Journal of Voice
Vol. 9, No. 3, pp. 235–248
© 1995 Lippincott-Raven Publishers, Philadelphia

Expression of Emotion in Voice and Music

Klaus R. Scherer

Department of Psychology, University of Geneva, Geneva, Switzerland

Summary: Vocal communication of emotion is biologically adaptive for socially living species and has therefore evolved in a phylogenetically continuous manner. Human affect bursts or interjections can be considered close parallels to animal affect vocalizations. The development of speech, unique to the human species, has relied on the voice as a carrier signal, and thus emotion effects on the voice become audible during speech. This article reviews (a) the evidence on listeners' ability to accurately identify a speaker's emotion from voice cues alone, (b) the research efforts trying to isolate the acoustic features that determine listener judgments, and (c) the findings on actual acoustic concomitants of a speaker's emotional state (real or portrayed by actors). Finally, based on speculations about the joint origin of speech and vocal music in nonlinguistic affect vocalizations, similarities of emotion expression in speech and music are discussed. **Key Words:** Affect vocalization—Emotional expression—Prosody—Singing—Nonverbal communication—Emotion effects on voice.

Emotions produce pervasive, although generally short-lived, changes in the organism as a whole. They represent reactions to events of major significance to the individual and mobilize all resources to cope with the respective situation, positive or negative. The most significant feature of the emotion mechanism is that it produces specific action readiness while providing a latency period that allows adaptation of the behavioral reactions to the situational demands (1–3). One of the major uses of this latency period in socially living species is to predict the likely reaction of others to an action that is "ready" for execution as the result of a particular emotional state. For example, before actually engaging in an aggressive act, I can first shout angrily at someone who has offended me and back down if it turns out that the other is likely to retaliate violently.

Accepted February 3, 1994.
Address correspondence and reprint requests to Dr. K. Scherer at Department of Psychology, University of Geneva, 9, route de Drize, CH-1227 Carouge-Geneva, Switzerland.
This article represents the written version of a keynote address to the 22nd Annual Symposium on Care of the Professional Voice, Philadelphia, June 7–12, 1993.

Thus, as first demonstrated in Darwin's classic work on the expression of emotion in humans and animals (4), emotional expression serves the vital function of externalizing an individual's reaction and action propensity and of communicating this information to the social environment. Just as emotion is phylogenetically continuous, found in more or less rudimentary form in many, particularly mammalian, species, so is emotional expression, particularly in species in which social life is based on complex interactions among individuals. All expressive modalities, particularly body posture, facial features, and vocalization, are involved in emotion communication. As far as facial expression is concerned, behavioral and social scientists have made much progress in gathering evidence on the phylogenetic continuity (5), the universality across cultures (6,7), and the rich information content of emotion in facial expressions (8).

Although less frequently studied than facial expression, the vocal communication of emotion has also been under scrutiny in the biological and psychological sciences. The large majority of all animal vocalizations are affective in nature while, at the same time, serving representational functions

(9,10). Research on animal communication has demonstrated that in many species affective states, generally linked to changes in physiological arousal, are externalized in vocalizations and serve specific communication functions, often involving acoustic patterns that are similar across species (11–15). In close parallel to animal affect vocalizations, we still find rudiments of nonlinguistic human affect vocalizations, often referred to as "interjections," such as "ouch," "ai," "oh," "yuck," etc. Kleinpaul (16) claimed that these reflexive "nature and feeling sounds" sound very much the same when uttered by speakers in different cultures. He insisted on a sharp distinction between spontaneously occurring interjections or exclamations expressing an emotional state and calls or shouts intentionally uttered for communicative reasons. Wundt (17) traced these vocalizations back to inarticulate screams and cries accompanying very intense feelings of aversion, rage, and fear. He distinguished between (a) primary interjections, defined as "nature sounds" and (b) secondary interjections, which become assimilated into the language. Kainz (18) claimed that as civilization advances, emotions are less and less frequently expressed by means of pure "nature sounds" but rather by interjections that have been assimilated into language.

Linguists interested in speech have also discussed these "prelinguistic fragments" in the flow of speech. James (19) reviews the respective writings of some of the classic authors in the field (Bloomfield, Fries, Jespersen, Sapir) who all agreed on the affective significance of interjections, their "primitive" and noncommunicative nature, and their lack of grammatical structure. Scherer (20) focused on the *functions* of these affect vocalizations. One can show that these serve practically all of the semantic, syntactic, pragmatic, and dialogical functions of nonverbal behavior in conversation [see Scherer (21) for a more detailed discussion of these functions]. Goffman (22) has provided an analysis of interjections from an interactionist view. He defines exclamatory, nonlexicalized, discrete interjections as "response cries," expressions that he sees as "a natural overflowing, a flooding up of previously contained feeling, a bursting of normal restraints, a case of being caught off-guard." An extensive discussion of such affect vocalizations or "facial/vocal affect bursts" can be found in Scherer (23).

Historically, there have been many suggestions by both philosophers, e.g., Rousseau and Herder, and natural scientists, e.g., Helmholtz, that both protospeech and protomusic have evolved from such primitive affect vocalizations. Scherer (24) has suggested following up on these early suggestions. He cites ethological work having shown that expression and impression are closely linked (25,26), suggesting that, in the process of conventionalization and ritualization, expressive signals may have been shaped by the constraints of transmission characteristics, limitations of sensory organs, or other factors. The resulting flexibility of the communication code may have fostered the evolution of more abstract, symbolic language and music systems. It is noted that this development is likely to have occurred in close conjunction with the evolution of the brain. Just as newer neocortical structures with highly cognitive modes of functioning have been superimposed on older "emotional" structures such as the limbic system, the evolution of human speech as a digital system of information encoding and transmission (and of musical scales and conventions for singing) has made use of the more primitive, analogue vocal affect signaling system as a carrier signal.

As vocalization, which remained a major modality for analog emotion expression, became the production system for the highly formalized, segmental systems of language and singing, both of these functions needed to be served at the same time. Thus, in speech, changes in fundamental frequency (F_0), formant structure, or characteristics of the glottal source spectrum can, depending on the language and the context, serve to communicate phonological contrasts, syntactic choices, pragmatic meaning, or emotional expression. Similarly, in music, melody, harmonic structure, or timing may reflect the composer's intentions, depending on specific traditions of music, and may simultaneously induce strong emotional moods (27–29). This fusion of two signal systems, which are quite different in function and in structure, into a single underlying production mechanism, vocalization, has proven to be singularly efficient for the purpose of communication.

In this article it is claimed that emotions differentially shape human vocal expression in speech and music and that listeners are capable of correctly inferring a speaker's emotional state (or an actor/singer's attempt to portray such a state) from the voice alone. It is instructive to start with the perception of emotion. If it is demonstrated that emotion can be correctly diagnosed from the voice, then clearly the emotions must differentially affect the

vocalization mechanism and, in consequence, yield demonstrable differences in acoustic patterning of the resulting sound waves. That the human voice not only permits judging the speaker's emotion but can also *induce* affect in the listener has been held as self-evident throughout history. In particular, ever since antiquity, different schools of rhetoric have insisted on the powerful effect of emotional expression in the voice on the listener (Cicero, Quintilian).

Even though such effects may seem evident, behavioral scientists require empirical evidence that, indeed, listeners are able to correctly recognize the speaker's emotional state from vocal cues alone, independent of information from situational context or other expressive cues, such as facial expressions, gestures, or posture. The evidence to date is reviewed in the following sections.

JUDGING EMOTIONAL STATES FROM THE VOICE

During the last 50 years, many studies have examined listeners' ability to correctly recognize or infer speaker affect state or attitude from voice samples. In general, researchers have asked speakers (often actors, both amateur and professional) to vocally portray different emotional states while producing a standard utterance (e.g., numbers, letters of the alphabet, nonsense syllables, or standardized sentences). The voice samples are recorded and later presented to lay judges, who are to identify which emotion is expressed in each of the different portrayals.

In a comprehensive review of these studies, Scherer (30) reported an average accuracy of ~60%. This is well above what one would expect to obtain if the listener judgments were based exclusively on guessing, i.e., on chance (~12%). The degree of recognition accuracy is impressive given that some of the studies included emotions such as love, pride, or jealousy, which are not part of the set of basic or fundamental emotions (e.g., anger, joy, sadness, fear).

Two recent studies confirm the earlier estimate on vocal emotion recognition accuracy. In a study of disgust, surprise, shame, interest, joy, fear, sadness, and anger, von Bezooijen (31) found a mean accuracy of 65%. Based on another recent series of studies of five emotions—fear, joy, sadness, anger, and disgust—using different types of listener groups, Scherer et al. (32) reported a mean accuracy of 56%. In consequence—after correction for chance guessing and sampling error—the recognition of emotion from standardized voice samples, using actor portrayals, seems to lie at ~50%, approximately four to five times higher than what would be expected by chance.

Not all of the different emotions are identified equally well. Sadness and anger are best recognized, followed by fear and joy. Disgust is the worst, with the accuracy barely above chance. The data show the need to analyze the recognizability of different emotions separately using confusion matrices, because errors are not randomly distributed and the patterns of misidentification provide important information on the judgment process. Also, cross-cultural studies are needed to elucidate the role of language and culture in this process. Frick (33) reviewed work in this area and concluded on the basis of the data that the vocal expression of at least some emotions seems to be universal.

IDENTIFYING THE ACOUSTIC CUES USED IN EMOTION INFERENCE FROM VOICE

If listeners are able to recognize vocally portrayed emotions with better than chance accuracy, one should be able to determine which acoustic cues they perceive and use in the process of attributing emotion to a speaker. In fact, even if the listeners' emotion inference is *inaccurate*, one would still want to know which vocal cues are attended to and how they are interpreted with respect to the underlying emotional state. Different research strategies have been used to determine the importance of various acoustic cues in the judgment process.

1. It is possible to use electroacoustic or digital equipment and/or voice experts to measure the acoustic and/or phonatory–articulatory characteristics of the vocal emotion portrayals, and to then correlate these with the listeners' judgments of underlying emotion or attitude of the speaker. Several studies of this type have yielded information on which vocal characteristics affect the judges' inference (31,34).
2. Another approach consists of partially or completely masking particular verbal/vocal cues to identify which information is carried by these characteristics. One of the best known techniques is low-pass filtering of the speech sample (at ~300–400 Hz, restricting the informa-

tion to the fundamental frequency range). This eliminates the intelligibility of speech and allows studying the extent to which vocal cues in the very low-frequency range (particularly temporal structure, dynamic loudness, and intonation cues) carry affective information (35,36). Scherer et al. (37) have compared different masking techniques (electronic filtering, randomized splicing, playing backwards, pitch inversion, and tone-silencing coding). Each of these techniques removes and/or preserves different combinations of acoustic characteristics of a vocal expression. In consequence, the systematic use of several techniques permit determination of which acoustic cues carry which type of emotional information. Because intelligibility is removed by all of these masking procedures, their application permits the use of natural speech material in judgment and analysis studies, including excerpts from real social interactions with "real-life" rather than artificially posed emotions.

For example, Scherer et al. (38) used a corpus of affectively laden utterances by civil servants in interaction with citizens to determine which acoustic cues are used by listeners to infer speaker emotion and attitude. Different groups of judges were presented with full audio records, with electronically filtered, randomly spliced, and reversed versions, as well as with verbal transcripts. Virtually all of the affective information was contained in the vocal expression samples (textual cues allowing only an inference of aggressiveness contained in the utterance). Arousal and uncertainty of the speakers were well captured in several masking conditions and seemed to be communicated by F_0 variability and F_0 mean, respectively.

One of the major issues explored in this study was the nature of the relationship between vocal cues and pragmatic or semantic information. A number of acoustic parameters, in particular F_0 mean or floor and F_0 variability, varied quite continuously with speaker arousal, suggesting a linear covariation or correlation between psychophysiological speaker state and listener judgment. However, this was not the case for one of the prosodic variables studied—intonation contour—where there seemed to be an interaction with the linguistic form of the respective utterance.

3. The recent development of electronic sound synthesis procedures allows systematic experimental manipulation of different acoustic cues. In an early study, Lieberman and Michaels (39) studied the effect of systematic variations of F_0 and envelope contour on emotion inference. Using the MOOG synthesizer, Scherer and Oshinsky (40) studied the effects of amplitude variation, pitch level, contour and variation, tempo, envelope, harmonic richness, tonality, and rhythm on emotional attributions to sentence-like sound sequences and musical melodies. Table 1 shows the results of this study. The systematic design of the study allows statistical determination of the importance or power of the respective cue for the listener judgment. Tempo of the sounds in the sequence and filtration level (i.e., number of audible higher harmonics) were by far the most powerful cues.

TABLE 1. *Emotional attributions significantly associated with acoustic parameters (reproduced from ref. 40, p. 339)*

Acoustic parameters of tone sequences	Direction of effect	Emotion rating scales listed in decreasing order of associative strength
Amplitude variation	Small	Happiness, pleasantness, activity
	Large	Fear
Pitch variation	Small	Disgust, anger, fear, boredom
	Large	Happiness, pleasantness, activity, surprise
Pitch contour	Down	Boredom, pleasantness, sadness
	Up	Fear, surprise, anger, potency
Pitch level	Low	Boredom, pleasantness, sadness
	High	Surprise, potency, anger, fear, activity
Tempo	Slow	Sadness, boredom, disgust
	Fast	Activity, surprise, happiness, pleasantness, potency, fear, anger
Envelope	Round	Disgust, sadness, fear, boredom, potency
	Sharp	Pleasantness, happiness, surprise, activity
Filtration cutoff	Intermediate (few)	Pleasantness, boredom, happiness, sadness
Level (number of harmonics)	High (many)	Potency, anger, disgust, fear, activity, surprise

Computer-based copy synthesis (or resynthesis) techniques make it possible to take natural voices and systematically change different cues via digital manipulation of the sound waves. In a number of recent studies by Scherer and collaborators, F_0 level, contour variability and range, intensity, duration, and accent structure of real utterances have been systematically manipulated (41,42).

These manipulated voice samples were presented to listeners who judged the apparent speaker attitude and emotional state for each stimulus. The results showed strong direct effects for all of the variables manipulated on these ratings. Relatively few effects due to interactions between the manipulated variables were found. This implies that the synthesized variables independently influenced judges' ratings. Only very minor effects for speaker and utterance content were found, indicating that the results are likely to generalize over different speakers and utterances.

Of all variables studied, F_0 range had the most powerful effect on judgments. Narrow F_0 range was seen as a sign of sadness or of absence of specific speaker attitudes. Wide F_0 range was consistently judged as expressing high arousal, producing attributions of strong negative emotions such as annoyance or anger, or for the presence of strongly developed speaker attitudes such as involvement, reproach, or emphatic stress. Furthermore, the data supported the hypothesis, derived from the earlier study by Scherer et al. (38), that these effects should be continuous, i.e., yielding a linear relationship between the size of F_0 range and the strength of emotion attribution (Fig. 1). High intensity was interpreted in terms of negative affects or aggressive speaker attitudes. Short voiced segment duration (fast tempo) was correlated with inferences of joy, and long duration (slow tempo) was correlated with inferences of sadness (see refs. 41 and 42 for details of the procedures and results).

In a further study (experiment IV in ref. 41), checking on the stability of the effects obtained, a highly trained speaker was asked to portray different emotions while producing a standard utterance. Emotion-specific acoustic characteristics were digitally resynthesized in such a way as to turn the utterance into a neutral rendering. The relative success of this procedure attests to the important role of F_0 range, intensity, and tempo on the judgment of speaker emotions and attitudes by listeners.

So far, the emphasis has been placed on the *recognition* of a speaker's emotion from the voice, assuming that there is a clear criterion for the nature of the emotion present (or, as in most research studies, of an actor's encoding intention). However, the early rhetoricians have claimed that powerful emotional oratory, using voice effects in addition to verbal appeal, is able to *induce* emotion. Unfortunately, there has been very little research on this issue. A possible induction mechanism might be represented by a process of empathy via motor mimicry as suggested by Lipps (43, pp. 228–231; see also refs. 44, 45).

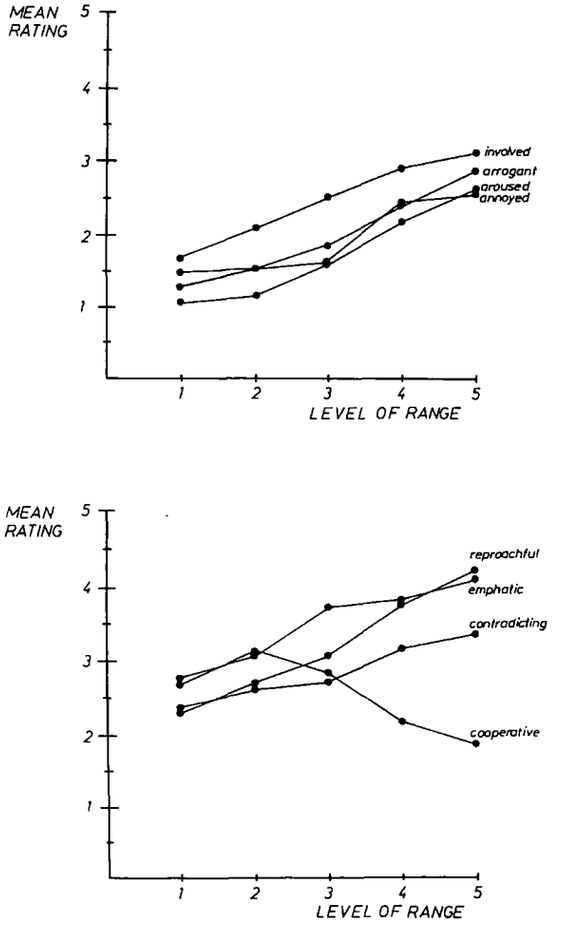

FIG. 1. Speaker emotionality and attitude ratings as a function of extent of resynthesized F_0 range (reproduced from ref. 42, p. 441).

THE EXTERNALIZATION OF EMOTION IN THE VOICE

The review of the evidence on accurate decoding of emotion from the voice has shown that the infor-

mation is obviously there, confirming the early hunches from classical rhetoric. It remains to be shown, then, how emotional arousal affects the vocal organs and to identify emotion-specific acoustic patterning.

Emotion has been defined as consisting of a synchronization of changes in all organismic subsystems, thus accounting for the pervasiveness and power of emotional states (46). Because vocalization reflects the activity of many different aspects of the functioning of the nervous and somatic systems, one would expect many different determinants of emotion effects on the voice. Some of the major mechanisms involved will be briefly outlined below (see refs. 3, 47, 48 for more detailed discussions).

Figure 2 shows some of the main effects of the major neurophysiological structures on the voice production mechanisms. Speech production is mostly controlled by the neocortex. Specific motor commands produce appropriate phonatory and articulatory movements for the desired sequence of speech sounds, including intentionally produced prosodic features (intonation, voice quality). The intended vocal effects are mostly produced by *phasic* activation of the muscles serving phonation and articulation.

The effects of emotional arousal on the vocalization process are primarily controlled by the limbic system (11,49). They are generally produced via *tonic* activation in the somatic nervous system (in particular the striated musculature) and sympathetic as well as parasympathetic activation of the autonomous nervous system. In addition, direct sympathetic or parasympathetic effects such as respiration changes and the secretion of mucus can affect the production of the vocalization.

Given the manifold determinants of voice production processes, even slight changes in physiological regulation will produce variations in the acoustic pattern of the speech waveform. This is shown by the fact that even if a speaker attempts to reproduce a particular utterance in exactly the same way immediately after having spoken it for the first time, some changes are likely to occur. The enormous sensitivity of the acoustic output to minor changes in voice production settings provides a sensitive and rapidly responding system for monitoring emotional arousal, but also introduces a high degree of noise into the system.

So far, it may seem as though acoustic patterns automatically mirror emotion-produced changes of the internal physiological system. However, because vocalization has developed in part as a social communicative signaling system, the simple externalization of internal states has been supplemented by display mechanisms producing a specific impression in the listener (independent of internal state). For example, in highly constrained, formal situations requiring politeness or cheerfulness, one will produce the appropriate pleasant voice quality in spite of internally boiling with rage . In the course of the evolution of expressive communication systems, impression models have molded the nature of expression (see ref. 26, for a thorough discussion of this point) and human vocalization is also partially determined by such social "display rules" (8,17). Scherer et al. (14,50,51) have introduced the distinction between *push* effects and *pull* effects to distinguish among the determining factors that operate on vocalization.

Push effects are produced by the physiological changes that accompany emotional arousal and that consequently change the voice production mechanism in predictable ways (e.g., increased tension of the laryngeal muscles producing higher fundamental frequency of the voice). Pull effects, on the other hand, are independent of the internal physiological processes in the organism. Their origin is found in

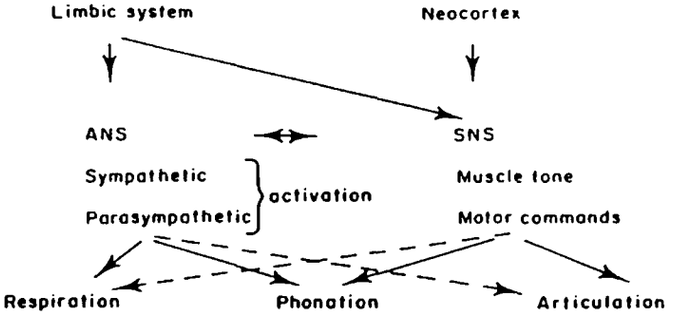

FIG. 2. Effects of neurophysiological structures on voice production mechanisms (reproduced from Scherer, 1989, p. 169).

external factors, such as ritualized or conventionalized acoustic signal patterns, that are required to ensure information transfer, constraints on the acoustic signal structure imposed by a communication channel or the environment, or the need for self-presentation (given the impression formation rules of the listeners). In most cases, the acoustic nature of a vocalization, particularly in humans, is determined by both types of effects: the effects of emotion-related physiological changes internal to the organism, and effects of external constraints or social target patterns. Given the difficulty of disentangling the two types of effects, most studies so far have not differentiated between push and pull effects.

Empirical studies on the acoustic patterns of emotional expression in the voice, using objective measurement, have been reported early on (see refs. 3, 30, 33, 47, 48, 52, and 53 for reviews of this early work). Many of these studies have been motivated by the obvious diagnostic implications of vocal analysis, particularly with respect to detecting stress (54–56), differentiating types of emotional disorders (depression, schizophrenia), and evaluating treatment effects (see ref. 48 for a review).

Work on the acoustic measurement of natural emotion expression has been handicapped by the practical and ethical difficulties of producing strong emotions in the laboratory. There are also few examples of impromptu recorded instances of vocal emotion expression in real life (such as the famous recording of a radio reporters' emotional commentary on the explosion of the Zeppelin "Hindenburg" in 1937; see ref. 52). In consequence, many researchers in this area have used actors as subjects, asking them to vocally portray different emotions, and have analyzed the acoustic features of the recorded portrayals. Although the actors in some of the studies used Stanislavski-like induction techniques, trying to at least partially produce the emotion to be portrayed, one cannot be sure that the results correspond to genuine emotion expression in the voice (see refs. 32 and 34 for a discussion of this point). The present state of the evidence, which is mostly based on studies with actors, can be summarized as follows (see ref. 57 for details):

Anger: Anger is vocally expressed by an increase in mean F_0 and mean intensity. Some studies, which may have been measuring "hot" rather than "cold" anger, also claim higher F_0 variability and a wider range of F_0. Further anger signs seem to be increases in high- frequency energy and downward-directed F_0 contours. The rate of articulation usually increases in anger.

Fear: One expects a very high arousal level for fear. Consistent with this hypothesis, the data show increases in mean F_0, in F_0 range, and in high-frequency energy. Rate of articulation increases. In some studies, higher mean F_0 is also reported for the weaker forms of fear (i.e., worry or anxiety).

Sadness: In sadness, mean F_0, F_0 range, and mean intensity all decrease, and F_0 contours are generally downward directed. High-frequency energy and rate of articulation decrease.

Joy: Studies consistently show increases in mean F_0, F_0 range, F_0 variability, and mean intensity. There is some evidence of an increase in high-frequency energy and in rate of articulation.

Whenever the findings in the literature converge, they are related to autonomic arousal. The vocal cues that indicate arousal or activation are similar across different emotions. So far, there is relatively little evidence for the vocal differentiation of individual emotions on other dimensions such as valence (3). However, because judges are able to recognize the individual emotions on the basis of vocal cues alone, there must be acoustic characteristics that differentiate the various emotions in addition to indicating arousal. So far, the number of acoustic cues that are measured in studies in this area has been very limited indeed. More parameters will have to be assessed, particularly measures of energy distribution in the spectrum. Most importantly, researchers in this area need to differentiate emotional states much more precisely, particularly with respect to *subdued* and *aroused* types of the same emotion family. For example, the *anger* family contains both *cold anger,* a subdued form, and *hot anger,* a highly aroused form of anger with highly different physiological and expressive characteristics (3,50). It is to be expected that there will be a sizable increase in the discriminative power of the acoustic parameters when both of these requirements are met.

Most of the studies in this area have attempted to find acoustic correlates for categorically labeled affect states. It is possible that emotional states are not that clearly differentiated. In fact, many emotion psychologists have commented on the strong probability that most real-life emotions are actually blends of different *basic* emotions or episodic processes with rapidly changing affective states (1,2). This could be equally true for emotions induced in the laboratory or portrayed by actors. Cosmides

(58) has conducted a study that does not limit actors to portray a specific emotion label but requires them to express affective responses in complex emotional situations taken from fiction. Her results show that despite the absence of clear labels, the specificity of the acoustic rendering of different affect situations is far stronger than individual differences in the portrayals. This approach suggests a promising alternative to studying categorically labeled states, an alternative that is very much in line with recent theoretical work that views emotional episodes as dynamic processes of continuous adaptation of the organism to a changing environment (as based on appraisal or evaluation of events in terms of their significance to the individual; see refs. 3 and 50).

The need to look at emotional expression in the voice in terms of dynamic temporal sequences, rather than as aggregate steady states over a complete utterance, is highlighted by a recent study by Tischer (59). This researcher is one of the first who has looked in detail at the microchanges in vocal expression over the course of a short utterance, particularly with respect to effects on listener judgments. He is able to show that for several emotions, the role of the acoustic characteristics changes dramatically depending on their temporal locations in the utterance.

EMOTION EXPRESSION IN VOCAL MUSIC

The notion that music is "the language of the emotions" has a very long tradition. Composers, musicologists, as well as philosophers and psychologists interested in music, have attempted—often in a highly controversial fashion—to delineate the factors that are responsible for emotional expression in music (60–67).

Helmholtz, one of the pioneers of musical acoustics, wrote: "An endeavour to imitate the involuntary modulations of the voice, and make its recitation richer and more expressive, may therefore possibly have led our ancestors to the discovery of the first means of musical expression, just as the imitation of weeping, shouting, or sobbing, and other musical delineations may play a part in even cultivated music (as in operas), although such modifications of the voice are not confined to the actions of free mental motives, but embrace really mechanical and even involuntary muscular contractions" (68, p. 371). Like Helmholtz, many scholars have given in to the temptation to speculate about a phylogenetic continuity for affect expression, speech, and music (24).

If the origin of music is indeed to be sought in the emotional expressions of the human voice, it should be human vocal music—singing—that should be most prone to evoke strong emotional feelings in the listener. A special case is represented by opera singing. To the emotional expressivity of the musical score and the affective message of the singing voice, opera adds all the emotional power of a dramatic plot and of the personality of the protagonist. An appreciation of the emotional expressivity of opera can be found in the writings of many composers and musicologists alike (27–29,63,66) and seems to be shared by the opera-going public at large (although there may be a category of music listeners who feel alienated by the musical and dramatic circumstances in opera and who react more emotionally to other types of vocal music).

There are very few studies that analyze the role of the interpreter, the singer, in vocally projecting the affective state of the character enacted on the operatic stage. Obviously, one of the reasons for this is that there is an enormous amount of factors involved: the music itself, the psychological interpretation of both the action and the characters by the director, the timing of the conductor, as well as the intuition of the singer, his/her empathy with the character, and the atmosphere created by the audience. It seems difficult to approach this question via standard research designs as they are used in the work on emotional expression in speech.

Three of the large number of potential factors will be outlined below. First, as in the case of a speaker's respiration, phonation, and articulation being affected by emotion-driven physiological changes, we can assume that the singer's own underlying emotional state during a performance will have similar effects on the singing voice. It seems reasonable to assume that specific physiological changes will have similar effects on the speaking and the singing voice. Unfortunately, as for affect expression in the speaking voice, there are virtually no systematic studies on naturally occurring, strong emotional states in singers and their effects on the voice. Such studies would be even more difficult to conduct than in the speech domain, because most singing activity seems to be confined to the performing arts—contrary to speech, which is used systematically for interpersonal communication in most social situations. The performance aspect of singing,

even if it is not necessarily linked to the stage, may limit (through regulation and control) the range and intensity of the emotions singers may experience during their vocal production; yet the singer's momentary, physiologically based, emotional state may be a powerful factor in determining various vocal features.

Second, an important source of emotion effects on the singing voice is the artist's interpretation of the mood or emotion expressed in a song or of the emotionality characterizing a role in a libretto. One of the most essential features for a successful singing performance is a credible portrayal of the emotional stance required by the respective piece. Although technical prowess and beauty of the voice are obviously decisive factors in a singer's career, they are rarely sufficient to move the audience to tears—or to standing ovations. Many singers have commented on the fact that a particularly successful performance is often brought about by strong emotional involvement, enabling them to slip into the role of the protagonist and to forcefully portray the appropriate emotional state (there is some dispute on whether the emotional involvement has to be genuine or can be simulated).

Third, emotional expression in singing is largely determined by the musical score as written by the composer. Although actors' performances in the speech theatre are constrained by the playwright's text (often highly emotional in itself), their vocal productions with respect to voice quality and prosodic features are largely chosen by themselves (or the director) to portray the affective state of the character. A singer, however, particularly in the lyric arts, is much more constrained by the composer's prescription of the large majority of the acoustic effects to be produced by the voice, leaving fewer degrees of freedom for the artist's personal emotional interpretation of the role.

The musicological literature abounds with accounts of how the composer manages to convey emotional meaning through a multitude of musical devices including melody, temporal patterning, and rhythm, as well as instrumentation (see refs. 63 and 66 for comprehensive historical surveys). To cite just one of a number of early accounts: In 1885 Hausegger published a manuscript entitled "Die Musik als Ausdruck" (Music as expression), much attacked by Hanslick and his school, in which he attempted to use the most recent psychobiological insights on emotional expression (including Darwin's book on emotional expression and the newest work in physiology) to explain the musical expression of emotion (with particular emphasis on the Wagnerian *Gesamtkunstwerk*).

Hausegger starts from the assumption, widely held at that time, of music having evolved from nonlinguistic affect sounds (see previous discussion herein). Using Darwin's functional explanation of emotional expression, he develops an argument about the immediate and powerful effect of emotional music, particularly due to the great importance of emotional communication for humans as social animals. Vocal music as a language of emotion is supposed to have immediate symbolic significance for affective content and can thus evoke affective feeling.

It is particularly interesting to look at the way in which Hausegger analyzes Donna Anna's aria *Or sai qui l'onore* from Mozart's *Don Giovanni*, bringing together elements from Darwin's explanation of the expression of anger and physiological evidence on the relationship between respiration and cardiac function. Based on Mozart's score, Hausegger diagnoses Donna Anna's state as one of sustained wrath rather than sudden, violent rage—citing the trembling and the sudden, angular movements that one would expect for the former state on the basis of Darwin's description, and showing the equivalence in Mozart's music. He even computed the respiration pattern required by the timing of the notes in the aria and claims that the state of arousal of the interpreter of Donna Anna must rise to a pulse of 120 beats/s as compared with a normal 72 (cited after ref. 63, p. 61).

To summarize, then, the study of emotional expression in singing needs to take into account the composer's emotional script, the singer's artistic interpretation and projection of a character's personality and affective state, and, finally, the singer's own physiologically based emotional state at the time of the performance.

As in speech research, three separate issues need to be addressed: (a) *inference*, i.e., the ability of the listener to *perceive* emotional quality in a lyrical piece or in the singer's performance (including the question concerning the ability of the performance to *induce* a specific emotional state in the listener); (b) the nature of the *acoustic cues* used in this inference process; and (c) the underlying *externalization* or *encoding* of the emotion, i.e., the actual covariation of a composer's emotional script, the director's or the performer's interpretation, and the singer's personal affect state and the specific vocal

production with its resulting acoustical waveform. Although empirical studies on either issue have been few and far between, there are a number of studies that can be cited for each of the three issues mentioned:

Inferring emotional meaning from the singing voice

Contrary to research on listeners' decoding of vocal expression of emotion in speech, it is less obvious to examine whether listeners can *correctly* infer an emotional state from vocal music. Arias and recitativos, Lieder, chansons, rock music lyrics, and even folk songs are always forms of art, removed from the reality of everyday emotions. Even if a particular, well-defined and labeled emotional expression were intended by the composer or by the performer's interpretation, the audience (or the researcher) would rarely have access to that intention. It is to be expected, however, that in most cases the artistic intentions are much more complex, defying a simple labeling in terms of fundamental emotion categories. Most likely, the desired expression will consist of subtle, highly context-bound mixtures or blends of different emotion categories.

However, it might be possible to design research procedures with singers that mirror the designs used with actors, i.e. ask professional singers to portray different types of "basic" or "fundamental" emotions by singing pure sounds or nonsense syllables (69). Clearly, the choice of particular melodic and dynamic features, independent of variations in voice quality, could be seen as corresponding to relatively arbitrary choices (which is quite similar in speech research, given that actors are also free to choose intonation and other dynamic speech features in their emotion portrayals). These stimuli could then be used in judgment studies to determine to what extent judges are able to infer the encoding intention of the singer.

To study the more subtle variations in emotional meaning in the context of a particular dramatic action, one might choose different pieces of vocal music with a large variety of affective content and use judgment studies to see to what extent judges are able to infer the expressive intention of the composer. Because the criterion is not very obvious here, one would need to use *agreement among listener judges* rather than accuracy as a measure of the nature of emotional inference.

With respect to the performer's interpretation, one could think of an experiment in which several performers would be asked to sing the same piece with different emotional interpretations of the affect state of the protagonist (e.g., the different types of emotional portrayals that could be given to the aforementioned Donna Elvira's aria—outrage, lust for revenge, righteous anger, fury mixed with desperation, etc.). Again, judgment studies could be used to determine to what extent listeners are able to correctly distinguish among these different projections on the basis of the acoustic cues alone.

It would be more difficult to settle the question whether listeners can distinguish the singer's personal affect state from acoustic cues. As in the speech area, it seems impossible, for practical and ethical reasons, to induce strong affect states in singers experimentally to obtain the necessary material for judgment studies. One possible avenue might be to record live performances of singers in a repertory company, asking them to keep a diary of their affective states for the same performance on different nights. Such material would be invaluable for getting at the listener's ability to detect very subtle variations in the singers' mood states independent of dramatic interpretation. It might be questionable, however, whether personal mood state and dramatic interpretation can be separated out very cleanly. There is a high probability that the performer's mood might interact with the dramatic interpretation chosen for a particular performance.

It is interesting to note that many more studies than in the speech area have been devoted to examining the *induction* of emotional states via music (70–74). This has become a particularly important issue in the psychology of emotion, where researchers are always in search of ethical methods to induce emotional states in the laboratory (75). Another important application is music therapy (76). An important future step in an empirical research program designed to understand how music *produces* affect would be to go beyond verbal assessment of emotional impact by directly measuring affective reactions of the listener—using micromeasurement of facial expression or psychophysiological recording.

Acoustic cues used in the inference process

Although the study of the *accuracy* of the emotion inference makes little sense in the case of vocal music, it is possible, and most interesting, to isolate the factors that determine the perception of emotional quality. What are the characteristics of the vocal performance, more specifically the acoustic

parameters in the respective voice signal, that induce the listener to attribute a particular emotion to the character impersonated by the singer? A number of empirical studies have been devoted to this issue, although only a few have dealt directly with the singing voice (64,77–84).

One possibility is to use a correlational method, as used in speech research, to establish the link among different acoustic configurations as found in emotionally expressive utterances via objective analysis, and correlate these with listener judgments. A viable approach, although obviously limited in scope, might be to take different recorded interpretations of famous scenes, preferably highly emotional, to study the effect of different voice qualities and dramatic interpretations used by various performers to portray emotional states and to analyze audience reactions. Siegwart and Scherer (87) have conducted digital acoustic analyses of two excerpts from the cadenza in the "mad scene" in Donizetti's *Lucia di Lammermoor* (using recorded performances by Toti dal Monte, Maria Callas, Renata Scotto, Joan Sutherland, and Edita Gruberova). They are able to show that the acoustic measures allow the prediction of a high percentage of the variability in listener judgments of preference and emotional meaning, providing some concrete hypotheses as to how certain acoustic features might affect the perception of emotional meaning in music.

Given that music is an art form that allows for virtually infinite variations, it is most productive to systematically vary different acoustic features and determine their effect on listener evaluation. With respect to instrumental music, in the study by Scherer and Oshinsky (40) cited previously herein, the MOOG synthesizer was used to vary the acoustic features of a small speech-like melody and of a Beethoven melody systematically. Very similar patterns of emotional judgment were found for the two types of melodies (Table 1). For the Beethoven melody, major and minor mode were also manipulated. As predicted on the basis of many such claims in the literature, major mode was shown to be seen as significantly more indicative of pleasantness and happiness than the minor mode, which favors impressions of disgust and anger.

The cleanest form of systematic variation would obviously be achieved via synthesis or resynthesis techniques, as in speech research. Recent advances in the synthesis of the singing voice (85,86) might allow the conducting of studies that, parallel to the speech area, systematically manipulate various expressive aspects of the singing voice and determine their role on audience appreciation of affective quality. Similar techniques might become available for resynthesis (copy synthesis) of natural voices. Similar to the study by Bergmann et al. (41) mentioned previously, one might use this technology to determine which acoustic features, in a Callas aria, for example, might need to be manipulated (and in which manner) to rob the interpretation of its emotional impact.

It seems, then, that the issue concerning the emotional impact of music might well be amenable to empirical study by linking listener judgments to objective data concerning the perceptual input. This is not to deny that cognitive, and specifically attributional, factors need to be taken into account (as Bever suggests: see ref. 60). However, it is unlikely that music, and opera in particular, will only amplify whatever affect is felt before listening to a piece.

Acoustic correlates of emotional expression in singing

As in the domain of emotion expression in speech, it is of major interest to describe the configurations of acoustic cues characteristic for the expression of certain affect states in singing. One of the few studies in this area has been conducted by Kotlyar and Morozov (69). These researchers asked 11 professional singers to sing different pieces in such a way as to portray happiness, sorrow, fear, and anger. They then carried out acoustic analyses of these stimuli. The results partly show patterns rather similar to what is found in speech research: fast tempo for fear, slow for sorrow, high vocal energy for anger. Other features seem quite different, such as longer pauses between syllables for fear compared with other emotions. These authors also used a synthesis procedure to evaluate the role of the acoustic features found to differentiate the emotion encoding in *decoding* or inference. They used electronically generated signals to manipulate the respective features and found that judges were able to identify the underlying affect rather well, except in the case of joy (a detailed description of the results is also provided in ref. 83, pp. 152–153).

Unless such portrayal or encoding methods are used, studies of the acoustic correlates of emotional state or projection intention are difficult to do because of the lack of a clear criterion for the nature of

the underlying state or expression intention (given the near impossibility of experimental induction); yet some of the aforementioned sampling methods might be used to good effect to generate material that could be used to study the acoustic correlates of emotion in singing intensively. Further work in this rather neglected area of research is likely to yield very promising results.

CONCLUSIONS

This article has attempted to give an overview of the theoretical foundations and the major research trends in the vocal expression of emotion. It is obvious that the voice, because of its physiological basis and its communicative functions, is a primary instrument for emotional expression. In consequence, from antiquity, there is no lack of literary commentary on the importance of emotional expression in speech and singing. Given the manifold difficulties of studying this phenomenon empirically, or even experimentally, it is not surprising that the research evidence so far is sparse. However, the data that are available do show reasonable convergence and allow the formulation of concrete hypotheses for theory-guided research. The electroacoustic instrumentarium for storing voice samples and the digital methods for acoustic analysis are becoming extremely sophisticated and affordable—particularly with the marketing of personal computers with multimedia capacity. There are also signs of increasing awareness of the need for interdisciplinary work in this area—bringing together speech scientists, psychologists, electrical engineers, musicologists, and performing artists, in an attempt to better understand the many facets of the intriguing issue of emotion expression in the voice (88,89). In this spirit, it might well be profitable to link the study of emotion expression in the speaking and the singing voice much more closely than has been the case to date.

Acknowledgment: I acknowledge contributions by Ursula Scherer and Johan Sundberg.

REFERENCES

1. Frijda NH. *The emotions*. Cambridge and New York: Cambridge University Press, 1986.
2. Scherer KR, Ekman P, eds. *Approaches to emotion*. Hillsdale, New Jersey: Erlbaum, 1984.
3. Scherer KR. Vocal affect expression: a review and a model for future research. *Psychol Bull* 1986;99:143–65.
4. Darwin C. *The expression of the emotions in man and animals*. London: Murray, 1872. (Reprinted Chicago: University of Chicago Press, 1965).
5. Redican WK. Facial displays of emotion by monkeys and apes: an evolutionary perspective on human facial displays. In: Ekman P, ed. *Emotion in the human face*. Cambridge and New York: Cambridge Univer. it'' Press, 1982.
6. Ekman P. Darwin and cross-cultural studies of facial expression. In: Ekman P, ed. *Darwin and facial expression*. New York: Academic Press, 1973:1–83.
7. Izard CE. *The face of emotion*. New York: Appleton-Century-Crofts, 1971.
8. Ekman P, Friesen WV, Ellsworth PC. *Emotion in the human face: guidelines for research and an integration of findings*. New York: Pergamon Press, 1972.
9. Marler P. Animal communication: affect or cognition? In: Scherer KR, Ekman P, eds. *Approaches to emotion*. Hillsdale, New Jersey: Erlbaum, 1984:345–68.
10. Scherer KR. On the symbolic functions of vocal affect expression. *J Lang Soc Psychol* 1988;7:79–100.
11. Jürgens U. Vocalization as an emotional indicator: a neuroethological study in the squirrel monkey. *Behaviour* 1979; 69:88–117.
12. Marler P, Tenaza R. Signaling behavior of apes with special reference to vocalization. In: Sebeok TA, ed. *How animals communicate*. Bloomington, Indiana: Indiana University Press, 1977:965–1033.
13. Morton ES. On the occurrence and significance of motivational-structural rules in some bird and mammal sounds. *Am Naturalist* 1977;111:855–69.
14. Scherer KR. Vocal affect signalling: a comparative approach. In: Rosenblatt J, Beer C, Busnel M, Slater PJB, eds. *Advances in the study of behavior*. New York: Academic Press, 1985:189–244.
15. Tembrock G. Die Erforschung des tierlichen Stimmausdrucks (Bioakustik). In: Trojan F, ed. *Biophonetik*. Mannheim: Bibliographisches Institut, 1975:51–67.
16. Kleinpaul R. *Sprache ohne Worte. Idee einer allgemeinen Wissenschaft der Sprache*. The Hague: Mouton (Orig. ed. Leipzig, 1888), 1972.
17. Wundt W. *Völkerpsychologie. Eine Untersuchung der Entwicklungsgesetze von Sprache, Mythos und Sitte*. Leipzig: Kröner, 1900.
18. Kainz F. *Psychologie der Sprache. I. Band. Grundlagen der allgemeinen Sprachpsychologie*, 3rd ed. Stuttgart: Enke, 1962.
19. James D. The syntax and semantics of some English interjections. *Papers Linguistics* 1974;1:3.
20. Scherer KR. Affektlaute und vokale Embleme. In: Posner R, Reinecke HP, eds. *Zeichenprozesse—Semiotische Forschung in den Einzelwissenschaften*. Wiesbaden: Athenaion, 1977:199–214.
21. Scherer KR. The functions of nonverbal signs in conversation. In: St. Clair R, Giles H, eds. *The social and psychological contexts of language*. Hillsdale, New Jersey: Erlbaum, 1980:225–44.
22. Goffman E. Response cries. In: Cranach MV, Foppa K, Lepenies W, Ploog D, eds. *Human ethology: claims and limits of a new discipline*. Cambridge: Cambridge University Press, 1979:203–40.
23. Scherer KR. Affect bursts. In: van Goozen SHM, van de Poll NE, Sergeant JA, eds. *Emotions: essays in emotion theory*. Hillsdale, New Jersey: Erlbaum 1994:161–196.
24. Scherer KR. Emotion expression in speech and music: In: Sundberg J, Nord L, Carlson L, eds. *Music, language, speech, and brain*. London: Macmillan, 1991:146–56.
25. Andrews RJ. The information potentially available in mammal displays. In: Hinde RA, eds. *Nonverbal communication*. Cambridge: Cambridge University Press, 1972:129–206.

26. Leyhausen P. Biologie von Ausdruck und Eindruck (Teil I). *Psychol Forsch* 1967;31:113–76.
27. Copeland A. *How to listen to music*. New York: Bantam (orig. ed. 1937), 1959.
28. Meyer LB. *Emotion and meaning in music*. Chicago: University of Chicago Press, 1956.
29. Seashore C. *The psychology of music*. New York: Dover (orig. ed. New York: McGraw Hill, 1938), 1967.
30. Scherer KR. Speech and emotional states. In: Darby JK, ed. *Speech evaluation in psychiatry*. New York: Grune & Stratton, 1981:189–220.
31. van Bezooijen R. *The characteristics and recognizability of vocal expression of emotions*. Dordrecht, The Netherlands: Foris, 1984.
32. Scherer KR, Banse R, Wallbott HG, Goldbeck T. Vocal cues in emotion encoding and decoding. *Motiv Emotion* 1991;15:123–48.
33. Frick RW. Communicating emotion: the role of prosodic features. *Psychol Bull* 1985;97:412–29.
34. Wallbott HG, Scherer KR. Cues and channels in emotion recognition. *J Pers Soc Psychol* 1986;51:690–9.
35. Kramer E. Elimination of verbal cues in judgments of emotion from voice. *J Abnorm Soc Psychol* 1964;68:390–6.
36. Starkweather JA. Content-free speech as a source of information about the speaker. *J Pers Soc Psychol* 1956;35:345–50.
37. Scherer KR, Feldstein S, Bond RN, Rosenthal R. Vocal cues to deception: a comparative channel approach. *J Psycholinguist Res* 1985;14:409–25.
38. Scherer KR, Ladd DR, Silverman KEA. Vocal cues to speaker affect: testing two models. *J Acoust Soc Am* 1984; 76:1346–56.
39. Lieberman P, Michaels SB. Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech. *J Acoust Soc Am* 1962;34:922–7.
40. Scherer KR, Oshinsky J. Cue utilization in emotion attribution from auditory stimuli. *Motiv Emotion* 1977;1:331–46.
41. Bergmann G. Goldbeck T, Scherer KR. Emotionale Eindruckswirkung von prosodischen Sprechmerkmalen. *Z Exp Angew Psychol* 1988;35:167–200.
42. Ladd DR, Silverman KEA, Tolkmitt F, Bergmann G, Scherer KR. Evidence for the independent function of intonation contour type, voice quality, and F0 range in signaling speaker affect. *J Acoust Soc Am* 1985;78:435–44.
43. Lipps T. *Psychologische Untersuchungen*. Leipzig: Engelmann, 1907.
44. Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. The perception of the speech code. *Psychol Rev* 1967;74:431–61.
45. Scherer KR. Interpersonal expectations, social influence, and emotion transfer. In: Blanck PD, ed. *Interpersonal expectations: theory, research, and application*. Cambridge and New York: Cambridge University Press, 1993:316–36.
46. Scherer KR. Neuroscience projections to current debates in emotion psychology. *Cogn Emotion* 1993;7:1–41.
47. Scherer KR. Non-linguistic indicators of emotion and psychopathology. In: Izard CE, ed. *Emotions in personality and psychopathology*. New York: Plenum Press, 1979:495–529.
48. Scherer KR. Vocal assessment of affective disorders. In: Maser JD, ed. *Depression and expressive behavior*. Hillsdale, New Jersey: Erlbaum, 1987:57–82.
49. Robinson BW. Anatomical and physiological contrasts between humans and other primate vocalizations. In: Dolhinow P, Washburn SL, eds. *Perspectives on human evolution*. New York: Holt, Rinehart, & Winston, 1972:438–43.
50. Scherer KR. Vocal correlates of emotion. In: Manstead A, Wagner H, eds. *Handbook of psychophysiology: emotion and social behavior*. London: Wiley, 1989:165–97.
51. Scherer KR, Kappas A. Primate vocal expression of affective states. In: Todt D, Goedeking P, Newman E, eds. *Primate vocal communication*. Heidelberg: Springer, 1988:171–94.
52. Williams CE, Stevens KN. On determining the emotional state of pilots during flight: an exploratory study. *Aerospace Med* 1969;40:1369–72.
53. Williams CE, Stevens KN. Emotions and speech: some acoustical correlates. *J Acoust Soc Am* 1972;52:1238–50.
54. Hollien H, Geison L, Hicks JWJ. Vocal stress evaluators and lie detection. *J Forensic Sci* 1987;32:405–18.
55. Scherer KR. Voice, stress, and emotion. In: Appley MH, Trumbull R, eds. *Dynamics of stress*. New York: Plenum, 1986:159–81.
56. Tolkmitt FJ, Scherer KR. Effects of experimentally induced stress on vocal parameters. *J Exp Psychol Hum Percept Perform* 1986;12:302–13.
57. Pittam J, Scherer KR. Vocal expression and communication of emotion. In: Lewis M, Haviland JM, eds. *Handbook of emotions*. New York: Guilford Press, 1993:185–98.
58. Cosmides L. Invariances in the acoustic expression of emotion during speech. *J Exp Psychol Hum Percept Perform* 1983;9:864–81.
59. Tischer B. *Die vokale Kommunikation von Gefühlen*. [Habilschrift], Universität München, 1991.
60. Bever TG. A cognitive theory of emotion and aesthetics in music. *Psychomusicology* 1988;7:165–75.
61. Clynes M. *Sentics: the touch of emotions*. New York: Anchor/Doubleday, 1977.
62. Cooke D. *The language of music*. London: Oxford University Press, 1959.
63. De la Motte-Haber H. *Handbuch der Musikpsychologie*. Laaber: Laaber-Verlag, 1985.
64. Fonagy I. Emotions, voice and music. In: Sundberg J, ed. *Research aspects on singing*. Stockholm and Paris: Royal Swedish Academy of Music and Payot, 1981:51–79.
65. Fonagy I. *La vive voix*. Paris: Payot, 1983.
66. Kivy P. *Osmin's rage: philosophical reflections on opera, drama, and text*. Princeton: Princeton University Press, 1988.
67. Lundin RW. *An objective psychology of music*. New York: Ronald Press (orig. ed. 1953), 1967.
68. Helmholtz HLF. *On the sensations of tone as a psychological basis for the theory of music*. New York: Dover (Orig. German ed. Braunschweig: Vieweg, 1863), 1954.
69. Kotlyar GM, Morozov VP. Acoustical correlates of the emotional content of vocalized speech. *Sov Phys Acoust* 1976; 22:208–11.
70. Bochkaryov LL. Adequacy of perception of music depending on listener's psychic condition. *Novye Issledovaniya Psikhologii* 1981;2:61–4.
71. Collins SC. Subjective and autonomic responses to Western classical music [Ph.D. Thesis]. Manchester: Department of Psychology, University of Manchester, 1989.
72. Goldstein A. Thrills in response to music and other stimuli. *Physiol Psychol* 1980;8:126–9.
73. Harrer G, Harrer H. Musik, Emotion und Vegetativum. *Wien Med Wochenschr* 1968;45/46:966–71.
74. Payne E. Musical taste and personality. *Br J Psychol* 1967; 58:133–8.
75. Gerrards-Hesse A, Spies K, Hesse FW. Experimental inductions of emotional states and their effectiveness: a review. *Br J Psychol* (in press).
76. Davis WB. *An introduction to music therapy: theory and practice*. Dubuque, Iowa: Brown, 1992.
77. Brown RW. Music and language. In: *Documentary Report of the Ann Arbor Symposium*, Reston, Virginia: 1981:233–65.
78. Gundlach RH. Factors determining the characterization of musical phrases. *Am J Psychol* 1935;47:624–43.

79. Hevner K. Expression in music: a discussion of experimental studies and theories. *Psychol Rev* 1935;42:186–204.
80. Hevner H. Experimental studies of the elements of expression in music. *Am J Psychol* 1936;48:248–68.
81. Osborne JW, Kennedy L. An empirical validation of Schopenhauer's theory of music through analysis of listeners' experiences of Mahler's Ninth Symphony. *J Phenomenol Psychol* 1985;16:13–38.
82. Petroushin VI. Simulation of emotions by means of music. *Voprosy Psikhologii* 1988;5:141–4.
83. Sundberg J. *The science of the singing voice*. DeKalb, Illinois: Northern Illinois University Press, 1987.
84. Wedin L. A multidimensional study of perceptual-emotional qualities in music. *Scand J Psychol* 1972;13:241–57.
85. Risset JC. Speech and music combined: an overview. In: Sundberg J, Nord L, Carlson R, eds. *Music, language, speech, and brain*. London: Macmillan, 1991:368–79.
86. Sundberg J. Synthesis of singing. *Swed J Musicol* 1978;60: 102–12.
87. Siegwart H, Scherer KR. Acoustic concomitants of emotional expression in operatic singing: The case of Lucia in *Ardi gli incensi*. *J Voice* 1995;9:000–000.
88. Sundberg J, Nord L, Carlson L, eds. *Music, language, speech, and brain*. London: Macmillan, 1991.
89. Dalmonte R, Baroni M, eds. *Secondo convegno europeo di analisi musicale*. Trento, Italia: Università degli Studi di Trento, 1992.