Chapter 7

Vocal Correlates of Emotional Arousal and Affective Disturbance

Klaus R. Scherer
University of Geneva

ABSTRACT

Human vocalization simultaneously reflects digital features of the linguistic phonetic system and analog features of physiologically mediated emotional and motivational states. Consequently, acoustic parameters of voice and speech constitute reliable indicators of emotional arousal as well as of sociocommunicative behaviour. This chapter describes the physiological determinants of voice production and evaluates the acoustic parameters that yield useful indicators of emotional arousal states and of affective disturbance. A review of pertinent literature establishes the validity of vocal measurement as a diagnostic tool. In a final section of the chapter, theoretically based predictions of vocal correlates of different emotional states are proposed.

VOCAL PARAMETERS AS INDICATORS OF PSYCHOPHYSIOLOGICAL PROCESSES

Human speech is an extraordinary communication system in that it uses as its vehicle a phylogenetically much older vocal expression system, which functions in most mammals as well as in man as a physiologically mediated indicator of emotional arousal. Similar to the evolution of the brain, where 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 encoding information has been grafted upon, and makes use of, the analogue nonverbal system of affect vocalization.

It has long been accepted that, in most animals, vocalization primarily serves the function of signalling motivational and emotional states of the animal. This view has recently been attacked by Marler and his associates (Marler, 1984) who, on the basis of recent studies on alarm calls in monkeys (Seyfarth and Cheney, 1982), argue that animal calls may also have symbolic representational functions, for example in drawing attention to different types of predators via acoustically differentiated calls. However, it can be shown that even in these cases the calls are clearly linked to an evaluation of the antecedent situation which produces differentiated emotions (Scherer and Kappas, 1988) and that such calls are only produced in response to motivationally and emotionally relevant stimuli. The evolutionary basis of vocalization, then, seems to be very closely linked to motivational and emotional arousal. Consequently, vocalization can be considered to be one of the most direct correlates, and consequently one of the most powerful indicators, of motivational and emotional processes.

As we shall demonstrate below, voice production mechanisms are controlled by a large number of physiological processes, many of which are considered to be important psychophysiological variables (such as blood pressure, muscle tension, etc.). Given the nature of the production mechanism, we are able to infer variations in the underlying physiological processes from changes in the acoustic structure resulting from the respective vocalization patterns. To the extent that acoustically assessed vocalization reflects underlying physiological parameters, it is possible to use vocal parameters, which can be easily and unobtrusively registered and analyzed, in lieu of cumbersome and intrusive psychophysiological recording. Much of this chapter will be concerned with this approach, i.e. the attempt to use vocal analysis as a measure of motivational and emotional change.

However, an important caveat has to be introduced at this point. Contrary to physiological systems which mainly serve internal regulation functions for the organism, vocalization has developed as a social communicative signalling system. Consequently, the physiological factors determining the expression of the internal state of the organism [which is of fundamental importance for social interaction, see Scherer (1984, 1985)] have been supplemented by pressures exerted by the demands for certain patterns of impression in the receiver of the message. While a naive model of communication might hold that impression is totally determined by expression, it can be shown that in the course of the evolution of expressive communication systems, impression models have moulded the nature of expression [see Leyhausen (1967) for a thorough discussion of this point]. Scherer (Scherer, 1985; Scherer and Kappas, 1988) has introduced the distinction between push-effects and pull-effects to distinguish between the determining factors which operate on vocalization. Push-effects are produced by the physiological changes which accompany emotional arousal and which, 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 internal physiological processes in the organism. Their origin is found in factors external to the organism, such as ritualized or conventionalized acoustic signal patterns, which 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 receivers).

In most cases, the acoustic structure 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.

This dual determination of the acoustic structure of vocalization is the major difference between using vocal parameters of emotional arousal and using psychophysiological, particularly autonomic, indices. Whereas the latter are only marginally affected by the organism's regulation attempts, the former, given their important role in expression and communication, may, to a very large extent, be determined by factors that are independent of the underlying physiological state. One of the major tasks for researchers in the area of vocal expression of emotion, then, is to disentangle push-effects and pull-effects on the basis of the observable acoustic pattern. On the one hand, this is a very difficult chore and a problem far from being resolved—the use of vocal parameters to infer underlying emotional arousal is not a straightforward measurement operation. On the other hand, the assessment of pull-effects on emotional vocalization is of major importance in its own right. Since emotion is generally a highly social phenomenon, and since it is often difficult to distinguish emotional reaction from attempts to cope with this reaction and from consequent social regulation (see Scherer, Wallbott, and Summerfield, 1986), vocalization changes induced by such social regulation attempts may provide important information on the nature of the emotional episode as a whole.

In this chapter we shall deal mainly with push-effects on vocalization, in keeping with the general psychophysiological approach in this handbook. Using both theoretical arguments and empirical evidence, we shall attempt to show which acoustic parameters can be used as indicators of various types of emotional arousal and of affective disturbance. First, we shall briefly review the major production mechanisms of vocalization and the physiological and neurological systems that control these mechanisms. Figure 7.1 gives a general outline of the structures involved in voice production.

In the process of respiration, the lungs produce the energy required for vocalization in building up subglottal pressure in the trachea below the closed glottal folds. This subglottal air pressure, together with motor commands to the laryngeal musculature, brings about phonation—the regular vibration of the

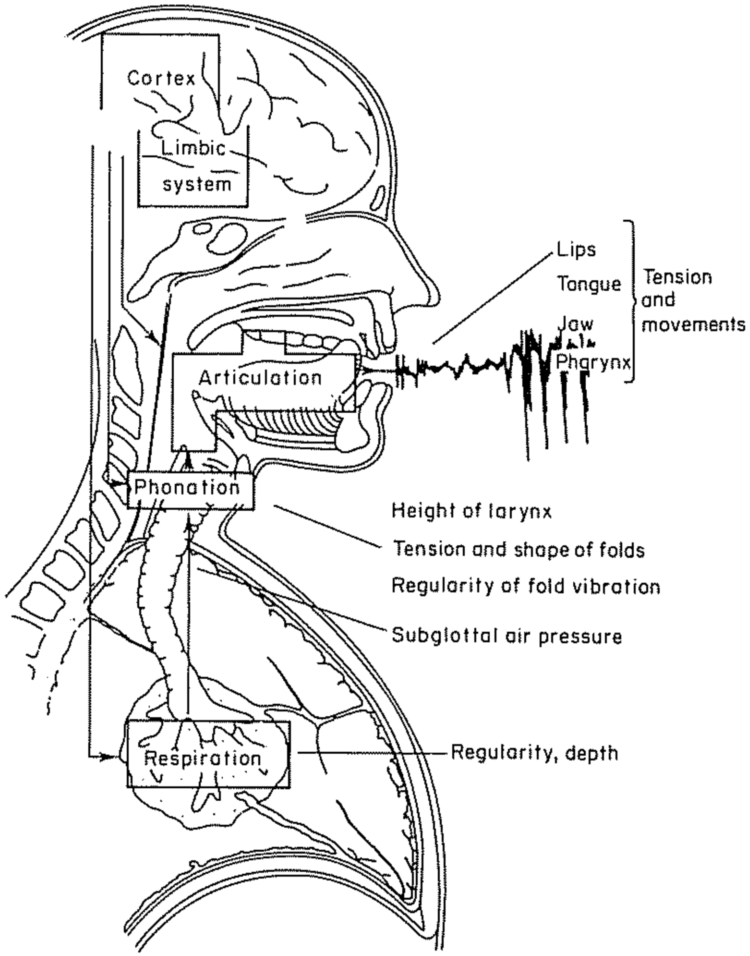


Figure 7.1. Overview of the voice production system and its major determinants.

vocal folds which releases air pulses into the supraglottal vocal tract. This rapid series of pulses, which constitutes a primitive, triangular-shaped wave-form, is then modified and filtered by the shape and the length of the supralaryngeal tract. Changes in the shape and length of the tract are mostly produced by motor commands to movable structures such as the tongue, the lips, and the jaw. These variations in the shape of the vocal tract produce articulation, which is responsible for the production of the basic linguistic units, the phonemes, and also for emotionally relevant variations of the filter or transfer function. For detailed surveys of the voice production process see Daniloff, Schuckers, and Feth (1980), Fry (1979), Scherer (1982), and Zemlin (1968).

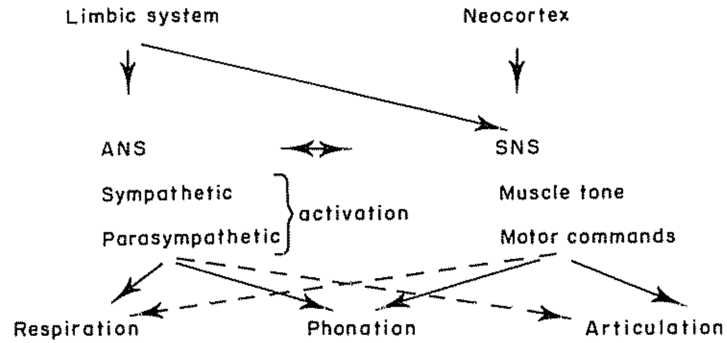


Figure 7.2. Effects of neurophysiological structures on voice production mechanisms

Figure 7.2 shows, in a highly simplified manner, some of the main effects of the major neurophysiological structures on the voice production mechanisms shown in Figure 7.1. Speech production in the linguistic sense is obviously mostly controlled by the neocortex, a process that works mainly via specific motor commands which produce appropriate phonatory and articulatory movements for the desired sequence of speech sounds. However, some of the pull-effects discussed above, such as attempts at self-presentation using certain conventionalized patterns of vocalization, may also be produced in this manner. 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 that are primarily controlled by the limbic system (Jürgens, 1979; Robinson, 1972) are much more diffuse and complicated. They are mostly produced via tonic activation of the somatic nervous system (in particular the striated musculature), and sympathetic, as well as parasympathetic, activation of the autonomic nervous system. Given the predominance of the striated musculature in producing vocalization, many of these emotional effects are also likely to be mediated via the somatic nervous system. However, direct sympathetic or parasympathetic effects, such as respiration changes and the secretion of mucus, will also affect the nature of the vocal output.

Vocalization is a very sensitive output system—even slight changes in physiological regulation will produce very noticeable changes in the acoustic pattern. Even if a speaker attempts to reproduce a particular utterance in exactly the same way, in exactly the same state of arousal, some changes are likely to occur. The enormous sensitivity of the acoustic output to minor changes in voice production settings is both advantageous, since it provides a sensitive and rapidly responding system for monitoring emotional arousal, and disadvantageous, given the high degree of noise that is introduced by minor changes in respiration, phonation, and articulation, in addition to strong individual differences. A somewhat different, but equally serious, problem is the fact that a number of acoustic patterns can be produced by very different phonatory and articulatory settings. Consequently, it is very difficult to infer the exact nature of the underlying production process from a given acoustic pattern.

VOCALIZATION PARAMETERS IN MEASUREMENT OF EMOTION

The type of parameter to be used in assessing emotion from vocalization depends on the level of measurement chosen (see also Scherer, 1982). At the most basic level, one can directly measure some of the physiological determinants of the voice production process. An example would be EMG measurement of laryngeal muscles involved in phonation, or the amount of mucus secreted. In other words, one would use similar physiological variables to those used in psychophysiological research generally—with the constraint that the variables involved also play a role in the physiological process underlying vocalization. Since most of the neurophysiological structures involved in vocalization cannot be accessed readily, this approach is unlikely to be useful in most cases of emotion research. However, further research on the exact nature of the physiological processes involved in voice production would clearly advance our understanding of the details of the mechanism and would provide a better basis for inferring how emotion-related physiological change influences vocalization.

At the next level of analysis, one can assess a variety of variables related directly to the motor effects of the underlying physiological processes. For example, using appropriate apparatus, one can measure respiration rate, tongue movements, vocal fold vibration, etc. On this level, the actual movement patterns involved in voice production and filtering are investigated. While many of these parameters are highly valid indicators of emotional arousal, and can be measured very objectively and reliably, the process of measurement generally requires a rather sophisticated experimental set-up, in terms of apparatus and speaker-cooperation. Given the inconvenience involved in most of these measurement procedures, it is likely that the negative affect generated by the procedures will overwhelm any other emotion experienced by the speaker.

While the first two levels of measurement are based on direct observation of the voice production mechanisms, the remaining three levels, and their respective parameters, are all based on the acoustic waveform that is radiated from the mouth as a result of the voice production process. The speech sound-wave can be used in very different ways for the extraction of parameters.

At a level of analysis that is still close to voice production, one can try using experts to infer the nature of the voice production movements and settings that have yielded the acoustic output under study. An example of this approach is the attempt to establish a coding system for different voice qualities proposed by Laver (1980).

A different type of analysis is concerned with the detailed description of the acoustic waveform resulting from the vocalization. In general, electro-acoustic equipment or digital computers are used for the extraction of a large number of acoustic parameters that can serve to describe the sound-wave. We shall deal with these parameters in detail below. Another approach is to use judges in order to obtain ratings of acoustic features, generally using natural language categories for certain types of acoustic parameters. This approach is hampered by the fact that there are no precise natural language concepts for particular acoustic parameters. In addition, the nature of the human auditory sense organs, and in particular the tendency to integrate independent acoustic dimensions (e.g. the auditory impression of voice pitch is affected not only by the fundamental frequency of the vocal fold vibrations but also by energy distribution in the spectrum), renders this procedure problematic if physical acoustic description is desired.

However, if the communicative aspect of the vocalization is under study, the use of judges in order to assess the perceptually and communicatively relevant dimensions of the acoustic waveform may well be considered. This is particularly true for the final level of analysis, where the acoustic output of the speaker's vocalization is used as the basis for auditory judgements of the valence of the acoustic pattern, given a certain number of criteria of the listener, such as aesthetic preference or motivational relevance. At this level of analysis, it is not the nature of the acoustic sound-wave or the types of effects that produced it that are of interest, but rather the impression created by the particular vocalization on the receiver.

Which of these parameters should be used in a particular study depends on the questions asked and the constraints posed by the operationalization involved. However, the level of objective acoustical analysis of vocalization is of central importance for a number of empirical approaches. The acoustic waveform is the signal which is actually transferred in the communication process. It reflects both the production mechanism and its physiological determinants, and it is the basis for the various types of listener inferences (both expert and naive). Consequently, for many studies of the effects of emotion on vocalization, the acoustic analysis of speech patterns will be of major importance. Furthermore, the acoustic parameters can be obtained objectively, economically, and unobtrusively, from tape-recordings of the speaker's utterances. Given the central importance of acoustic measurement of the emotional vocalization, we shall devote most of our attention in this chapter to this type of analysis and the parameters that seem most useful in this connection. The emphasis on acoustic parameters also seems justified by the fact that whereas, until recently, the ability to extract the relevant parameters has been restricted to laboratories with access to sophisticated computer equipment, the appropriate techniques are now becoming available for use with standard microcomputers.

We shall now turn to a detailed description of the major acoustic parameters and a brief discussion of the appropriate measurement procedures and of the

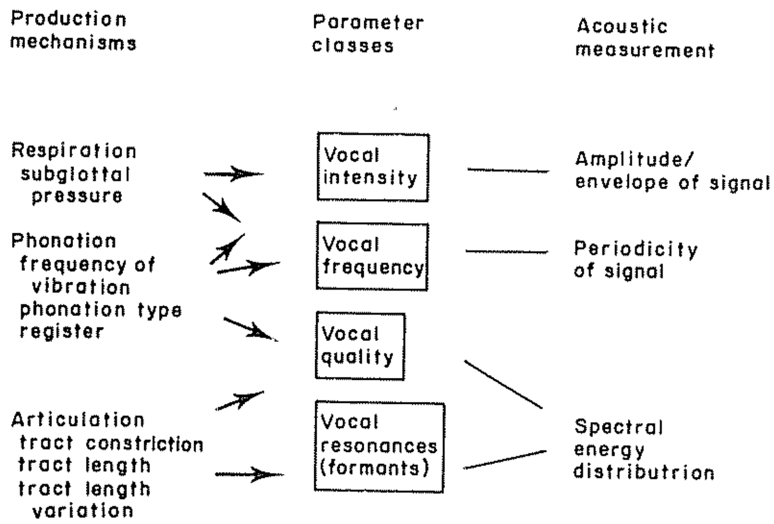


Figure 7.3. Effects of voice production mechanisms on acoustic parameters.

factors that are primarily responsible for variation in the respective parameter. Figure 7.3 shows the four major classes of vocal parameters, together with the production mechanisms that are responsible for changes in these parameters and the acoustic variables used in their measurement. We shall discuss each of these classes in turn. Figure 7.4 is provided to help to visualize the parameters discussed in terms of their measurement and graphic display based on acoustic analysis of the vocal signal [readers who are unfamiliar with acoustic phonetic material may find the following references useful: Daniloff *et al.* (1980), Fry (1979), Hollien (1981), Lieberman, (1977), Scherer (1982), and Zemlin (1968)].

Vocal intensity, subjectively heard as loudness of the voice, is jointly determined by respiratory and phonatory action. Higher intensity is generally due to increased subglottal pressure and greater laryngeal tension (see Daniloff *et al.*, 1980; Zemlin, 1968). The acoustic measurement of intensity is rather straightforward since it is directly related to the amplitude (or envelope) of the speech waveform. There are a large number of instruments available to measure amplitude of the acoustic waveform, either in voltage or in dB (see Hollien, 1981; Scherer, 1982). If recorded speech is being used, care has to be taken to allow for differential gain in recording and reproduction. Amplitude measurements can be reliably obtained for very short segments of the speech-wave, down to about 100 mseconds. The resulting amplitude values can be averaged over various portions of the speech utterance and a number of dispersion statistics, such as standard deviation and range, can be computed. The steepness of intensity rises (attack) and falls (decay) for voiced speech segments (see Figure 7.4b) can be obtained from the time signal or from the

envelope. Graphically, amplitude can be plotted over time to yield the envelope or amplitude contour of the signal (see Figure 7.4e).

Vocal frequency, subjectively heard as pitch of the voice, is primarily determined by the frequency of vibration of the vocal folds, a process which is determined jointly by differential innervation of the laryngeal musculature and the extent of subglottal pressure. Acoustically, vocal frequency is measured by the fundamental frequency of the speech waveform, i.e. the lowest periodic cycle component of the complex waveform (that is, number of periods per second, see Figure 7.4c), and the harmonics (multiples of fundamental frequency). Fundamental frequency of a speech signal can be measured with a variety of devices now available (see Hollien, 1981). Of particular importance is the use of digital fundamental frequency extraction, for which a variety of different algorithms have been proposed (Hess, 1983). While fundamental frequency (F_0) extraction is theoretically straightforward, there are numerous practical difficulties that require particular caution in the extraction and interpretation of fundamental frequency. While, in principle, individual periods can be determined, most hardware F_0 detection devices and most automatic F_0 extraction algorithms require a period of about 100 mseconds for reliable F_0 extraction. As for amplitude, F_0 values can be averaged over different portions of the speech utterance, and the usual dispersion parameters can be computed. A graphical plot of the sequence of F_0 values across an utterance yields an intonation contour (see Figure 7.4d), which together with the envelope of the signal is one of the most important parameters for prosodic analysis. Of particular interest for studies of the vocal expression of emotion is F_0 perturbation, or pitch jitter, a variable that refers to the degree of variability in the length of adjoining periods (see Lieberman, 1961; Sorensen and Horii, 1984).

Since fundamental frequency is directly proportional to the length of the vocal folds, males, females and children have rather different modal fundamental frequencies [with a modal value of 128 Hz for males and 260 Hz for females, see Daniloff *et al.* (1980), p. 203]. Because of these important sex differences, great care has to be taken in comparing the F_0 measurements of males and females.

Vocal quality, subjectively heard as timbre, is determined by phonation type [see Daniloff *et al.* (1980), Ch. 6] and phonation register [pulse, modal, or falsetto registers, see Laver (1980)]. In addition, the general tension and the specific configuration produced by the articulatory setting has an impact on vocal quality. The acoustic measurement of vocal quality is still hotly debated in the literature. In general terms, the most direct effect of vocal quality is to produce changes in the energy distribution and the power spectrum (see Figure 7.4e). For example, shrill voices have a much higher proportion of energy in the upper frequency range (i.e. above 500 Hz.) whereas resonant voices in general have a greater proportion of energy in the lower frequency range. In general, researchers are still somewhat uncertain about which exact variable to use to capture differential energy distribution in the spectrum. Among some of the basic variables proposed are frequency range, i.e. the difference between F_0 and the highest point in the

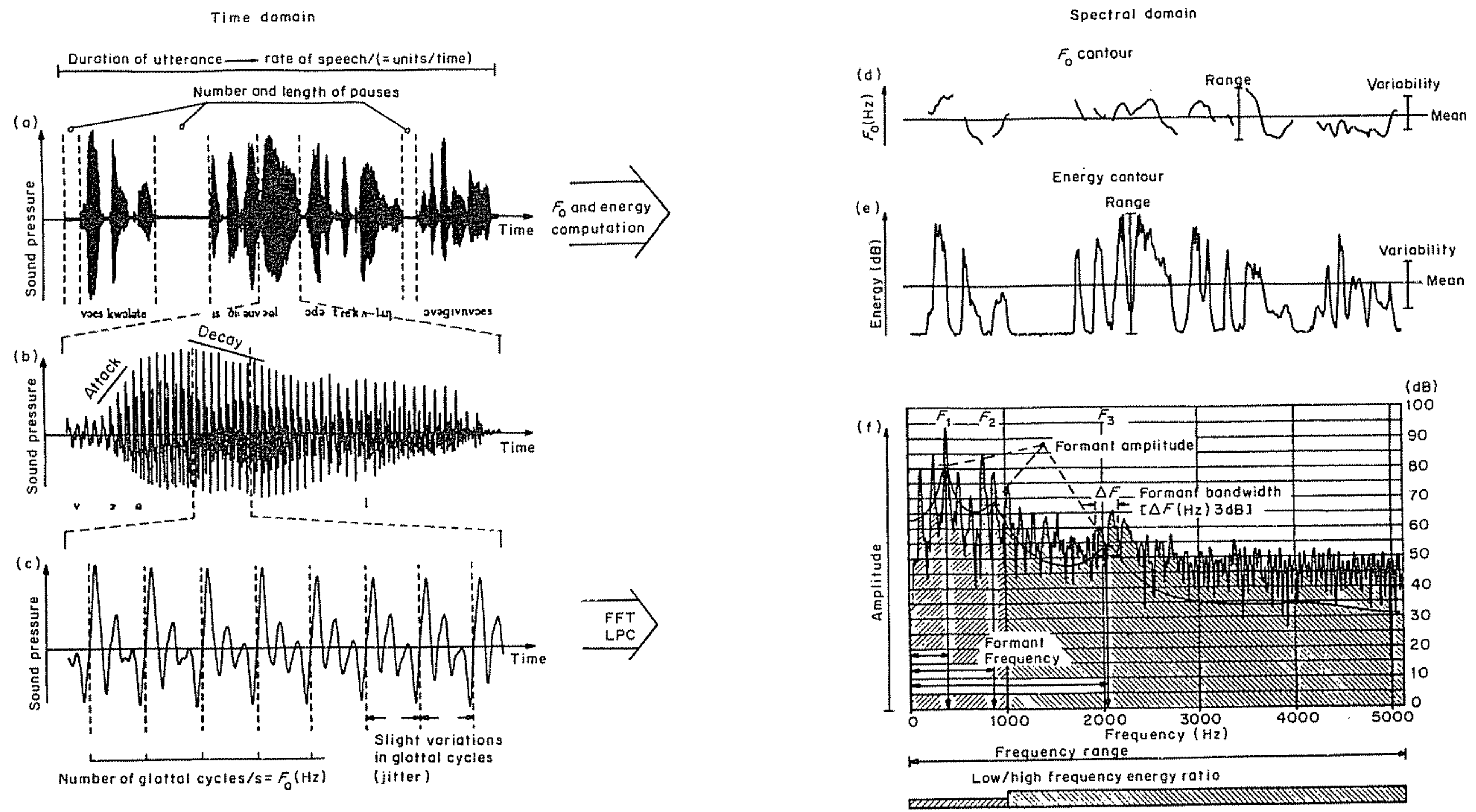
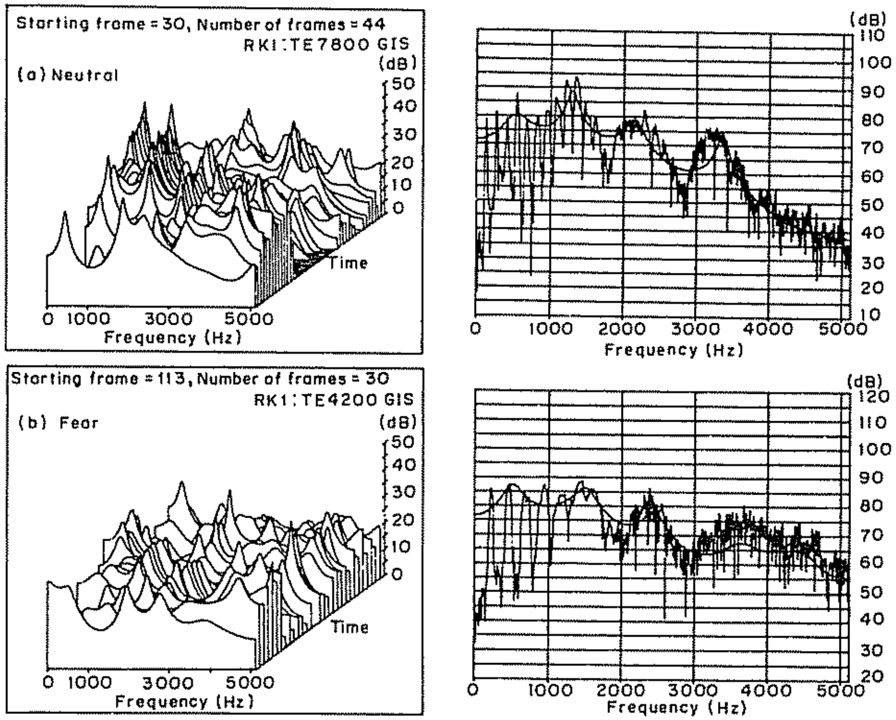


Figure 7.4. Major acoustic parameters and their measurement.

frequency spectrum where there is still measurable energy, and various proportions of low versus high energy concentration (i.e. energy below 500 Hz in relation to above 500 Hz, or below 1000 Hz in relation to above 1000 Hz; see Van Bezooijen (1984), for review).

Of further interest is the so-called *spectral noise*, the aperiodic energy components in the spectrum. While the definition and measurement of the specific variables still present problems, it is rather straightforward to obtain power spectra for relatively short periods of speech (down to about 20 mseconds). The power spectrum can be displayed graphically either for a particular window (i.e. a sampling period of the speech-wave) or a sequence of these individual power spectra can be displayed in a 3D-perspective (see Figure 7.5). Another possibility is to use differential degree of shading to indicate the relative energy of different



frequency components as is done in the spectrogram, the classic way of obtaining graphic displays of energy distribution in the spectrum (see the example in Figure 7.6).

Finally, the *vocal resonances*, or *formants*, which are used for the differential pronunciation of different sounds, particularly speech sounds such as vowels and diphthongs, can be measured. With the help of the various articulatory organs, such as the palate, the velum, the tongue, the lips, and the pharynx, the vocal tract can be constricted in several places and its length can be varied. As predicted by acoustic theory (see Fant, 1960) the site of constriction, the degree of constriction, and the relative degree of lip opening will determine the specific resonances of the acoustic tube formed by the vocal tract (see Figure 7.7).

Acoustically, the vocal resonances are measured as formants, i.e. specific energy peaks in the power spectrum, using methods similar to those used in the measurement of voice quality. A number of digital alogorithms are available for tracking formants in the spectrum. The parameters used are the frequency of a

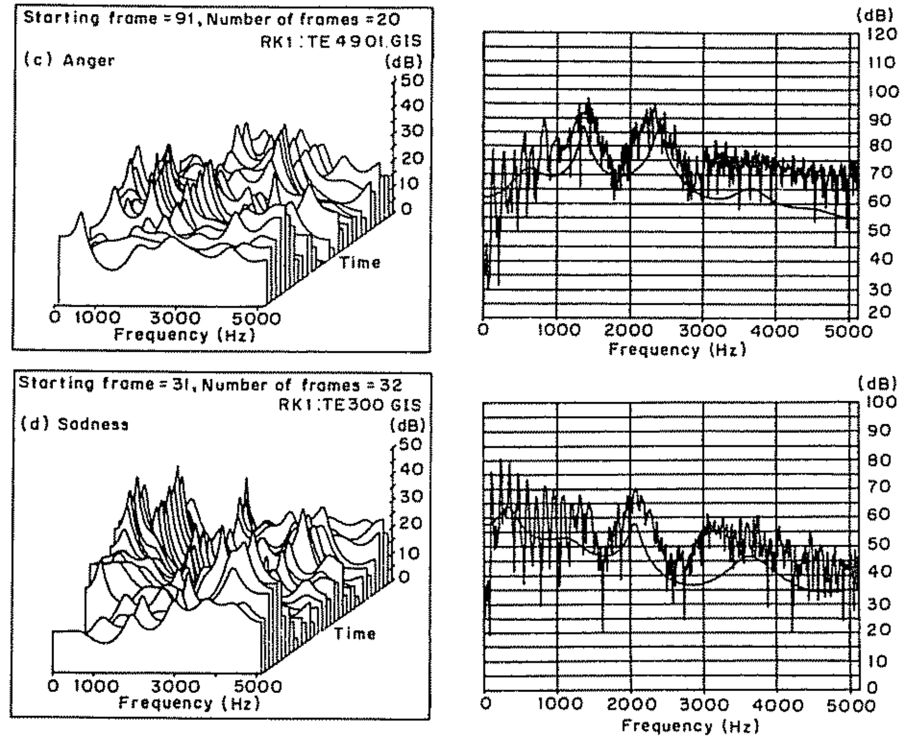


Figure 7.5. 3D-spectra (left) and power spectra with smoothed envelope showing the formants (right) for a logatome taken from a short standard utterance produced by an actor under four different emotional portrayal conditions.

formant (generally only the first three formants are used), its amplitude, and its bandwidth (which is defined as the area below the curve containing a specific proportion of energy) (see Figure 7.4e). Again, while in theory formant measurement is rather straightforward, in reality a large number of problems related to speaker, recording, and speech material may prevent accurate measurement. In terms of summary statistics, great care has to be taken to average only those formant values that are directly comparable, i.e. mean formants for particular vowels as well as the respective dispersion measures. The graphic display of formant contours is frequently used in psycholinguistic studies.

Table 1 provides a summary of the vocal parameters discussed so far, including the most important dispersion measures used. This overview lists those variables which have been used most frequently in research to date. It should be kept in mind, however, that most of these parameters have been developed in the context of research on speech transmission. It is quite possible that these measures need to be supplemented by others, yet to be developed, which are more specifically

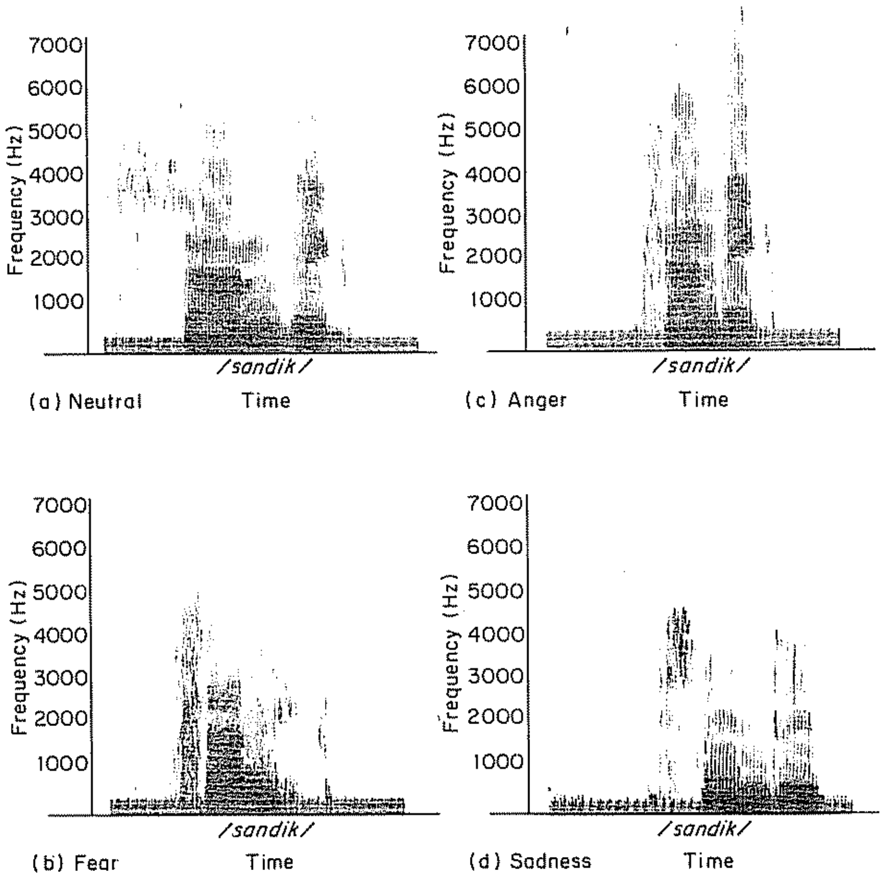


Figure 7.6. Sonagram, or spectrogram, of the logatomes shown in Figure 7.5

oriented toward an assessment of the physiological substratum of voice production.

REVIEW OF THE EVIDENCE IN THE LITERATURE TO DATE

In the following section we shall review the results of the major studies that have been conducted with a view to determining the effect of emotion on vocal behavior. As mentioned in earlier reviews (Scherer, 1979, 1986a) the studies in this area are not very satisfactory since, in general, actor-portrayed emotional utterances have been studied, as opposed to naturally occurring emotional

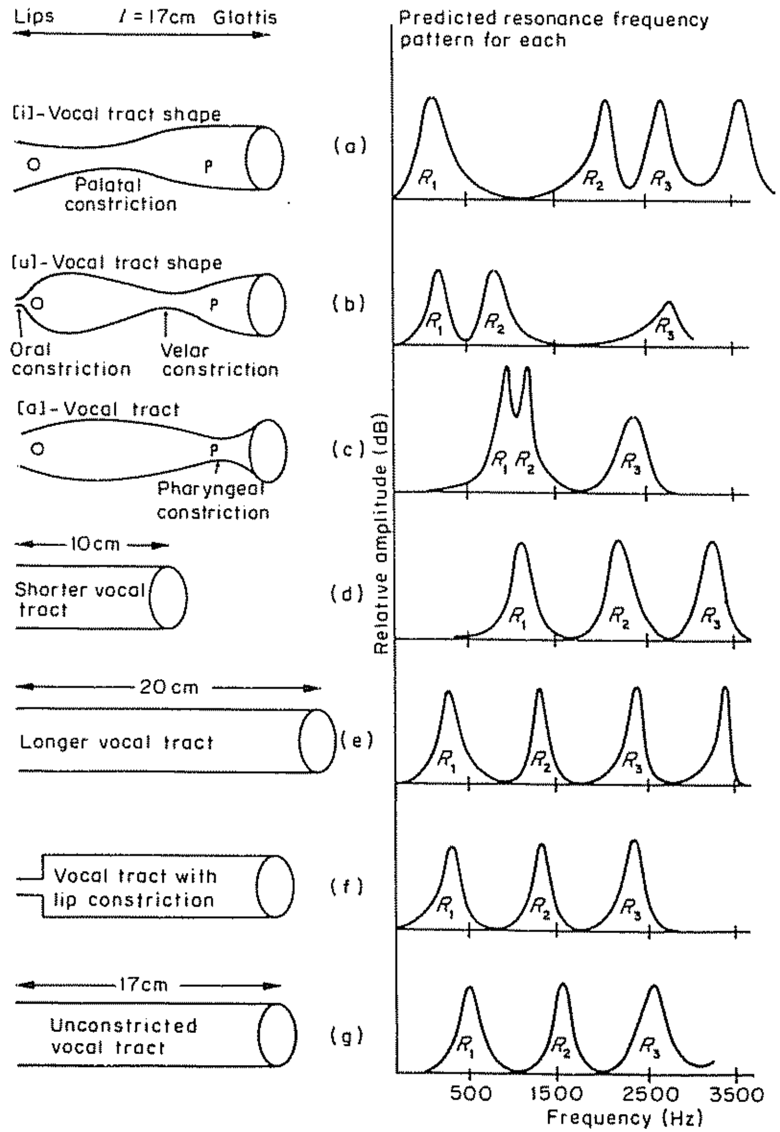


Figure 7.7. Resonance frequency patterns for vocal tracts of differing length, or constricted at differing places along the tract
Note: all resonance patterns should be compared with that for (G), which is the resonance pattern for an unconstricted vocal tract.
(From Daniloff, Schuckers, and Feth (1980) *The Physiology of Speech and Hearing: An Introduction*, © 1980, pp. 17, 174, 203. Reprinted by permission of Prentice-Hall, Inc., Englewood Cliffs, NJ.)

Table 7.1. Overview of major acoustic parameters

Parameter	Description
F_0 perturbation	Slight variations in the duration of glottal cycles
F_0 mean	Fundamental frequency (vibration rate of vocal folds as averaged over a speech utterance)
F_0 range	Difference between highest and lowest F_0 in an utterance
F_0 variability	Measure of dispersion (e.g. standard deviation of F_0)
F_0 contour	Fundamental frequency values plotted over time (intonation)
F_1 mean	Frequency of the first (lowest) formant (significant energy concentration in the spectrum) averaged over an utterance
F_2 mean	Mean frequency of the second formant
Formant bandwidth	Width of the spectral band containing significant formant energy
Formant precision	Degree to which formant frequencies attain values prescribed by phonological system of a language
Intensity mean	Energy values for a speech sound wave averaged over an utterance
Intensity range	Difference between highest and lowest intensity values in an utterance
Intensity variability	Measure of dispersion of intensity values in an utterance (e.g. standard deviation)
Frequency range	Difference between F_0 and highest point in the frequency spectrum where there is still speech energy
High-frequency energy	Relative proportion of energy in the upper frequency region (e.g. > 1 kHz)
Spectral noise	Aperiodic energy components in the spectrum
Speech rate	Number of speech segments per time unit

Note: F_0 = fundamental frequency; F_1 = first formant; F_2 = second formant.
Source: Scherer (1986a), p. 149. (© 1986 by the American Psychological Association. Reprinted by permission of the publisher and author.)

vocalizations. Furthermore, in these encoding studies a number of variables have been only imperfectly controlled, such as the number of speakers, the type of emotions studied, the instructions for portrayal, the verbal material used, etc. (see Scherer, 1986a; Wallbott and Scherer, 1986). However, since there is a rather strong degree of convergence in most of these findings in spite of the methodological shortcomings, it seems useful to take stock of the results reported in the literature. This review will cover discrete emotions, emotional disorder (in particular depression), and stress.

VOCAL INDICATORS OF DISCRETE EMOTIONS

As has been argued elsewhere (Scherer, 1986a), one of the major problems in trying to reconcile some of the contradictions found in the literature is the fact that most researchers have used rather broad emotion categories—often including in the same category rather mild or passive forms of a particular emotion as well as

strong and active forms. For example, this is true for anger; one must assume on the basis of the methods described in the relevant studies that sometimes cold anger, or irritation, (which can be rather subdued) has been studied, while at other times flaring rage has been studied. Although it is difficult on the basis of descriptions in the literature to separate these different emotional states, it seems important to do so in order to understand better the pattern in the data and to provide a broader basis for future work. Consequently, the following review will use fairly differentiated emotion categories (see also Scherer, 1986a). In the following, we shall restrict coverage of particular acoustic parameters and their relation to particular types of emotional arousal to those variables that have repeatedly been found to correlate with certain emotions (i.e. findings from a single study which have not been replicated are not reported).

Boredom/indifference

As one might expect on the basis of the assumption of lowered activation level, the results point to a decrease in mean F_0 (Davitz, 1964; Fairbanks and Pronovost, 1939) and mean intensity (Bortz, 1966; Davitz, 1964; Müller, 1960).

Displeasure/disgust

The results are not very consistent for this emotion, since three studies (Plaikner, 1970; Scherer, 1979; Scherer, Wallbott, Tolkmitt and Bergmann, 1985) find an increase in mean F_0 whereas two others (Kaiser, 1962; Van Bezooijen, 1984) find a lowering of mean F_0. It is possible that variation of the induction procedure used in these studies is partly responsible for this discrepancy. Studies finding an increase of F_0 induced displeasure/disgust with unpleasant films, whereas in the studies finding a decrease actors were asked to simulate disgust. This discrepancy raises the point that affect control or display strategies are likely to affect induction and portrayal studies in very different ways.

Contempt/scorn

This emotion has been included in the emotion sets studied in four of the relevant reports (Costanzo, Markel, and Costanzo, 1969; Fairbanks and Hoaglin, 1941; Fairbanks and Pronovost, 1939; Van Bezooijen, 1984). There is no consistent evidence for any one acoustic parameter; mainly due to the fact that each of the four studies has investigated different acoustic parameters.

Irritation/cold anger

The distinction made here between cold and hot anger has not been made in the literature. The following considerations are, therefore, based on a tentative classification of the research literature according to type of induction procedure or portrayal used. Cold anger seems to lead to an increase in mean F_0 (Eldred and Price, 1958; Roessler and Lester, 1976), mean intensity (Costanzo, Markel, and Costanzo, 1969; Eldred and Price, 1958), high frequency energy (Kaiser, 1962; Roessler and Lester, 1976) and a tendency for intonation contours F_0 contours to be directed downwards (Höffe, 1960; Kaiser, 1962).

Rage/hot anger

As for cold anger, one finds increases in mean F_0 (Davitz, 1964; Fairbanks and Pronovost, 1939; Fonagy, 1978; Höffe, 1960; Wallbott and Scherer, 1986; Williams and Stevens, 1969, 1972) and mean intensity (Bortz, 1966; Davitz, 1964; Höffe, 1960; Kotlyar and Morozov, 1976; Müller, 1960; Van Bezooijen, 1984; Williams and Stevans, 1969). However, by contrast with cold anger, what is particularly noticeable are increases in F_0 variability and total range over the utterances studied (Fairbanks and Pronovost, 1939; Havrdova and Moravek, 1979; Höffe, 1960; Williams and Stevens, 1969). Since there has been no study trying to compare cold anger and hot anger systematically, it is very difficult to try to use the existing evidence to discriminate these two varieties of anger. This problem is exacerbated by the fact that the state of measurement and reporting in this area does not permit one to make parametric comparisons in order to determine, for example, whether an F_0 change, is of small, medium, or great size.

Sadness/dejection

We find strong consensus on the acoustic correlates for this emotion. This may be due to the fact that there is less variability in the eliciting situations and the social strategies for display or control of this emotion (see Scherer, Wallbott, and Summerfield, 1986). Sadness seems generally to decrease mean F_0 (Coleman and Williams, 1979; Davitz, 1964; Eldred and Price, 1958; Fairbanks and Pronovost, 1939; Fonagy, 1978; Kaiser, 1962; Sedlacek and Sychra, 1963; Wallbott and Scherer, 1986; Williams and Stevens, 1969), F_0 range (Fairbanks and Pronovost, 1939; Fonagy, 1978; Sedlacek and Sychra, 1963; Van Bezooijen, 1984; Williams and Stevens, 1969; Zuberbier, 1957; Zwirner, 1930) and downward directed F_0 contours (Fairbanks and Pronovost, 1939; Kaiser, 1962; Sedlacek and Sychra, 1963; Zwirner, 1930). Similarly, mean intensity is lowered (Davitz, 1964; Eldred and Price, 1958; Hargreaves, Starkweather and Blacker, 1965; Huttar, 1968; Kaiser, 1962; Müller, 1960; Skinner, 1935; Zuberbier, 1957). Though based on very few studies, there is suggestive evidence for a decrease in high frequency energy (Hargreaves *et al.*, 1965; Kaiser, 1962; Skinner, 1935) and in the precision of articulation as indexed by formant precision (Zuberbier, 1957; Van Bezooijen, 1984).

Grief/desperation

As for anger, a differentiation between a passive reaction to loss, i.e. sadness and dejection, and a more active desperate grief reaction is rarely made in the emotion literature, and even more rarely studied. As will be shown below, one would, in theory, expect rather different acoustic correlates for grief/desperation from those found for sadness/dejection.

Worry/anxiety

As in the earlier distinctions between milder, more passive, and stronger, more active, forms of an emotion, we distinguish between worry/anxiety and fear/terror. Anxiety has been very frequently studied in relation to verbal and temporal measures (Siegman and Feldstein, 1987) but only rarely in terms of vocal parameters. A number of studies (Bonner, 1943; Hicks, 1979; Höffe, 1960; Plaikner, 1970) seem to suggest that mean F_0 increases.

Fear/terror

As for sadness, we find a strong consensus on acoustic correlates of fear. As one might expect on the basis of the very strong parasympathetic arousal, there is evidence for increase in mean F_0 (Coleman and Williams, 1979; Duncan, Laver and Jack, 1983; Fairbanks and Pronovost, 1939; Fonagy, 1978; Höffe, 1960; Kuroda, Fujiwara, Okamura, and Utsuki, 1976; Niwa, 1971; Roessler and Lester, 1976; Sulc, 1977; Utsuki and Okamura, 1976; Williams and Stevens, 1969), F_0 range (Fairbanks and Pronovost, 1939; Utsuki and Okamura, 1976; Williams and Stevens, 1969), F_0 variability (Fairbanks and Pronovost, 1939; Williams and Stevens, 1969), and high frequency energy (Roessler and Lester, 1976, 1979; Simonov and Frolov, 1973).

Enjoyment/happiness

Most studies have investigated active joy or elation rather than quiet enjoyment/happiness. Consequently, there is little or no empirical evidence relating to this state of peaceful enjoyment.

Joy/elation

Again, we find remarkably consistent increases in mean F_0 (Coleman and Williams, 1979; Davitz, 1964; Fonagy, 1978; Havrdova and Moravek, 1979; Höffe, 1960; Huttar, 1968; Kaiser, 1962; Sedlacek and Sychra, 1963; Skinner, 1935; Van Bezooijen, 1984), F_0 range (Fonagy, 1978; Havrdova and Moravek, 1979; Höffe, 1960; Huttar, 1968; Sedlacek and Sychra, 1963; Skinner, 1935), and F_0 variability (Havrdova and Moravek, 1979; Sedlacek and Sychra, 1963; Skinner, 1935). Mean intensity also seems to increase (Davitz, 1964; Höffe, 1960; Huttar, 1968; Kaiser, 1962; Kotlyar and Morosov, 1976; Müller, 1960; Skinner, 1935; Van Bezooijen, 1984).

In reviewing this pattern of findings, the reader will notice that, in general, emotion-induced voice changes seem to be determined by a single dimension formed by several of the parameters. Emotions with high arousal and activity are characterized by increased mean F_0, range, and variability, as well as intensity, whereas the opposite is true of more passive, withdrawn emotions. As noted previously (Scherer, 1981, 1986a) this dimension seems to reflect a general sympathetic response syndrome. This could be construed as evidence for the old argument that the major feature of emotional responding is sympathetic arousal and that it would be difficult to distinguish further the differentially labelled emotion states on the basis of physiological or vocal parameters.

However, it can be shown that judges asked to identify particular emotions on the basis of vocal speech samples attain a very high accuracy (mean accuracy across 45 studies equal 54.4%—or 46.7% if correction for guessing, taking the number of answer alternatives into account, is applied). This is the case even if the speech samples are masked with regard to both content and a large number of specific acoustic variables (Scherer, 1986a). Consequently, the acoustic correlates of the various emotions must be rather robust, quite redundant, and certainly highly differentiated. The reason why empirical research has so far been unable to pinpoint the differentiating acoustic variables may be related to the fact that some important vocal parameters, such as voice quality, which may be important for emotional dimensions such as pleasantness or control rather than straight activation, have so far not been systematically investigated (Scherer, 1986a, p. 145). An early study by Williams and Stevens (1972), in which these authors tried to measure a very large number of acoustic parameters, can still be regarded as justification for the hope that it will be possible to find very distinctive features for specific emotions in the voice. Below, we outline a series of theoretical predictions that might be useful for further research in this area.

VOCAL INDICATORS OF EMOTIONAL DISTURBANCE

So far, most work on emotional disturbance has been directed towards the study of depressive patients. Unfortunately, the general nosological difficulties in the field of psychiatry have had an impact on research in this area. One of the major problems is the lack of differentiation between different patient groups which, given the nature of their disorder, may exhibit very different vocal behaviour. This is true, for example, of the difference between endogenously depressed patients and biphasic, manic-depressive patients. Obviously, for a biphasic patient, one would expect very different vocal behaviour in a depressed phase as compared with a manic phase.

Disregarding these difficulties, we find that there is still a rather remarkable consistency in the findings reported in connection with the difference between normal and depressed speech and those pertaining to change after therapy. This is particularly true for intensity, where there is evidence that depressive patients speak with relatively low intensity (Eldred and Price, 1958; Moses, 1954; Whitman and Flicker, 1966; Zuberbier, 1957); similarly, intensity tends to increase after therapy (Hargreaves and Starkweather, 1964, 1965; Ostwald, 1961, 1963; Tolkmitt, Helfrich, Standke, and Scherer, 1982). Dynamic range, i.e. the range of intensity changes and variability, also seems to be very low in depression (Zuberbier, 1957).

As far as F_0 is concerned, there is much less convergence among the data. In several studies a rather low mean F_0 for depressives (Bannister, 1972; Eldred and Price, 1958; Moses, 1954; Roessler and Lester, 1976) is found, which is somewhat surprising given findings from our own laboratory (Tolkmitt *et al.*, 1982; Klos, Ellgring, and Scherer, 1988) which show that F_0 seems to decrease after therapy. This latter finding has been linked to a decrease in general tension in the striated musculature (Scherer, 1979). It seems possible that depressive patients have lower F_0 in comparison with normal speakers, but that F_0 level is elevated during depressive phases. However, there is at least one study in which F_0 was found to increase with the severity of depression (Whitman and Flicker, 1966).

Most likely, however, the apparent discrepancy can be explained in terms of lack of homogeneity in patient groups, as mentioned above, and differences in measurement methodology. In summarizing the research findings we have referred to F_0 (which is an objective measure of vibration frequency of the vocal folds); however, a number of studies on both emotion and depression use subjective ratings of pitch, rather than objective analyses. Obviously, this is a procedure that does not necessarily yield a reliable estimate of F_0. As mentioned above, pitch judgement is often influenced by other acoustic variables, such as energy distribution in the spectrum and variability of pitch. Since several studies find a rather narrow range and restricted variability for F_0 pitch in depressive speech (Bannister, 1972; Hargreaves *et al.*, 1965; Newman and Mather, 1938; Ostwald, 1964; Zuberbier, 1957), it is possible that judges are influenced by the reduced range and other acoustic factors in subjectively assessing pitch.

In many psychiatry textbooks one finds the observation that depressive speech sounds monotonous. Several studies (Moses, 1954; Newman and Mather, 1938; Zwirner, 1930) have found that patients tend to employ fairly stereotypic downward directed intonation contours that are frequently repeated. It is possible that the impression of monotonousness results from the lack of variability in the choice of intonation contours rather than the flatness of the contour.

Another interesting parameter is the precision of articulation, a factor which could be linked to general muscle tone. There is suggestive evidence that depressive speech is characterized by rather lax articulation (as reflected in formant precision) and that precision of articulation improves after therapy (Tolkmitt *et al.*, 1982; Zuberbier, 1957).

Most of the data in this area are still very preliminary and in need of replication. It would be particularly helpful to define more precisely the different nosological subgroups of emotional disturbance and to base further research on concrete hypotheses concerning the types of vocal changes expected for each subgroup. A series of theoretical predictions of this sort has been presented in Scherer (1987).

VOCAL INDICATORS OF STRESS

The possibility of measuring stress through vocal analysis has become very popular due to the notion of 'voice lie detection'. Unfortunately, the evidence lags far behind the bold claims that some of the commercial firms marketing 'voice lie detectors' (sic!) have made. So far there is no convincing empirical evidence that lying can be detected through simple electroacoustic devices (Hollien, 1981; Scherer, 1981). It is very important to distinguish between the detection of lying and the detection of stress. While lying could occur without any accompanying stress, and would therefore not be detectable in the voice, it is likely that stress does effect vocal parameters. Again, one of the major determinants of stress effects on the voice seems to be sympathetic arousal. As one might expect from the discussion above, this should primarily result in an increase in fundamental frequency and F_0 variability. This has been the typical finding in studies investigating this effect (cf. Ekman, Friesen, and Scherer, 1976; Scherer, 1981; Streeter, Krauss, Geller, Olson, and Apple, 1977; Williams and Stevens, 1969).

One of the major problems in the empirical assessment of stress effects is the powerful effect of individual differences. As is well known from studies of physiological correlates of stress, there seems to be an enormous degree of response specificity. In a series of experimental studies we were able to show that F_0 increase, as a consistent correlate of stress response, may be limited to persons using particular coping styles, as measured by personality scales. In a major experimental study 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 F_0 increase (Scherer, Wallbott, Tolkmitt and Bergmann, 1985)

In these studies we were also able to show that the investigation of stress effects on the voice must take into account the type of *coping strategies* of the subject. In analysing the precision of formant values, our data indicated that a consistent effect for female anxiety-deniers (repressors) takes the form of a tendency to increase precision of articulation with increasing *cognitive* stress and to decrease this precision with mounting *emotional* stress. Thus, personality factors, coping style, and type of stress induction, all seem to determine the vocal response. It can be argued (Scherer, 1986b) that stress responses could profitably be analysed within the framework of a general theory of emotion, assuming that stress occurs in cases where a problem cannot be solved through normal emotional responding (with return to baseline within a standard time-frame).

A THEORETICAL MODEL OF EMOTION EFFECTS ON VOCAL PRODUCTION

One of the major problems in this area is that almost all of the studies that have been conducted are purely correlational in nature. In other words, researchers attempt to observe the covariation of particular vocal parameters with various types of affective state. In addition, many of these studies are based on work with actors' portrayals of emotion. Consequently, much of the definition of the emotional state that underlies the changed vocal performance rests on the verbal labelling of an emotional state, which, as we have seen, tends to be rather gross. Furthermore, this type of approach does not allow one to elucidate the type of physiological mechanism that underlies the observed correlation between emotional change and change in vocal production. In order to understand better the ways in which the psychophysiological processes accompanying emotional arousal

Table 7.2. Sequence of stimulus evaluation checks (SECs)

1. Novelty check. Evaluating whether there is a change in the pattern of external or internal stimulation, particularly whether a novel event has occurred or is to be expected.
2. Intrinsic pleasantness check. Evaluating whether a stimulus event is pleasant, inducing approach tendencies, or unpleasant, inducing avoidance tendencies; based on innate feature detectors or on learned associations.
3. Goal/need significance check. Evaluating whether a stimulus event is relevant to important goals or needs of the organism (relevance subcheck), whether the outcome is consistent with, or discrepant from, the state expected for this point in the goal/plan sequence (expectation subcheck), whether it is conducive or obstructive to reaching the respective goals or satisfying the relevant needs (conduciveness subcheck), and how urgently some kind of behavioural response is required (urgency subcheck).
4. Coping potential check. Evaluating the causation of a stimulus event (causation subcheck) and the coping potential available to the organism, particularly the degree of control over the event or its consequences (control subcheck), the relative power of the organism to change or avoid the outcome through fight or flight (power subcheck), and the potential for adjustment to the final outcome via internal restructuring (adjustment subcheck).
5. Norm/self compatibility check. Evaluating whether the event, particularly an action, conforms to social norms, cultural conventions, or expectations of significant others (external standards subcheck), and whether it is consistent with internalized norms or standards as part of the self-concept or ideal self (internal standards subcheck).

Source: Scherer (1986a) p. 147. (© 1986 by the American Psychological Association. Reprinted by permission of the publisher and author.)

affect voice production and the resulting acoustic output, it is necessary to develop specific hypotheses concerning the relationship between affect and vocalization.

Unfortunately, apart from Darwin's early speculations, there are very few attempts to predict specific physiological changes in the voice-producing organs as a result of specific emotions. This is also basically true for facial expression, where the patterns that have been observed for discrete emotions are generally linked to rather ill-defined neural programs (Izard, 1977a, b; Tomkins, 1962, 1963).

Recently, an attempt has been made to 'decompose' the classic emotional states into a number of components which may allow more specific predictions. Proposing a 'component process' theory, Scherer (1984, 1986a) has argued that emotional states are produced by the outcomes of a series of five stimulus evaluation checks. In other words, the organism is seen to scan the environment constantly and to evaluate information on the basis of the five criteria or checks listed in Table 7.2 As is shown in more detail elsewhere, most of the major emotions which are referenced by verbal labels can be conceived of as resulting from a specific constellation of outcomes of these five checks. Table 7.3 shows a theoretical prediction table, proposing an outcome profile as an explanatory construct for emotional states habitually labelled with the emotion word in question.

How does this approach generate specific predictions for emotional impact on voice production? The major difference from discrete emotion theories is that in the context of the component process model, a 'component patterning' theory (Scherer, 1984, 1986a) proposes specific changes in the various subsystems of the organism which are seen to subserve emotion (physiological responses, motor expression, motivational tendencies, subjective feeling states). Thus, the outcome of each check is seen to affect all the different emotion components in a 'value-added' function. Given that the organism constantly evaluates and reevaluates ongoing stimulation on the basis of these checks, one can expect constant modifications of the state of the various subsystems on the basis of the sequences of changes in the outcomes of the checks.

The changes expected for the various subsystems for each type of outcome are based on functional considerations with a strong phylogenetic bias. Table 7.4 shows the predictions made for the major components of emotion.

More specifically, on the basis of literature in acoustic phonetics and vocal physiology one can draw up a very detailed pattern of predictions for the voice production domain. This is shown in Table 7.5. Clearly, many of these predictions are rather gross and quite preliminary, given the state of our knowledge in this area. Doubtless, this type of table will have to be revised repeatedly as new evidence becomes available. On the other hand, it does present an initial basis for generating a number of hypotheses concerning vocal changes in emotion. Based on Tables 7.3 and 7.5 we can draw up a detailed prediction table for specific acoustic parameters, which can then be empirically investigated. This is shown in Table 7.6.

Table 7.3. Hypothetical outcomes of stimulus checks for selected emotional states

			Goal/need significance				Coping potential			Norm compatibility	
Emotional state	Novelty	Pleasantness	Relevance	Expectation	Conduciveness	Urgency	Control	Power	Adjust	External	Internal
Enjoyment/happiness	Low	High	Medium	Consistent	High	Very low	—	—	High	High	High
Elation/joy	High	High	High	Discrepant	High	Low	—	—	Medium	High	High
Displeasure/disgust	Open	Very low	Low	Discrepant	Low	Medium	Open	Open	High	Low	—
Contempt/scorn	Open	Low	Low	Discrepant	Low	Low	Open	High	High	Low	—
Sadness/dejection	Low	Low	High	Discrepant	Obstruct	Low	None	—	Medium	—	—
Grief/desperation	High	Low	High	Discrepant	Obstruct	High	Low	Low	Low	—	—
Anxiety/worry	Low	Open	Medium	Discrepant	Obstruct	Medium	Open	Low	Medium	—	—
Fear/terror	High	Low	High	Discrepant	Obstruct	Very high	Open	Very low	Medium	—	—
Irritation/cold anger	Low	Open	Medium	Discrepant	Obstruct	Medium	High	Medium	High	Low	Low
Rage/hot anger	High	Open	High	Discrepant	Obstruct	High	High	High	High	Low	Low
Boredom/indifference	Very low	Open	Low	Consistent	Obstruct	Low	Medium	Medium	High	—	—
Shame/guilt	Low	Open	High	Discrepant	Obstruct	Medium	High	Open	Medium	Very low	Very low

Source: Scherer (1986a) p. 147. (© 1986 by the American Psychological Association. Reprinted by permission of the publisher and author.)

Table 7.4. Component patterning theory predictions of SEC outcome effects on subsystems

SEC outcome	Organismic functions	Social functions	Support system	Action system: Muscle tone	Face	Voice	Instrumental	Posture	Locomotion
Novelty									
Novel	Orienting Focusing	Alerting	Orienting response	Local changes	Brows/lids up Open orifices	Interruption Inhalation	Interruption	Straightening Raising head	Interruption
Old	Homeostasis	Reassuring	No change	No change	No change	No change	No change	No change	No change
Intrinsic pleasantness									
Pleasant	Incorporation	Recommending	Sensitization of sensorium	Slight decrease	Expanding orifices 'sweet face'	Wide voice	Centripetal movement	Expanding Opening	Approach
Unpleasant	Expulsion Rejection	Warning Decommending	Defense response: desensitization	Increase	Closing orifices. 'sour face'	Narrow voice	Centrifugal movement	Shrinking Closing in	Avoidance Distancing
Goal/need significance									
Consistent	Relaxation	Announcing stability	Trophotropic shift	Decrease	Relaxed tone	Relaxed voice	Comfort position	Comfort position	Rest position
Discrepant	Activation	Announcing activity	Ergotropic dominance	Increase	Corrugator	Tense voice	Task-dependent	Task-dependent	Task-dependent
Coping potential									
No control	Readjustment	Indicating withdrawal	Trophotropic dominance	Hypotonus	Lowered eyelids	Lax voice	No activity or slowing	Slump	No movement or slowing
High power control	Goal assertion	Dominance assertion	Ergo-tropho balance Noradrenaline discharge Respiration volume up	Slight decrease Tension in head and neck	Baring teeth Tensing mouth	Full voice	Agonistic movement	Anchoring body, lean forward	Approach
Low power control	Protection	Indicating submission	Ergotropic dominance Adrenaline discharge Peripheral vasoconstriction Respiration rate up	Hypertonus Tension in locomotor areas	Open mouth	Thin voice	Protective movement	Readiness for locomotion	Fast locomotion or freezing

Source: Scherer (1985), p. 216.

Table 7.5. Component patterning theory predictions of vocal changes after different SEC outcomes

Novelty check	
Novel	**Old**
Interruption of phonation Sudden inhalation Silence Ingressive (fricative) sound with a glottal stop (noise-like spectrum)	No change
Intrinsic pleasantness check	
Plesant	**Unpleasant**
Faucal and pharyngeal expansion, relaxation of tract walls Vocal tract shortened by mouth, corners retracted upward More low-frequency energy, F_1 falling, slightly broader F_1 bandwidth, velopharyngeal nasality Resonances raised *Wide voice*	Faucal and pharyngeal constriction, tensing of tract walls Vocal tract shortened by mouth corners retracted downward More high-frequency energy, F_1 rising, F_2 and F_3 falling, narrow F_1 band-width, laryngopharyngeal nasality Resonances raised *Narrow voice*
Goal/need significance check	
Relevant and consistent	**Relevant and discrepant**
Shift toward trophotropic side: overall relaxation of vocal apparatus, increased salivation F_0 at lower end of range, low-to-moderate amplitude, balanced resonance with slight decrease in high-frequency energy *Relaxed voice* If event conducive to goal: relaxed voice + wide voice If event obstructive to goal: relaxed voice + narrow voice	Ergotropic dominance: overall tensing of vocal apparatus and respiratory system, decreased salivation F_0 and amplitude increase, jitter and shimmer, increase in high-frequency energy, narrow F_1 bandwidth, pronounced formant frequency differences *Tense voice* If event conducive to goal: tense voice + wide voice If event obstructive to goal: tense voice + narrow voice

Table 7.5 (Contd.)

Coping potential check	
Control	No control
Ergotropic dominance: (see tense voice)	Trophotropic dominance: hypotension of the musculature in the vocal apparatus and respiratory system
(see tense voice)	Low F_0 and restricted F_0 range, low amplitude, weak pulses, very low high-frequency energy, spectral noise, formant frequencies tending toward neutral setting, broad F_1 bandwidth
Tense voice	*Lax voice*
Power	No power
Deep, forceful respiration; chest register phonation	Rapid, shallow respiration; head register phonation
Low F_0 high amplitude, strong energy in entire frequency range	Raised F_0, widely spaced harmonics with relatively low energy
Full voice	*Thin voice*
Norm/self compatibility check	
Standards surpassed	Standards violated
Wide voice + full voice	Narrow voice + thin voice
+ Relaxed voice (if expected)	+ Lax voice (if no control)
+ Tense voice (if unexpected)	+ Tense voice (if control)

Source: Scherer (1986a), p. 156. (© 1986 by the American Psychological Association. Reprinted by permission of the publisher and author.)

Table 7.6. Changes predicted for selected acoustic parameters on the basis of the voice type predictions in Table 7.4

	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	>	<			>	<		<		<		

Note: 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/coldanger; 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; > = increase; < = decrease; 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), p. 158. (© 1986 by the American Psychological Association. Reprinted by permission of the publisher and author.)

Much of the foregoing has been highly speculative in nature. However, it is argued that the approach taken here will help to establish closer links between theories of emotion, psychophysiological research, and the acoustic measurement of vocal behaviour. Should this approach prove feasible, it would help researchers to go beyond the fairly atheoretical correlational approach that presently characterizes the field, and would thereby provide a better understanding of the physiological mechanisms involved.

REFERENCES

Bannister, M. L. (1972) An instrumental and judgemental analysis of voice samples from psychiatrically hospitalized and nonhospitalized adolescents. Unpublished Ph.D thesis, University of Kansas.

Bonner, M. R. (1943) Changes in the speech pattern under emotional tension. *American Journal of Psychology*, **56**, 262–73.

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

Coleman, R. F., and Williams, R. (1979) Identification of emotional states using perceptual and acoustic analyses. In V. Lawrence and B. Weinberg (Eds), *Transcript of the Eighth Symposium: Care of the Professional Voice*, Part I. New York: The Voice Foundation.

Costanzo, F. S., Markel, N. N., and Costanzo, P. R. (1969) Voice quality profile and perceived emotion. *Journal of Counseling Psychology*, **16**, 267–70.

Daniloff, R., Schuckers, G., and Feth, L. (1980) *The Physiology of Speech and Hearing: An Introduction*. Englewood Cliffs, NJ: Prentice-Hall.

Davitz, J. R. (1964) *The Communication of Emotional Meaning*. New York: McGraw-Hill.

Duncan, G., Laver, J., and 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.

Ekman, P., Friesen, W. V., and Scherer, K. R. (1976) Body movement and voice pitch in deceptive interaction. *Semiotica*, **16**, 23–7.

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

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

Fant, G. (1960) *Acoustic Theory of Speech Production*. The Hague: Mouton.

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

Fry, D. B. (1979) *The Physics of Speech*. Cambridge: Cambridge University Press.

Hargreaves, W. A., and Starkweather, J. A. (1964) Voice quality changes in depression. *Language and Speech*, **7**, 84–8.

Hargreaves, W. A., and Starkweather, J. A. (1965) Vocal and verbal indicators of depression. Unpublished manuscript, San Francisco.

Hargreaves, W. A., Starkweather, J. A., and Blacker, K. H. (1965) Voice quality in depression. *Journal of Abnormal Psychology*, **70**, 218–20.

Havrdova, Z., and Moravek, M. (1979) Changes of the voice expression during suggestively influenced states of experiencing. *Activitas Nervosa Superior*, **21**, 33–5.

Hess, W. (1983) *Pitch Determination of Speech Signals*. Berlin: Springer.

Hicks, J. W. (1979) An acoustical/temporal analysis of emotional stress in speech. *Dissertation Abstracts International*, **41** (4-A).

Hollien, H. (1981) Analog instrumentation for acoustic speech analysis. In J. Darby (Ed.), *Speech Evaluation in Psychiatry*, pp. 79–103. New York: Grune & Stratton.

Huttar, G. L. (1968) Relations between prosodic variables and emotions in normal American English utterances. *Journal of Speech and Hearing Research*, **11**, 481–7.

Höffe, W. L. (1960) Über Beziehungen von Sprachmelodie und Lautstärke. *Phonetica*, **5**, 129–59.

Izard, C. E. (1977a) The emergence of emotions and the development of consciousness in infancy. In J. M. Davidson, R. J. Davidson, and G. E. Schwartz (Eds), *Human Consciousness and its Transformations: A Psychobiological Perspective*. New York: Plenum Press.

Izard, C. E. (1977b) *Human Emotions*. New York: Plenum Press.

Jürgens, U. (1979) Vocalization as an emotional indicator. A neuroethological study in the squirrel monkey. *Behaviour*, **69**, 88–117.

Kaiser, L. (1962) Communication of affects by single vowels. *Synthese*, **14**m 300–19.

Klos, Th., Ellgring, H., and Scherer, K. R. (1988) Vocal changes in depression. Unpublished manuscript, University of Geneva.

Kotlyar, G. M., and Morozov, V. P. (1976) Acoustical correlates of the emotional content of vocalized speech. *Sov. Phys. Acoust.*, **22**, 208–11.

Kuroda, I., Fujiwara, O., Okamura, N., and Utsuki, N. (1976) Method for determining pilot stress through analysis of voice communication. *Aviation, Space, and Environmental Medicine*, **47**, 528–33.

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

Leyhausen, P. (1967) Biologie von Ausdruck und Eindruck, Teil I. *Psychologische Forschung*, **31**, 113–76.

Lieberman, P. (1961) Perturbations in vocal pitch. *Journal of the Acoustical Society of America*, **33**, 597–603.

Lieberman, P. (1977) *Speech Physiology and Acoustic Phonetics: An Introduction*. New York: Macmillan.

Marler, P. (1984) Animal communication: affect or cognition? In K. R. Scherer and P. Ekman (Eds), *Approaches to Emotion*, pp. 345–68. Hillsdale, NJ: Erlbaum.

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, Universität Göttingen.

Newman, S. S. and Mather, V. G. (1938) Analysis of spoken language of patients with affective disorders. *American Journal of Psychiatry*, **94**, 913–42.

Niwa, S. (1971) Changes of voice characteristics in urgent situations. (2) *Reports of the Aeromedical Laboratory, Japan Air Self Defense Force*, **11**, 246–51.

Ostwald, P. F. (1961) The sound of emotional disturbance. *Archives of General Psychiatry*, **5**, 587–92.

Ostwald, P. F. (1963) *Soundmaking: The Acoustic Communication of Emotion*. Springfield: C. C. Thomas.

Ostwald, P. F. (1964) Acoustic manifestations of emotional disturbance. *Disorders of Communication—Research Publications*, **42**, 450–65.

Plaikner, D. (1970) Die Veränderungen der menschlichen Stimme unter dem Einfluss psychischer Belastung. Unpublished doctoral dissertation, Universität Innsbruck.

Robinson, W. P. (1972) *Language and Social Behavior*. Harmondsworth: Penguin.

Roessler, R., and Lester, J. W. (1976) Voice predicts affect during psychotherapy. *Journal of Nervous and Mental Disease*, **163**, 166–176.
Roessler, R., and Lester, J. W. (1979) Vocal patterns in anxiety. In W. E. Fann, A. D. Pokorny, I. Koracau, and R. L. Williams (Eds), *Phenomenology and Treatment of Anxiety*. New York: Spectrum.
Scherer, K. R. (1979) Nonlinguistic vocal indicators of emotion and psychopathology. In C. E. Izard (Ed.), *Emotions in Personality and Psychopathology*, pp. 493–529. New York: Plenum Press.
Scherer, K. R. (1981) Vocal indicators of stress. In J. Darby (Ed.), *Speech Evaluation in Psychiatry*, pp. 171–87. New York: Grune & Stratton.
Scherer, K. R. (1982) Methods of research on vocal communication: paradigms and parameters. In K. R. Scherer and P. Ekman (Eds), *Handbook of Methods in Nonverbal Behaviour Research*, pp. 136–98. Cambridge: Cambridge University Press.
Scherer, K. R. (1984) On the nature and function of emotion: a component process approach. In K. R. Scherer and P. Ekman (Eds), *Approaches to Emotion*, pp. 293–318. Hillsdale, NJ: Erlbaum.
Scherer, K. R. (1985) Vocal affect signalling: a comparative approach. In J. Rosenblatt, C. Beer, M.-C. Busnel, and P. J. B. Slater (Eds), *Advances in the Study of Behavior*, Vol. 15, pp. 189–244. New York: Academic Press.
Scherer, K. R. (1986a) Vocal affect expression: a review and a model for future research. *Psychological Bulletin*, **99**, 143–65.
Scherer, K. R. (1986b) Voice, stress, and emotion. In M. H. Appley and R. Trumbull (Eds), *Dynamics of Stress*, pp. 159–81. New York: Plenum Press.
Scherer, K. R. (1987) Vocal assessment of affective disorders. In J. D. Maser, (Ed), *Depression and Expressive Behavior*, pp. 57–82. Hillsdale, NJ: Erlbaum.
Scherer, K. R. (1988, in press) Toward a dynamic theory of emotion: the component process model of affective states.
Scherer, K. R., and Kappas, A. (1988, in press) Vocal affect expression in primates. In D. Todt, P. Goedeking, and E. Newman, (Eds), *Primate vocal communication*, pp. 171–94. Heidelberg: Springer.
Scherer, K. R., Wallbott, H. G. and Summerfield, A. B. (Eds) (1986) *Experiencing Emotion: A Cross-Cultural Study*. Cambridge: Cambridge University Press.
Scherer, K. R., Wallbott, H. G., Tolkmitt, F., and Bergmann, G. (1985) *Die Stress reaktion: Physiologie und Verhalten*. Göttingen: Hogrefe.
Sedlacek, K., and Sychra, A. (1963) Die Melodie als Faktor des emotionellen Ausdrucks. *Folia Phoniatrica*, **15**, 89–98.
Seyfarth, R. M. and 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, and M. R. Petersen (Eds), *Primate Communication*, pp. 239–52. Cambridge: Cambridge University Press.
Siegman, A. W., and Feldstein, S. (1987) *Nonverbal Behavior and Communication*, 2nd edn. Hillsdale, NJ: Erlbaum.
Simonov, P. V., and Frolov, M. V. (1973) Utilization of human voice for estimation of man's emotional stress and state attention. *Aerospace Medicine*, **44**, 256–8.
Skinner, E. R. (1935) A calibrated recording and analysis of the pitch, force and quality of vocal tones expressing happiness and sadness. *Speech Monographs*, **2**, 81–137.
Sorensen, D. and Horii, J. (1984) Directional perturbation factors for jitter and shimmer. *Journal of Communication Disorders*, **12**, 143–57.
Streeter, L. A., Krauss, R. M., Geller, V., Olson, C., and Apple, W. (1977) Pitch changes during attempted deception. *Journal of Personality and Social Psychology*, **35**, 345–50.
Sulc, J. (1977) To the problem of emotional changes in the human voice. *Activitas Nervosa Superior*, **19**, 215–16.

Tolkmitt, F. J., Helfrich, H., Standke, R., and Scherer, K. R. (1982) Vocal indicators of psychiatric treatment effects in depressives and schizophrenics. *Journal of Communication Disorders*, **15**, 209–22.
Tomkins, S. S. (1962) *Affect, Imagery, Consciousness. Vol. 1. The Positive Affects*. New York: Springer.
Tomkins, S. S. (1963) *Affect, Imagery, Consciousness. Vol. 2. The Negative Affects*. New York: Springer.
Utsuki, N., and Okamura, N. (1976) Relationship between emotional state and fundamental frequency of speech. *Reports of the Aeromedical Laboratory, Japan Air Self Defense Force*, **16**, 179–88.
Van Bezooijen, R. (1984) *The Characteristics and Recognizability of Vocal Expressions of Emotion.*. Dordrecht: Foris.
Wallbott, H. G., and Scherer, K. R. (1986) Cues and channels in emotion recognition. *Journal of Personality and Social Psychology*, **51**, 690–9.
Whitman, E. N. and Flicker, D. J. (1966) A potential new measurement of emotional state: A preliminary report. *Newark Beth Israel Hospital*, **17**, 167–172.
Williams, C. E., and Stevens, K. N. (1969) On determining the emotional state of pilots during flight: An exploratory study. *Aerospace Medicine*, **40**, 1369–72.
Williams, C. E., and Stevens, K. N. (1972) Emotions and speech: Some acoustical correlates. *Journal of the Acoustical Society of America*, **52**, 1238–1250.
Zemlin, W. R. (1968) *Speech and Hearing Science* (2nd edn. 1981). Englewood Cliffs, NJ: Prentice-Hall.
Zuberbier, E. (1957) Zur Schreib- und Sprechmotorik der Depressiven. *Zeitschrift für Psychotherapie und Medizinische Psychologie*, **7**, 239–49.
Zwirner, E. (1930) Beitrag zur Sprache des Depressiven. *Phonometrie III, Spezielle Anwendungen I*, pp. 171–87. Basel: Karger.