Vocal cues in emotion encoding and decoding

Scherer, Klaus R.; Banse, Rainer; Wallbott, Harald G.; Goldbeck, Thomas

How to cite

SCHERER, Klaus R. et al. Vocal cues in emotion encoding and decoding. In: Motivation and Emotion, 1991, vol. 15, n° 2, p. 123–148. doi: 10.1007/BF00995674

This publication URL: https://archive-ouverte.unige.ch/unige:101980

Publication DOI: 10.1007/BF00995674

Motivation and Emotion, Vol. 15, No. 2, 1991

Vocal Cues in Emotion Encoding and Decoding[1]

Klaus R. Scherer[2] and Rainer Banse
University of Geneva

Harald G. Wallbott and Thomas Goldbeck
University of Giessen

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

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

[1]This research was supported by a grant from the Deutsche Forschungsgemeinschaft (Sche 156/8-5). The authors acknowledge collaboration of Westdeutscher Rundfunk, Cologne, in producing professional versions of actor emotion portrayals, and thank Kurt Balser, Theodor Gehm, Christoph Gierschner, Judy Hall, Ursula Hess, Alice Isen, and one anonymous reviewer for helpful comments and contributions.
[2]Address all correspondence, including requests for reprints, to Klaus R. Scherer, University of Geneva, FPSE, 24 rue General Dufour, 1211 Geneva 4, Switzerland.

0146-7239/91/0600-0123$06.50/0

real-life records of vocal expression of specific emotions. It is generally true that access to naturalistic affect expression in public settings is highly limited (see Scherer, 1986; Wallbott & Scherer, 1986b). However, recent technological advances in photography and video recording (using telephoto lenses, unobtrusively positioned cameras, etc.) permit researchers to obtain candid shots of facial and gestural emotion expression samples. In contrast, the limited pick-up range of microphones as well as the pervasive noise problem make it difficult to "eavesdrop" on naturally occurring vocal affect expressions.

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

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

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

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

Frequently actors are just asked to read a passage in such a way as to make the rendering *sound* representative of a particular emotion. A serious problem with this procedure is that the assignment to encode a particular emotional state consists only of a single emotion label. As has been discussed in detail by Scherer (1986), many of the "basic" emotions such as anger, joy, or fear, can take very different forms, depending on the antecedent situation and the type of reaction of the person concerned. This is the case for the difference between hot vs. cold anger, for example. These two types of anger are likely to entail very different types of expression patterns. If only a single label is used to specify the desired emotion, it is left to the actor to imagine a particular situation and to choose the variant of the emotion to be portrayed. Obviously, this procedure makes it rather unlikely that results can be replicated.

Furthermore, there is rarely concern with the quality or competence of actors and their specific ability to convey a particular emotional state convincingly (which is all the more worrisome when lay actors or acting students are used). Wallbott and Scherer (1986b) have shown that differences in competence between actors, particularly the tendency to "overact," may present serious problems in studies of this sort.

The series of studies reported in this paper attempts to improve the general research paradigm, in particular with respect to specific procedures of actor selection, portrayal instruction, and analysis of the stimulus material. Using two male and two female professional radio actors, a standardized emotion portrayal instruction based on a scenario approach, and a specifically developed type of speech sample, this study attempts to determine specific patterns of acoustic cues which characterize the vocal portrayal of particular emotions. Advanced digital speech analysis techniques are used to determine the nature of the acoustic cues used in the expression of the different emotional states.

One of the major purposes of the study was to provide an empirical investigation of the senior author's detailed predictions (based on the component process model of emotion; Scherer, 1984, 1986) concerning the different acoustic patterns expected for a number of specific emotional states. Table III in the Results section reproduces those predictions relevant to the emotions and the parameters studied in this research.[3] The predictions are based on the presumed effects of emotion-specific physiological changes on respiration, phonation, and articulation (see also Scherer, 1989). While it is unlikely that strong physiological changes occur in actors portraying an emotional expression, it is of interest to see whether the vocal cues they employ to encode the presence of an emotional state

[3]It should be noted that this study was designed before the senior author published predictions for more finely differentiated variants of the basic emotion categories. In consequence, the more global predictions (see Table III) for the broad categories fear, joy, sadness, and anger have been tested in this study.

correspond to those that are predicted to occur under conditions of natural elicitation of these emotions.

ENCODING STUDY

Method

Actors. Two male and two female professional actors regularly employed in radio, television, and stage work were recruited by the casting office of a major West German radio station (Westdeutscher Rundfunk, Cologne). All had graduated from professional acting schools and had many years of experience in acting and radio work. They were paid for their participation.

Scenarios. The following five emotions were studied: joy, anger, sadness, fear, and disgust. Two different emotion-eliciting scenarios were used for each emotion. The scenarios selected were based on the authors' intercultural studies on emotion experience in which representative emotion-eliciting situations were collected from over 3,000 respondents on all five continents (Scherer, Wallbott, & Summerfield, 1986; Wallbott & Scherer, 1986a).

The following example illustrates the kind of scenario used (see appendix):

Sadness:

I have to give my beloved pet to strangers, because we are moving and our new landlord doesn't allow pets.

I say sadly: . . .

The actors were asked to read the scenario and imagine to experience the situation described. They were then to speak a standard sentence as they would have uttered it in the respective situation.

Standard Sentence. Contrary to earlier work, in which series of numbers, letters of the alphabet, nonsense syllables, or standard speech material were used, we decided to construct a speech-like but meaningless sentence. With a view toward future cross-cultural studies in which the role of the language spoken by encoders and decoders is of major importance, one of the authors (a phonetician, TG) designed, in a systematic fashion, several synthetic sentences using different phonemic elements from various European languages. From a set of six such synthetic sentences, the following two were chosen for the study after preliminary judgment studies: /haet sandik prongnu ventsi/ and /fi goettlaich schongkill gostaer/. The advantage of this method, as opposed to nonsense syllables, is that the voice samples sound like a real speech utterance with appropriate intonation and rhythm. This is comparable to listening to an unknown language.

Recording the Utterances. After having prepared for the task at home, the actors recorded the utterances in the studios of Westdeutscher Rundfunk, Cologne, under supervision of a professional director. Professional studio equipment was used for the recording.

Design. In total, 88 stimuli were produced in this fashion. Eighty were based on the following design: five emotions by two actors by two genders by two scenarios by two synthetic sentences. In addition, each of the four actors produced the two sentences in a neutral manner, resulting in eight additional stimuli.

DECODING STUDIES

Method

Judgment Study 1. Using professional tape-recording equipment, the 88 utterances were copied in random order onto a stimulus tape for use in judgment studies. Each utterance was repeated once on the tape in order to give judges longer exposure to the vocal expressions. The stimulus tape was stopped after each stimulus and restarted when subjects had finished their rating of that stimulus. Twenty-nine psychology students (21 female, 8 male; mean age = 22.5 years) participated in the first judgment study which was designed to eliminate those stimuli in which the actor's portrayal did not communicate the intended emotion sufficiently clearly. The criterion used was that the judges' impression of the intended emotion should be stronger than that of any of the other emotions. Accordingly, judges were asked to rate each stimulus on all five emotions studied (joy, sadness, fear, anger, disgust), using 7-point scales from least to most intense. The intention had been to select about 40 stimuli (8 per emotion) satisfying the criterion for use in further rating studies. Only those items were retained which had the highest mean on the intended emotion and a difference of at least 0.50 scale point between the highest and the second highest rating. Only 30 item were found which satisfied this criterion. These were used in the further rating studies.

Judgment Study 2. Twenty judges (4 male, 16 female students) were paid to rate the 30 stimuli selected in Study 1. The same procedure as in Study 1 was used to obtain the judgments.

Judgment Study 3. In order to obtain data on the recognition accuracy for emotion portrayal from a sample which was more representative for the general population than student volunteers, it was decided to use an opportunity provided by an open university course (Funkkolleg) which was broadcast once weekly by most of the German radio stations. Within the context of course material on the psychobiology of communication, listeners were requested to participate in a short test of their capacity to recognize vocally portrayed emotions. The 10 stimuli which had been most clearly identified in the previous judgment studies were used. Appropriate rating scales had been printed in the written course material, which was distributed by mail before the radio broadcast. The written instructions specified that each stimulus was to be rated on a 3-point scale, indicating whether the stimulus represented each of the five emotions not at all, somewhat, or strongly. We chose this 3-point, labeled scale to ease the rating task for the radio listeners, who were probably not very experienced in the use of rating scales.

Three hundred and two participants in the open university course who had listened to the broadcast of the stimuli returned their rating sheet by mail. There were 196 female and 106 male subjects who participated (age distribution from 12 to 78 years with the mode in the age range between 34 and 54 years).

Judgment Study 4. In order to evaluate the difference between the radio listener and the student samples, the same 10 stimuli were presented to 98 university students at the University of Marburg, who were asked to rate each stimulus using the same rating scale as the radio listeners. Obviously, the judgment conditions were different in that for the radio listeners the rating procedure involved listening to the radio in their home (allowing no control of various environmental factors). In the case of the student rating study, stimuli were presented via loudspeakers in a university auditorium and the ratings were thus obtained in a more controlled but less motivating group setting.

Results

In this section, we analyzed the judges' ability to recognize the type of emotion intended by the encoder for a particular portrayal. Judges had been given the possibility to rate each stimuli on several emotion scales (see Decoding Studies Method section). The resulting mean profiles of ratings for all four judgment studies are shown in Table I. Since different types of rating scales had been used in the different studies (see Decoding Studies Method section), we transformed all ratings onto a common scale (see footnote *a* to Table I). Looking across the rows of Table I, one can evaluate the degree to which the set of stimuli representing an intended emotion has been rated as pure or blended with other emotions.

Table I. Mean Ratings in the Four Judgment Studies[a]

	Emotion scales				
Intended Emotions	Fear	Disgust	Joy	Sadness	Anger
		Study 1			
N = 29; 88 items					
Fear	.58	0	.16	.05	.38
Disgust	.12	.38	.05	.29	.33
Joy	.06	.08	.64	.06	.14
Sadness	.07	.03	.04	.92	.02
Anger	.02	.09	.02	.06	.92
Neutral	.03	.02	.10	.23	.13
		Study 2			
N = 20; 30 items					
Fear	1.09	.41	.19	.51	.48
Disgust	.36	.55	.17	.63	.61
Joy	.28	.27	.71	.23	.39
Sadness	.55	.13	.10	1.30	.16
Anger	.24	.42	.08	.22	1.32
		Study 3			
N = 307; 10 items					
Fear	1.63	.17	.02	.12	.16
Disgust	.15	.77	.11	.44	.12
Joy	.12	.10	1.47	.64	.05
Sadness	.12	.05	.01	1.65	.03
Anger	.06	.10	.11	.04	1.11
		Study 4			
N = 98; 10 items					
Fear	1.25	.39	.16	.57	.59
Disgust	.42	.57	.37	.72	.42
Joy	.23	.20	1.11	.38	.11
Sadness	.62	.34	.16	1.52	.23
Anger	.24	.25	.27	.25	1.14

[a] *Note:* To render the scale values comparable across studies, we transformed all ratings to a scale ranging from 0 to 2. The values of the 7-point scale (0–6) used in Studies 1 and 2 were divided by 3, and 1 was subtracted from the values of the 3-point scale (1–3) used in Studies 3 and 4. N = number of judges.

In order to obtain, in addition, dichotomous data on recognition, a criterion for the assessment of accuracy was required. We scored an item as correctly recognized if the scale value of the rating of the intended emotion was higher than any other rating. It should be noted that by this criterion ties are scored as incorrect, yielding a rather conservative measure of judgment accuracy. The accuracy data for all four judgment studies are reported in Table II.

This table reports both the raw accuracy percentages and the percentage expected to be correct if judges had guessed blindly. The latter value is computed in such a way to take into account the empirically found frequency of usage of each of the emotion categories proposed (the number of times category *x* is chosen divided by the total number of ratings). This is a more appropriate comparison value than the chance level reported normally (100 divided by the number of categories) since it reflects differences in category use which obviously affect accuracy.[4] The raw accuracy percentage should always be interpreted with respect to this reference value. We have not used the Accuracy Index (accuracy–chance level/chance level) suggested by Wagner (1990), as this index may attain unreasonably high values in extreme cases. For example, if a category is used very rarely, but on those occasions almost always correctly, the Accuracy Index attains a very high value. We believe that this index may be misleading in such cases since it does not take into account that disproportionately infrequent use of a category may also reflect an inability to recognize a certain class of stimuli. In addition, Table II shows the average intercorrelation between judges for each emotion scale, giving an estimate of interrater agreement for each scale.

Judgment Study 1

This study was designed to select the best specimens of the portrayals in terms of unambiguous encoding of the emotions, since one would expect to find many blended portrayals in an unselected set of stimuli. As Table I shows, this is particularly the case for fear and disgust. On the whole, the mean ratings are relatively low, indicating that some of the portrayals were not very intense.

The accuracy results show that, in spite of the large number of as yet unselected stimuli, recognition of all emotions largely exceeds the chance or guessing level (although the accuracy percentage is markedly lower in the case of disgust).

Interrater agreement is quite high, except in the case of disgust.

[4]It should be noted that accuracy percentages cannot be directly compared across different studies if the number of categories used varies. However, Rosenthal and Rubin (1989) have suggested an index, *pi*, that allows one to transform accuracy percentages to a standard dimension of dichotomous choice. This index is computed on the basis of the number of categories used and the raw accuracy percentage.

Table II. Results on Accuracy of Recognition in Four Studies[a]

	Fear	Disgust	Joy	Sadness	Anger	Mean/Sum
			Study 1, N = 29			
Stimuli = 80[b]	17	17	15	16	15	
Recognition	40.8%	27.4%	50.8%	66.8%	69.9%	51.1%
Chance level	11.8%	8.5%	14.9%	21%	25.6%	81.8%[c]
Intercorr.	.51	.22	.50	.52	.60	.48
			Study 2, N = 20			
Stimuli = 30	5	5	6	7	7	
Recognition	49%	21%	48.3%	68.6%	75%	52.4%
Chance level	13.5%	5.7%	13.5%	24.2%	26.8%	83.7%[c]
Intercorr.	.40	.27	.41	.68	.57	.48
			Study 3, N = 307			
Stimuli = 10	2	2	2	2	2	
Recognition	83.6%	43%	81.4%	91%	72.1%	74.2%
Chance level	19.2%	10.8%	19.2%	23.9%	16.6%	89.7%[c]
Intercorr.	.86	.56	.83	.83	.78	.79
			Study 4, N = 98			
Stimuli = 10	2	2	2	2	2	
Recognition	34.7%	18.4%	53.6%	61.7%	56.1%	44.9%
Chance level	10.4%	5.4%	16.3%	22.7%	15%	69.8%[c]
Intercorr.	.36	.16	.40	.52	.46	.39
			Means over the four studies			
Recognition	52%	27.5%	58.5%	72%	68.3%	55.7%
Chance level	13.7%	7.6%	16%	23%	21%	81.3%[c]
Intercorr.	.58	.31	.57	.66	.62	.56

[a] *Note:* Stimuli = number of stimuli, Recognition = percentage of emotion recognition, Chance level = expected chance level for correct guessing, and Intercorr = mean intercorrelation between raters.
[b] Since "neutral" was not included as an alternative in the judgment task, the eight neutral stimuli used in Study 1 do not appear in this table.
[c] The sum is given for chance level accuracy percentages since these add up to the percentage of valid judgments, excluding ties and missing data.

Judgment Study 2

As expected, after selection of relatively unambiguous stimuli, the values of the ratings generally increase, indicating more intense portrayals. However, many stimuli still receive high secondary ratings (on nonintended emotions), indicating potential blends or confusions.

Surprisingly, despite the selection, which theoretically should have improved the accuracy of recognition, the percentages were virtually identical with those in Study 1. This may be due to the fact that in this judgment group a second emotion was often tied with the target emotion (i.e., received an equally high rating, which was counted as not recognized, according to our conservative criterion).

Judgment Study 3

In this study with subjects from a radio listener sample, only the 10 portrayals that were most accurately judged in the earlier studies were used. As one would expect, emotion rating profiles showed fewer blends and confusions, and the interrater agreement as well as the accuracy percentages were much higher than in the earlier studies. This is partly due to a rather low percentage of ties in the ratings (3.8%).

Judgment Study 4.

Again, only the 10 portrayals that were most accurately judged in the earlier studies were presented in this study with student subjects. However, neither the rating profiles nor the accuracy data showed major differences from those of Studies 1 and 2, in which many more and more ambiguous stimuli had been used. The relative drop in accuracy rate is particularly due to the large percentage of ties in this study (19.4%).

General Comments

In general, the stability of both the size of the accuracy percentages and their patterns across emotions is quite remarkable. We did not expect, however, to find a higher degree of accuracy for the radio listener sample, in view of the large variability in age and occupation, than for the more homogeneous student sample in Study 4 (in which the same 10 stimuli and the same 3-point scale were used). This difference between the accuracy proportions is highly significant ($p < .001$; tested by z as suggested by Fleiss, 1981, p. 23-24). In general, the differences between the student samples and the radio listener sample raise interesting questions concerning the existence of different stereotypes, habitual judgment patterns, or differential aptitudes in different groups of the population. We checked on the possibility that the ability to recognize emotions on the basis of vocal expression increases with life experience in the radio listener sample. This is not the case. The correlation between age and percentage accuracy is $r = .11$, n.s. Further research is obviously needed to investigate the potential factors involved. It is possible that students use a more differentiated judgment pattern. The data show that student subjects tend to check more emotions for the same stimulus (leading to ties, which we considered as inaccurate) than the radio listeners. However, students also produced more missing data. Most likely, a motivational factor is also involved: The student samples were recruited in fulfillment of course requirements or received only small payments. The members of the radio listener sample, on the other hand, given the fact that they participated in an open university course and expended the effort to participate in the test, can be expected to have been strongly motivated to do well in the task.

It is of particular interest that, as before, the disgust stimuli were judged with a very low degree of accuracy, in spite of the preselection. This result is rather surprising since disgust is obviously one of the emotions which can be shown to have important biological functions (Rozin & Fallon, 1987) and one which seems to have a major effect of phonation characteristics (see Scherer, 1986; Trojan, 1975). It would be interesting to isolate the factors responsible for the lack of ability of our judges to identify the disgust portrayals. The problem may lie on the sender side. It cannot be excluded that actors are not able convincingly to portray disgust, an emotion rarely used on the stage and one that may also be less frequently used for strategic purposes in everyday social interaction. Another possibility is that disgust might be vocally expressed via interjections or vocal emblems (see Goffman, 1979; Scherer, 1977) rather than via particular types of voice quality changes in running speech.

ACOUSTIC ANALYSES

Method

In order to evaluate the nature of the acoustic differentiation of the vocal emotion portrayals, digital computer analyses of the speech signals were performed. A number of investigators have used objective analysis of voice and speech patterns in actor-produced emotion portrayals (Fairbanks & Hoaglin, 1941; van Bezooijen, 1984; Wallbott & Scherer, 1986b; Williams & Stevens, 1972). The design of these studies often does not allow multifactorial statistical analyses of acoustic differences between the portrayals of different emotions under different conditions. The design of the present study, involving two speakers of each gender, several basic emotions based on two different scenarios per emotion, and use of two different carrier sentences, was originally chosen to allow such systematic statistical analysis of the acoustic patterns. However, since only 30 of the 88 original stimuli were identified by the judges with satisfactory accuracy (see above), the complete design could not be used due to lack of appropriate stimuli. Given the rather unsatisfactory recognition accuracy for disgust, we excluded the disgust stimuli from this stage of the analysis. From the remaining 62 portrayals, 20 stimuli which had been found to be unambiguously representative of particular emotions (or neutral rendering) in judgment Studies 1 and 2 allowed a balanced ANOVA design (see below) to be constructed and were thus selected for acoustic analysis. Due to the differential recognition rate for the different actors, two actors were overrepresented in this sample. Scenario was not used as a factor in the analysis since a balanced design could not be constructed.

The digital analysis of the vocal emotion portrayals was performed with the Giessen Speech Analysis System (GISYS, based on a PDP 11/23 minicomputer) for parameter extraction (see Standke, 1991, for a detailed description of the parameters and the extraction algorithms).

Results

We will use two approaches to discuss the results of the acoustic analyses. First, we approach the data with the goal of description, in particular with respect to emotion and gender differences, by presenting the means for the acoustic parameters studied in this research. For these variables, 5 × 2 × 2 ANOVAS (Emotion × Gender × Sentence) were computed (since only one observation per cell was available, the three-way interaction was used for error estimation; see Winer, 1962, p. 272). Since main effects for sentence and interaction effects involving sentence reached significance only in a very small number of cases (not exceeding what one would expect by chance), these effects will not be discussed. It is hoped that the report of these data can provide the basis for replication in further studies in this area.

Second, we compare our findings with the predictions proposed by Scherer (1986) for some of the vocal effects of the emotions studied. We do not venture to consider this a "test" of the predictions, since these were made for the case of the vocal expression of real, naturally occurring emotions. In the present study, only actor portrayals were studied. Yet, we propose to examine the correspondence between predictions and our data. This procedure is based on the assumption that actors use implicit knowledge about the acoustic concomitant of particular emotion expressions (or reproduce states equivalent to natural emotions). Since possible gender differences had not been taken into account in the original predictions, we proceeded in the following manner to check on this correspondence: The speaker identity (which included the gender difference) was dummy-coded into four variables (corresponding to the four different speakers; using 0–1 coding). These dummy variables were regressed onto the acoustic parameters, and the residuals were used for a one-way ANOVA with contrast coding. Contrasts were constructed as shown in Table III, comparing differential emotion portrayal effects with neutral rendering, corresponding to the predictions made in Scherer (1986, Table VI). In addition to the t value, the probability, and the effect size (in standard deviation units, d) are reported for each variable. It should be noted that we did not assume equal variances of the variables across emotions and thus corrected the degrees of freedom for the t tests according to the formula proposed by Welch (Winer, 1962, p. 37). As a result, the df values are not necessarily integers.

Rate of Articulation

We decided to use length (duration) of voiced segments in the standard sentences as a measure of rate of articulation (tempo of speech). This index is more sensitive to differences in speed than the total duration of articulation (voiced and voiceless segments) or total length of the utterance (including pauses), since unvoiced segments and pauses in very short standardized phrases have generally little effect on overall speech rate.

The means are shown in Table IV. The ANOVA yielded significant main effects for emotion [$F(4, 4) = 7.085, p = .037$] and gender [$F(1, 4) = 12.849, p = .023$]. No other effect was significant. This gender difference (i.e., men speaking generally faster) confirms the limited evidence compiled by Hall (1984). However, since our sample of actors was quite small, we do not want to overinterpret this difference.

It was predicted that rate of articulation should be much higher than neutral in fear, higher than neutral in anger and joy, and lower in sadness. A one-way ANOVA of the residuals (see above), with the planned comparison contrasts specified in Table III, yielded a significant effect ($t = 3.32, df = 10.9, p = .007, d = 2.01$). This supports the hypothesis concerning the difference between sadness and the other emotions. However, contrary to expectation, the rate for fear is somewhat lower than for anger and joy.

It may seem odd that the neutral sentences had even higher means than sadness portrayals. Auditory inspection of the tapes shows that this is probably due to the fact that the actors spoke the neutral sentences with exaggerated slowness and clearness, which is hardly representative of the way that these sentences would be normally spoken in everyday speech situations characterized by low affective tone. In this sense the neutral utterances in this study may not represent a good reference level.

Table III. Predictions of Vocal Emotion Characteristics (Based on Scher, 1986) and the Corresponding Lambda Weights Used for Contrast Analysis in This Study[a]

	Fear	Joy	Sadness	Anger	Neutral
F0	++	+	+-	+-	0
	2	1	0	0	-3
F0 variance	+	+	-	+	0
	1	.5	-2	1	-.5
Intensity mean	+	+	- -	+	0
	.5	.5	-2	1	0
Intensity variance	+	+	-	+	0
	1	1	-3	1	0
High-frequency energy (R)	++	+-	+-	++	0
	-1	0	0	-1	2
Speech rate (R)	++	+	-	+	0
	-1	-.5	2	-.5	0

[a] *Note.* + = increase, - = decrease compared to neutral utterances; double symbols of the same type indicate the strength of the change, double symbols with opposite direction signs indicate that both increase and decrease are possible. (R) indicates that lambda weights were reversed, since the concept for which a prediction is made is inversely related to the parameter used for its measurement.

Intensity

Intensity is a difficult variable to measure since it depends highly on the distance and direction of the speaker's mouth to the microphone, the gain setting of the tape recorder, the equipment used, etc. However, since in the present study these variables were kept constant over speakers and emotions (because of the professional studio recording conditions), comparison between stimuli is possible. In this study, we determined the energy of the digitized speech samples as measured in decibels. Both the mean and the variance of this measure of the intensity are reported.

The means are shown in Table IV. The overall three-factor ANOVA for mean intensity yielded neither significant main effects nor interactions. It had been predicted that mean intensity would increase strongly for anger, increase moderately for joy and fear, and decrease for sadness. The one-way ANOVA of the residuals with the planned-comparison contrasts shown in Table III is significant ($t = 4.6$, $df = 8.3$, $p = .002$, $d = 3.2$). While this result supports the hypothesis that sadness differs from the other three emotions, the empirically observed means are not in accordance with the assumption that mean intensity increases more strongly for anger than for fear and joy (in the present sample of utterances joy had the highest mean intensity).

The overall ANOVA for intensity variance yielded an F value that can be considered as representing a trend [$F(4, 4) = 4.93, p = .076$; means shown in Table IV]. Scherer (1986) had predicted that intensity variance should increase for anger, joy, and fear, and decrease for sadness. A one-way ANOVA of the residuals with the appropriate contrasts (see Table III) revealed a significant effect ($t = 5.8$, $df = 7.8$, $p < .001$, $d = 4.15$).

Fundamental Frequency

The fundamental frequency of the voice (F0) represents the frequency of the vibration of the vocal folds during phonation (our auditory impression of voice pitch is most strongly determined by F0). The physiological mechanism expected to operate in changing F0 for these emotions is described in detail in Scherer (1986). Both mean fundamental frequency (F0) and low F0 quartile, i.e., the F0 value below which fall the lowest 25% of the F0 values extracted, were computed for each utterance. Although these variables are highly correlated ($r = .95$), both are reported here since mean F0 in short sentences is highly dependent on the accent structure of the sentence and may thus be less indicative of the physiological F0 setting than low F0 quartile which is closer to F0 floor (the physiologically determined baseline for a speaker's F0 level, defined as the lowest 5% of the F0 values extracted; see Scherer, Ladd, & Silverman, 1984).

The means for both variables are shown in Table IV. As expected, the overall ANOVA shows strong gender difference main effects for both F0 variables [F0 mean: $F(1, 4) = 22.8, p = .009$; F0 low quartile: $F(1, 4) = 131.2, p < .001$]. This is evidently due to the fact that the phonation structures are anatomically different in females (e.g., the vocal cords are generally shorter), producing higher F0 in comparison to adult males (which makes F0 one of the major gender markers, see Smith, 1979). F0 values are often logarithmically transformed (e.g., into semitones) to remove the male–female F0 differences. We did not do this in the present study, because we used gender as one of the factors in the three-factor ANOVA and because gender is partialed out in the analysis of the predicted contrasts.

In terms of emotion main effects we found the following results: F0 mean [$F(4, 4) = 6.0, p = .055$], low F0 quartile [$F(4, 4) = 22.35, p = .005$]. The Gender × Emotion interaction effect is significant for low F0 quartile [$F(1, 4) = 11.99, p = .017$] but not for F0 mean [$F(4, 4) = 3.1, p = .148$].

Table IV. Acoustic Parameter Means for Four Emotions and Neutral

Acoustic parameters		Fear	Joy	Sadness	Anger	Neutral
Duration of voiced segments, sec						
	Male	0.84	0.81	0.94	0.94	1.21
	Female	1.13	1.08	1.23	0.86	1.44
Intensity, dB						
Mean	Male	16.3	13.5	9.7	14.8	11.6
	Female	16.5	16.9	10.7	15.8	11.7
Variance	Male	38.5	31.5	23.2	36.3	27.2
	Female	38.8	38.8	25.9	35.6	26.5
Fundamental frequency, Hz						
Mean	Male	155.1	159.0	116.1	176.7	124.1
	Female	273.5	265.3	156.6	169.8	175.2
Low quartile	Male	131.9	125.7	111.1	135.2	102.9
	Female	257.0	206.7	140.2	158.7	154.2
Variance	Male	1821	2659	427	3773	2118
	Female	2305	5138	1074	1167	1806
Spectral energy distribution						
Low-frequency energy	Male	95.8	75.4	96.4	56.2	74.5
% below 635 Hz	Female	95.0	84.0	95.8	81.4	95.4
Spectral slope	Male	-5.27	-5.02	-5.87	-4.29	-6.08
dB/octave	Female	-5.48	-5.48	-5.97	-5.0	-6.56

This interaction effect may be due to the fact that the differences between the portrayals are much more pronounced for the female than for the male actors. Again, we feel that our sampling of actors of the two genders is too limited to speculate about possible explanations.

The senior author predicted a sharp increase of F0 mean for fear and a substantial increase for joy. No clear predictions were made for sadness and anger since F0 mean could be expected to increase or decrease depending on the relative strength of different determinants pushing phonation parameters in opposite directions. In the paper by Scherer (1986) no explicit prediction had been made for low F0 quartile (or F0 floor). However, it can be argued that these variables are even better indicators of the effect of physiological emotion effects on vocal fold vibration than F0 mean (see above) and that the predictions can be equally applied to these variables. In the one-way ANOVA on the residuals the theoretically predicted contrasts yielded the following results: F0 mean – $t = 3.75$, $df = 8, p = .006, d = 2.65$; low F0 quartile – $t = 3.15$, df $= 6.2, p = .019$, $d = 2.53$.

With respect to F0 variance we find the following effects: Emotion $F(4, 4) = 2.13, p = .24$; gender, $F(1, 4) < 1$; Emotion × Gender, $F(4,4) = 1.4, p = .37$.

It had been predicted that fear and anger would be characterized by strong, and joy by medium, increases in F0 variance, whereas a decrease was expected for sadness. This hypothesis was examined by testing for a specific contrast in a one-way ANOVA of the residuals. The result supports the hypothesis: $t = 3.71$, df $= 12.5, p = .003, d = 2.10$. While the planned comparison supported the overall prediction, joy, contrary to expectation, showed an even higher increase than fear or anger.

It would seem, then, that the predictions concerning the fundamental frequency of the voice, F0, the acoustic variable that seems to represent one of the most important vocal emotion indicators (see Scherer, 1979, 1989), were rather well supported by the data in this study. Since the actors in this study manipulated their F0 levels depending on the type of emotion portrayed, we have yet another piece of evidence suggesting that F0 plays a significant role in vocal emotion communication.

Spectral Energy Distribution

In the following, only a rather cursory introduction concerning this complex set of variables can be provided. The reader is referred to Scherer (1982, 1989) for further detail. The spectrum of a complex sound consists of a two-dimensional display of the distribution of the energy partials over the total frequency range. In voice analysis one usually works with a long-term spectrum which consists of the average of the individual spectra for all short-term (usually between 20 and 50 msec) speech segments, analyzed over a lengthy utterance. In every long-term spectrum the amplitude of the harmonics decreases from the low- toward the high-frequency range of the spectrum. This dropoff in energy amounts to about 6 db per octave in normal speech (Frokjaer-Jensen & Prytz, 1976; Nolan, 1983). The stronger this dropoff, the lower the proportion of energy found in the upper frequency range, relative to the lower one.

Figure 1 shows spectra for the vowel [a] as spoken by one of the actors in this study, segmented out of one of the two standardized utterances. Both a three-dimensional display of spectra for successive 20-msec frames over the duration of the vowel and the average spectrum over those frames (including the smoothed spectrum) are shown for the neutral utterance and for three of the four emotions studied here. This figure is intended as an illustration of the spectrum and of the differences (in the case of one single speaker and for one single phoneme) between emotions. For the statistical analyses, we used the average spectrum over the complete utterance. It should be noted that the brief utterance we used (the mean of the articulation time is 1.42 sec) would not be sufficiently long to obtain a stable *long-term spectrum* (which requires at least several minutes). However, since each utterance consisted of standardized sequences of phonemes the average spectra should have been directly comparable for each of the sentences individually.

While Fig. 1 illustrates the differences in energy distribution in the frequency domain between neutral and three emotions, the complete spectrum (which usually consists of 256 or 512 numbers) is too unwieldy to be used in univariate statistical analysis (even with a reduced set of values for third-octave bands). One hesitates to use multivariate procedures at the present stage of our knowledge, since the dependencies between the energy values that constitute the spectrum are not well understood. Therefore, a number of indices or proportions that provide a rough indication of the distribution of energy between the low- and high-frequency range has been proposed. It is possible to use the steepness of the energy decrease from low to high frequencies in the spectrum as a measure by computing the slope of the regression line of the spectral values above the F0 peak. We extracted this variable, which will be called "spectral slope," in what follows. Another measure frequently suggested in the literature is based on the relative proportion of total energy above and below a certain cutoff frequency. We computed the percentage of the total energy of the long-term spectrum found below 635 Hz. This cutoff level (which represents the upper limit of the fifth third-octave band) was chosen because it seems to represent a

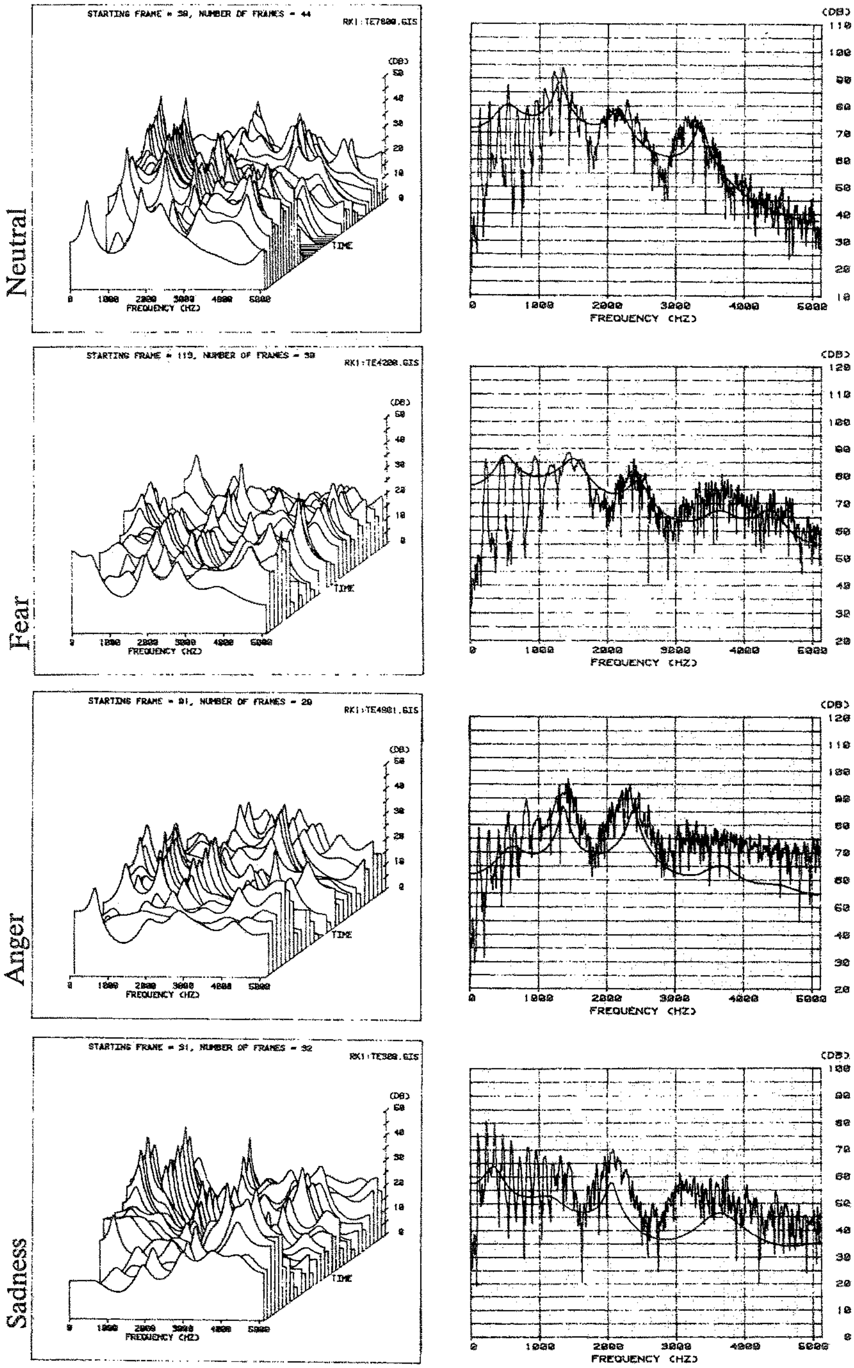

Fig. 1. Three-dimensional spectra (left) and average power spectra (right) for a vowel from one of the standard sentences as spoken by one of the male actors in the neutral and three emotion portrayals.

reasonable compromise between the values of 500 and 1000 Hz, both of which have been suggested in the literature (see Scherer, 1989). This variable is called "low-frequency energy."

We find a significant gender main effect [$F(1, 4) = 25.92, p < .007$] for low-frequency energy, and a significant Emotion × Gender interaction effect [$F(4, 4) = 6.57, p = .048$]. For joy, anger, and neutral the actresses show a greater percentage of energy in the low-frequency range of the spectrum than the actors. There is no significant gender or interaction effect for spectral slope.

Turning to emotion main effects, the F for spectral slope just misses the .05 level of significance [$F(4, 4) = 6.148, p = .053$], whereas the F for low-frequency energy is significant [$F(4, 4) = 23.78, p = .005$].

Scherer (1986) has predicted an increase of high-frequency energy (without specifying a concrete operationalization) for anger and fear, with variable patterns for joy and sadness. The one-way ANOVA of the residuals with planned comparison contrasts (see Table III), yields the following results: spectral slope – $t = 4.15, df = 6.1, p = .006, d = 3.36$; low-frequency energy – $t < 1$, ns.

In spite of partial support for the predictions, the pattern of empirically observed means suggests that the theoretical suggestions need to be reconsidered. Only the anger data, for both energy distribution variables in this study, clearly confirm the expectations. The prediction that fear, like anger, is characterized by a very high proportion of high-frequency energy is not supported. While spectral slope is steeper for fear than for sadness and neutral, as expected, it does not differ greatly from joy. With respect to low-frequency energy, the fear results depart even further from the predicted pattern (and come even closer to those for sadness). Joy seems to involve a stronger change in the direction of increased high-frequency energy than had been envisaged in the predictions (in which the influence of opposing tendencies had been surmised).

The central assumption underlying the predictions for spectral energy distribution had been (Scherer, 1979, 1986) that emotions involving strong arousal of the ergotropic system (corresponding to sympathetic arousal; see Gellhorn, 1970) would likely be characterized by more energy in the higher-frequency range. This pattern was indeed found for anger and joy. Since fear is generally considered as an emotion characterized by high ergotropic arousal, we expected the fear results to be closer to those of anger than to those of sadness (which could be considered more strongly determined by trophotropic predominance). If the results reported here are confirmed in future studies, these assumptions about the effects of the psychological changes accompanying fear on phonation characteristics will have to be revised.

Intercorrelations Between Acoustic Parameters

After having presented the results for the individual variables it seems useful to examine the intercorrelations between the acoustic parameters that were extracted in this study in order to be aware of potential overlap. The intercorrelation matrix is shown in Table V. The two measures of F0 level correlated very highly in this study. However, we decided to report the results for both variables, since F0 mean is often used in the literature, even though a measure approximating F0 floor might be more appropriate for the short sentences studied here (see above). The sizable correlations between F0 and intensity were to be expected on the basis of acoustic-phonetic descriptions of the phonation mechanism (e.g., subglottal pressure affecting both variables). Yet, since only about a quarter of the variance is shared, each variable obviously contributed independent information on speaker state. The near unity correlation between intensity mean and intensity variance is due to the nature of the variable: Since the lower end of the distribution of the values is always near zero intensity, any change in mean intensity will automatically result in a change of the variance. Thus, intensity variation does not furnish a good estimate of intensity variability and future studies will need to identify a more appropriate measure of variability. A negative correlation between the two special measures, low-frequency energy and spectral slope, is to be expected, since the stronger the harmonics in the low-frequency range, the higher (i.e., high negative value) should be the spectral slope in the spectrum. The remaining correlations do not require special discussion.

CONCLUSIONS

We interpret the results reported in this paper as showing that actor portrayals of emotion can be profitably used to further our understanding of both encoding and decoding of emotions using the vocal–auditory modality of communication. However, the results also show that this research paradigm and the associated methodology is in urgent need of refinement. Although in the present study, contrary to earlier approaches, a systematic design for the selection of the actors, the emotion scenarios, and the speech material was used, a number of problems remain. The most important concern is the capacity of different actors actually to convey emotional meaning in a highly differentiated manner. As in an earlier study (Wallbott & Scherer, 1986b) we are forced to conclude that, in spite of the fact that very professional and highly paid radio actors with many years of experience in different types of acting activities were used, there were sizable differences in the quality of the portrayals obtained. This is indicated by the fact that the final selection

Table V. Intercorrelations of Acoustic Parameters

	Mean F0	Low F0 quartile	F0 variance	Mean intensity	Intensity variance	Articulation rate	Low-frequency energy	Spectral slope
Mean F0	1.00	.95[a]	.53[b]	.57[b]	.63[b]	.15	−.02	.38
Low F0 quartile		1.00	.26	.55[b]	.59[b]	.12	.12	.25
F0 variance			1.00	.22	.29	.25	−.40	.52[b]
Mean intensity				1.00	.97[a]	−.31	−.17	.47
Intensity variance					1.00	−.31	−.23	.57[b]
Articulation rate						1.00	.17	−.45
Low-frequency energy							1.00	−.50
Spectral slope								1.00

[a] $p < .01$.
[b] $p < .05$.

of items to be used in the judgment studies was heavily biased toward some of the actors. Thus, studies in which a single actor or even several actors are used without subsequently assessing the degree to which the portrayals are consensually judged may not be very informative concerning either the expression patterns used for emotional communication or the ability of judges to infer emotions from expressive behavior. It would seem then, that a rather large number of actors has to be recruited and that extensive judgment studies need to be performed in order to select those portrayals that are correctly decoded. Furthermore, given the rather strong differences between male and female emotion portrayals found in this study, it would seem of great importance to include a reasonable number of both male and female actors in future studies.

This study, as several others before (see Scherer, 1981, 1986, and Siegman, 1987, for reviews), has confirmed that judges are able to identify emotions from vocal portrayals with much better than chance accuracy. So far, there has been little systematic work on any possible differences among different populations of judges, or the relative ability of the judges to identify particular emotions, nor has there been study of the factors responsible for such differences. The differences in accuracy between a broad cross section of the general population and a group of students reported here show that a generalization of the results obtained with students to the general population may not be appropriate.

Finally, the results of the acoustic analyses show the potential of this approach to test systematically the detailed predictions on vocal emotion expression which have been advanced by the senior author (Scherer, 1986). While a large number of these predictions have been supported, some of the results point to the need to revise the hypothetical links between emotion-specific physiological arousal and acoustic characteristics of vocal expression. Since these predictions are based on physiological mechanisms, the present approach represents an important first step in determining the extent to which acoustic patterns of naturally occurring expressions of emotions might be reflected in actor portrayals of emotion. Since actors use culturally shared strategies to simulate real emotions, such findings may help us to assess the extent to which conventional communication uses markers that are based on vocal correlates of the physiological concomitants in naturally occurring emotion processes.

A fair number of gender differences and Emotion × Gender interaction effects were found this study. Since only two actors of each gender were studied and since the individual differences in acting style were very pronounced, these results need to be replicated with larger samples of actors of each gender. There can be no doubt, in the light of the present results, that future studies need to pay much more attention to potential gender differences, requiring much more systematic sampling of encoders than has been the case to date. It may also be possible to develop a number of hypotheses based on emerging knowledge about gender differences in emotion. For example, it may be that women express anger differentially than men do, possibly because of social control or cultural display rules or role socialization (see Wallbott, 1988). It would be interesting to determine whether there are different types of stereotyped expressions of anger in the two genders and whether the female *experience* of anger differs from the corresponding male experience (see Scherer, Wallbott, & Summerfield, 1986).

APPENDIX

The following scenarios were used in this research (translated from the German originals):

Fear

1) While I am on a tour bus, the driver loses control of the bus while trying to avoid another car. The bus comes to a standstill at the edge of a precipice, threatening to go over.

2) I am going to an audition which will be decisive for getting the leading role.

Disgust

1) I have to clean up the vomit of a party guest who threw up in my apartment.

2) I have a summer job in a restaurant. Today I have to clean the toilets which are incredibly filthy and smell very strongly.

Joy

1) I am acting in a new play. From the start, I get along extremely well with my colleagues who even throw a party for me.

2) My uncle arrives for my birthday. He asks me to look out of the window. In front of the house is his gift: the car I had been dreaming of for a long time.

Sadness

1) I have to give my beloved pet a strangers, because we are moving and our new landlord doesn't allow pets.

2) I get a call to tell me that my favorite aunt suddenly died.

Anger

1) The director is again late for rehearsal and we have to work until very late at night. Once again I have to cancel a date.

2) I have sublet my apartment for a period of several months. Upon my return my place is in a real mess and the person did not keep to a single agreement.

REFERENCES

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

Ekman, P., Friesen, W. V., & O'Sullivan, M. (1988). Smiles when lying. *Journal of Personality and Social Psychology, 54,* 414-420.

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

Fleiss, L. J. (1981). *Statistical methods for rates and proportions.* New York: Wiley.

Frokjaer-Jensen, B., & Prytz, S. (1976). Registration of voice quality. *Bruel and Kjaer Technical Review, 3,* 3-17.

Gellhorn, E. (1970). The emotions and the ergotropic and trophotropic systems. *Psychologische Forschung, 34,* 48-94.

Goffman, E. (1979). Response cries. In M. von Cranach, K. Foppa, W. Lepenies, & D. Ploog (Eds.), *Human ethology* (pp. 203-240). Cambridge, England: Cambridge University Press.

Hall, J. A. (1984). *Nonverbal sex differences: Communication accuracy and expressive styles.* Baltimore: Johns Hopkins University Press.

Nolan, F. (1983). *The phonetic bases of speaker recognition.* Cambridge, England: Cambridge University Press.

Rosenthal, R., & Rubin, D. B. (1989). Effect size estimation for one-sample multiple-choice-type data: Design, analysis, and meta-analysis. *Psychological Bulletin, 106,* 332-337.

Rozin, P., & Fallon, A. (1987). A perspective on disgust. *Psychological Review, 94,* 23-41.

Scherer, K. R. (1977). Affektlaute und vokale Embleme. In R. Posner & H. P. Reinecke (Eds.), *Zeichenprozesse — Semiotische Forschung in den Einzelwissenschaften* (pp. 199-214). Wiesbaden: Athenaion.

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). Speech and emotional states. In J. Darby (Ed.), *Speech evaluation in psychiatry* (pp. 189-220). New York: Grune and Stratton.

Scherer, K. R. (1982). Methods of research on vocal communication: Paradigms and parameters. In K. R. Scherer, & P. Ekman (Eds.), *Handbook of methods in nonverbal behavior research* (pp. 136-198). Cambridge, England: Cambridge University Press.

Scherer, K. R. (1984). On the nature and function of emotion: A component process approach. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 293-317). Hillsdale, NJ: Erlbaum.

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin, 99,* 143-165.
Scherer, K. R. (1989). Vocal correlates of emotion. In H. Wagner & A. Manstead (Eds.), *Handbook of psychophysiology: Emotion and social behavior.* (pp. 165-197). London: Wiley.
Scherer, K. R., Ladd, D. R., & Silverman, K. E. A. (1984). Vocal cues to speaker affect: Testing two models. *Journal of the Acoustical Society of America, 76,* 1346-1356.
Scherer, K. R., Wallbott, H. G., & Summerfield, A. B. (Eds.). (1986). *Experiencing emotion: A cross-cultural study.* Cambridge, England: Cambridge University Press.
Siegmann, A. W. (1987). The telltale voice: Nonverbal messages of verbal communication. In A. W. Siegman & S. Feldstein (Eds.), *Nonverbal behavior and communication.* (2nd ed., pp. 351-434). Hillsdale, NJ: Erlbaum.
Smith, P. M. (1979). Sex markers in speech. In K. R. Scherer & H. Giles (Eds.), *Social markers in speech.* Cambridge, England: Cambridge University Press.
Standke, R. (1991). *Methoden der digitalen Sprachverarbeitung in der vokalen Kommunikationsforschung.* Unpublished doctoral dissertation, University of Giessen.
Trojan, F. (1975). *Biophonetik.* Zürich: Bibliographisches Institut.
Van Bezooijen, R. (1984). *The characteristics and recognizability of vocal expression of emotions.* Dordrecht, The Netherlands: Foris.
Wagner, H. L. (1990). The spontaneous facial expression of differential positive and negative emotions. *Motivation and Emotion, 14,* 27-43.
Wallbott, H. G. (1988). Big girls don't frown, big boys don't cry—Gender differences of professional actors in communicating emotion via facial expression. *Journal of Nonverbal Behavior, 12,* 98-106.
Wallbott, H. G., & Scherer, K. R. (1986a). How universal and specific is emotional experience? Evidence from 27 countries on five continents. *Social Science Information, 25,* 763-796.
Wallbott, H. G., & Scherer, K. R. (1986b). Cues and channels in emotion recognition. *Journal of Personality and Social Psychology, 51,* 690-699.
Williams, C. E., & Stevens, K. N. (1969). On determining the emotional state of pilots in flight: An exploratory study. *Aerospace Medicine, 40,* 1369-1372.
Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: Some acoustical correlates. *Journal of the Acoustical Society of America, 52,* 1238-1250.
Winer, B. J. (1962). *Statistical principles in experimental design.* New York: McGraw-Hill.
Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In L. Berkowitz (Ed.), *Advances of experimental social psychology,* (Vol. 14, pp. 1-59). New York: Academic Press.