech. *Perception and Psychophysics, 27*, 24–27.

Tembrock, G. (1971). *Biokommunikaton: Informationsübertragung im biologischen Bereich.* Berlin: Akademie-Verlag.
(1975). Die Erforschung des tierlichen Stimmausdruckes (Bioakustik). In F. Trojan (Ed.), *Biophonetik* (pp. 51–68). Mannheim: Bibliographisches Institut.
Titze, I. R. (1974). The human vocal cords: A mathematical model (pt. 2). *Phonetica, 29*, 1–21.
Titze, I. R., & Talkin, D. T. (1979). A theoretical study of the effects of various laryngeal configurations on the acoustics of phonation. *Journal of the Acoustical Society of America, 66*, 60–74.
Todt, D. (1988). Serial calling as a mediator of interaction processes: Crying in primates. In D. Todt, P. Goedeking, & D. Symmes (Eds.), *Primate vocal communication* (pp. 88–107). Berlin: Springer-Verlag.
Tolkmitt, F. J., Helfrich, H., Standke, R., & Scherer, K. R. (1982). Vocal indicators of psychiatric treatment effects in depressives and schizophrenics. *Journal of Communication Disorders, 15*, 209–222.
Tomkins, S. S. (1962). *Affect, imagery, consciousness: Vol. 1. The positive affects.* New York: Springer-Verlag.
Utsuki, N., & Okamura, N. (1976). Relationship between emotional state and fundamental frequency of speech. *Reports of the Aeromedical Laboratory, Japan Air Self Defense Force, 16*, 179–188.
Van Bezooijen, R. (1984). *The characteristics and recognizability of vocal expressions of emotion.* Dordrecht: Foris.
Wallbott, H. G. (1979). Beziehungen zwischen Signalsystemen: Einführung. In K. R. Scherer & H. G. Wallbott (Eds.), *Nonverbale Kommunikation* (pp. 187–192). Weinheim: Beltz.
(1988). Faces in context: The relative importance of facial expression and context information in determining emotion attributions. In K. R. Scherer (Ed.), *Facets of emotion: Recent research* (pp. 139–160). Hillsdale, NJ: Erlbaum.
Wallbott, H. G., & Scherer, K. R. (1986a). Cues and channels in emotion recognition. *Journal of Personality and Social Psychology, 51*, 690–699.
(1986b). How universal and specific is emotional experience? Evidence from 27 countries on five continents. In K. R. Scherer (Ed.), *Facets of emotion: Recent research* (pp. 31–56). Hillsdale, NJ: Erlbaum.
Whitehead, J. M. (1987). Vocally mediated reciprocity between neighbouring groups of mantled howling monkeys (*Alouatta palliata palliata*). *Animal Behavior, 35*, 1615–1627.
Whitman, E. N., & Flicker, D. J. (1966). A potential new measurement of emotional state: A preliminary report. *Newark Beth Israel Hospital, 17*, 167–172.
Williams, C. E., & Stevens, K. N. (1969). On determining the emotional state of pilots during flight: An exploratory study. *Aerospace Medicine, 40*, 1369–1372.
(1972). Emotions and speech: Some acoustical correlates. *Journal of the Acoustical Society of America, 52*, 1238–1250.
Wolf, J. J. (1981). Acoustic analysis and computer systems. In J. Darby (Ed.), *Speech evaluation in psychiatry* (pp. 105–110). New York: Grune & Stratton.
Wundt, W. (1903). *Grundzüge der physiologischen Psychologie* (Vol. 3). Leipzig: Wilhelm Engelmann.
Zaidel, S., & Mehrabian, A. (1969). The ability to communicate and infer positive

and negative attitudes facially and vocally. *Journal of Experimental Research in Psychology, 34*, 966–977.

Zuberbier, E. (1957). Zur Schreib- und Sprechmotorik der Depressiven. *Zeitschrift für Psychotherapie und Medizinische Psychologie, 7*, 239–249.

Zwirner, E. (1930). Beitrag zur Sprache des Depressiven. *Phonometrie III, Spezielle Anwendungen I* (pp. 171–187). Basel: Karger.