Cue utilization in emotion attribution from auditory stimuli

Scherer, Klaus R.; Oshinsky, James S.

How to cite

SCHERER, Klaus R., OSHINSKY, James S. Cue utilization in emotion attribution from auditory stimuli. In: Motivation and Emotion, 1977, vol. 1, n° 4, p. 331–346. doi: 10.1007/BF00992539

This publication URL: https://archive-ouverte.unige.ch/unige:101856

Publication DOI: 10.1007/BF00992539

Motivation and Emotion, Volume 1, Number 4, 1977

Cue Utilization in Emotion Attribution from Auditory Stimuli[1]

Klaus R. Scherer[2]
University of Giessen

James S. Oshinsky
University of Tennessee

Electronically synthesized tone sequences with systematic manipulation of amplitude and pitch variation, pitch level and contour, tempo, envelope, and filtration were rated on emotional expressiveness. The results show that two-thirds to three-quarters of the variance in the emotion attributions can be explained by the manipulation of the acoustic cues, and that a linear model of the judges' cue utilization seems to be a good approximation to their response system. Implications for phylogenetic and ontogenetic aspects of the vocal expression of emotion and for the psychology of music are discussed.

Although the vocal expression of emotion has received little attention in comparison with other modalities of human nonverbal communication, there is evidence from both encoding and decoding studies that there are a number of voice and speech cues which reliably characterize the vocal expression of particular emotional states (cf. review by Scherer, in press) and that human observers are able to infer the emotional state of a speaker with a fair degree of accuracy on the basis of such vocal cues (cf. Kramer, 1963; Davitz, 1964; Scherer, 1970). In Brunswikian terms, both the ecological validity of certain acoustic cues and the functional validity of observers' attributions of emotional states based on such cues (Brunswik, 1956) seem to be established. Yet there has been very little work on the nature of the

[1]This paper is based on research supported by NIMH grant MH 19-569-01 to the first author while at the University of Pennsylvania. The authors gratefully acknowledge comments and suggestions by Paul Ekman, Ursula Scherer, and two anonymous reviewers.

[2]Requests for reprints should be addressed to Dr. Scherer at Fachbereich Psychologie, Universität Giessen, Otto Behaghelstrasse 10/F, 6300 Giessen, West Germany.

cognitive inference processes that reflect the utilization of these cues in the attribution of emotion, using systematic experimental manipulations of the cues and cue combinations to which judges are exposed. The cognitive inference rules resulting from studies of this kind can be validated by predicting the outcome of actual attribution processes on the basis of the input to the inference system. In this fashion, Scherer (1972, 1976) has been able to predict personality attributions based on samples of voice quality by using independently assessed voice-personality inference rules.

Although there has been a lot of experimental research on cue utilization in information processing (cf. Rappoport & Summers, 1973), most of these studies use verbal labels or quantitative indices such as test scores rather than nonverbal cues. The experimental study of cue utilization in the attribution of personality and emotion from appearance and behavior is particularly difficult since, as Brunswik has shown, the relevant cues may not occur in isolation or may not be readily reproducible and/or combinable in a representative design (Brunswik, 1956). In an early study, Brunswik and Reiter (1937) used schematized faces to vary systematically a number of physiognomic features (mouth height, forehead height, eye separation, nose length) to study the utilization of these stimulus patterns by the responding system. In a series of studies of this kind the predictive power of a variety of physiognomic cues and cue combinations for personality attribution was established (cf. Rohracher, 1975, pp. 129-135).

Just as the physiognomy of the face can be approximated by line drawings, sequences of speech sounds can be approximated by sequences of pure sound wave bursts. The acoustic characteristics of such tone sequences relevant for the vocal expression of emotion such as loudness, pitch, timbre, as well as their range and variation, can be easily manipulated in a factorial design. Since an earlier study had shown that a minimal set of acoustic cues may be sufficient to communicate the evaluation, potency, and activity dimensions of emotional meaning (Scherer, Koivumaki, & Rosenthal, 1971), Scherer conducted a pilot study in which level and variation of both pitch and amplitude as well as tempo of synthesized tone sequences were systematically varied, and naive judges were asked to rate the emotional expressiveness of these stimuli (Scherer, 1974). The results showed highly consistent emotion attribution based on specific cues and cue configurations, providing a first indication of the nature of cue utilization in cognitive inferences of emotion from auditory cues.

The present study was designed to replicate these results in a more comprehensive design and to extend the scope of the investigation by including acoustic parameters relevant to the psychology of music (cf. Farnsworth, 1958; Lundin, 1953; Meyer, 1956). More specifically, the "impression values" of a number of acoustic cues and cue configurations in inferring emotional meaning from auditory information were obtained via subjective ratings of the emotional expressiveness of synthesized tone sequences with systematically manipulated acoustic parameters. An attempt is made to assess whether a linear model adequately describes the judgmental strategy of the raters, i.e., whether the emotion attributions can be considered to be based on linear combinations of the available cues (cf. Slovic & Lichtenstein, 1973, pp. 23-24).

METHOD

Design

Untrained raters were exposed to three series or types of synthesized tone sequences consisting of eight sawtooth wave bursts (tones). For these tone sequences major acoustic parameters were systematically manipulated by appropriate settings on a MOOG synthesizer.[3]

There were 128 Type I stimuli, resulting from the independent variation of seven two-level factors (*amplitude variation*—small and large; *pitch level*[4]—high and low; *pitch contour*—up and down; *pitch variation*—small and large; *tempo*—slow and fast; *envelope*—low attack/decay ratio, and equal attack/decay ratio;[5] *filtration cut-off level*—intermediate and high)[6], in a 2^7 factorial design.

Type II stimuli consisted of four of the 128 Type I tone sequences which underwent additional manipulations. Two three-level factors (*filtration level*—low, intermediate, and high; *tonality*—major, minor, and atonal) were varied independently in a 4 X 3 X 3 factorial design, yielding 36 tone sequences.

Type III stimuli consisted of manipulations of a Beethoven melody (*filtration level*—low, intermediate, and high; *tonality*—major and minor; *rhythm*—even and uneven; tempo—slow and fast), generating 24 tone sequences in a 3 X 2 X 2 X 2 factorial design.

[3]A detailed technical description of the procedure and the synthesizer settings used for the different parameters can be obtained by writing to the authors.

[4]It would be more correct to use "fundamental frequency" or "F_0" rather than pitch to refer to the physical parameter that was manipulated. However, pitch, the term for the perceptual representation of F_0, has been used throughout the paper for easier reading.

[5]Attack refers to the amount of time taken by an auditory signal to reach its maximum amplitude. Decay refers to the amount of time taken to drop from maximum amplitude to zero.

[6]Filtration refers to the low-pass filtering of the acoustic signal. The intermediate filter cutoff level passed the first three, the high level the first eight harmonics. In the "low" setting only the fundamental is passed.

The resulting 188 stimuli were divided into four sets, each set containing 32 stimuli from Type I, 9 from Type II, and 6 from Type III in randomized orders. The four stimulus sets were presented to four rating groups, each group hearing one and only one stimulus set. To test intrarater agreement and the effects of fatigue during the rating sessions, two sequences from the beginning of each group's set were repeated at the end of each rating session. In addition, the repeated sequences from each rating group's set were also inserted into the set orders of the other three rating groups (between series positions 15 and 16), to test intergroup agreement. Thus, eight sequences out of 188 were heard by all subjects. In all, each subject rated 55 sequences.

Each subject rated each tone sequence on three 10-point semantic differential scales with the following polar labels: pleasantness-unpleasantness, activity-passivity, and potency-weakness. Each subject also used a dichotomous 0-1 scale to indicate whether each sequence could (1) or could not (0) be an expression of any or all of the following emotions: anger, fear, boredom, surprise, happiness, sadness, and disgust. These emotion labels were chosen because of their frequent usage in studies on both facial and vocal expression of emotion (cf. Ekman, 1972; Scherer, in press).

Subjects

Subjects were 48 University of Pennsylvania undergraduates (25 male, 23 female), who were recruited from undergraduate psychology courses. Participation in the experiment was voluntary, and all subjects were paid for participating. Group assignment was determined by sign-up order. There were 13 subjects in group 1 (7 male, 6 female), 13 subjects in group 2 (8 male, 5 female), 12 subjects in group 3 (5 male, 7 female), and 10 subjects in group 4 (5 male, 5 female). In all groups, unequal *N*'s were due to absenteeism. Four rating sessions, one for each group, were conducted.

Procedure

Subjects had been told that the purpose of the research was to study emotional expression in speech and music and that their task would be to judge the kind of emotion that each of the synthesized tone sequences they were going to hear expressed. At the beginning of the session each subject was given an instruction booklet containing the rating scales and a list of the stimulus code numbers. Two practice sequences were played, and the experimenter answered questions. After hearing each new sequence for the first time, subjects rated it on the first three scales (semantic differential dimensions). The sequence was repeated, and then rated for the emotions. All stimuli were played back using a Revox A 77 tape recorder, a Pioneer SX-990 amplifier, and two KLH Model 17 loudspeakers.

RESULTS

Inter-Judge Agreement

The effective reliabilities of the ratings (cf. Rosenthal, 1973) based on mean interrater correlation coefficients are presented in Table I. These data show that the level of inter-judge agreement is similar across all four rater groups. Higher effective reliabilities are found for the ratings of the activity dimension. The fact that judges seem to agree more readily on the activity level expressed by the tone sequences is consistent with earlier results by Davitz (1964), showing a more consistent relationship between the auditory cues of pitch, loudness, rate, and timbre and the subjectively rated activity level, compared to valence and strength of a simulated emotional display.

Differences between Rater Groups

Four groups of raters, each rating a fourth of the total number of 188 stimuli generated by the factorial design, had been used in this study to avoid rater fatigue. An analysis of variance (rater group by control stimuli) with four groups of judges and the eight tone sequences which were rated by all groups was computed for the three semantic differential scales to check for systematic differences between groups. The results, reproduced in Table II, did not show a main effect for groups for any of the three scales. There was a significant but not very strong group by stimuli interaction for pleasantness, indicating disagreement between groups with respect to the perceived pleasantness for some of the control stimuli. On the whole, the results on interrater agreement (high effective reliabilities) and the absence of group differences for the semantic differential scales seemed to justify a pooling of all ratings across raters and groups. All of the analyses reported subsequently were carried out with mean ratings for each of the tone sequences.[7]

[7]Since no sex differences in the processing of these stimuli had been found in a pilot study Scherer, 1974, Study II), a joint mean for rater group, averaging across male and female raters, was computed.

Table I. Effective Reliabilities of the Semantic Differential Ratings[a]

Rating scales	Rater group 1 (13 judges)	2 (13 judges)	3 (12 judges)	4 (10 judges)	Mean
Pleasantness	.84	.85	.86	.85	.85
Activity	.90	.91	.92	.93	.92
Potency	.85	.79	.83	.72	.80

[a]Based on the mean interrater correlations for $N = 55$ stimuli (cf. Rosenthal, 1973).

Test-Retest Reliability

For each rater group two tone sequences were repeated at the end of the rating sessions to test the effects of fatigue. The results of a rater group by test-retest repeated measures ANOVA are shown in Table III. Main effects for rater groups are due to the fact that for each group two different stimuli were repeated. For the activity and potency ratings there is no significant change from first to second rating. For pleasantness there is a significant main effect: The control stimuli are rated as *more pleasant* at the end (mean rating across groups 0.42 at the first rating and 1.30 at the second rating). It seems unlikely that this effect is due to boredom or fatigue since, in that case, more negative ratings would be expected. Theories on the attitudinal effects of mere exposure (cf. Zajonc, 1968) would predict that the raters evaluate the stimuli generally in a more positive way after having listened to about 50 of them.

Analysis of Type I Stimuli Effects

Multiple regression analysis was used to assess the predictive strength of the acoustic parameters, manipulated in a factorial design, for the emo-

Table II. F Values for Interrater-Group ANOVA of Ratings of Eight Control Stimuli

Between effects	Rating scales: Pleasantness	Activity	Potency
Rater groups[a]	1.51	1.37	1.78
Stimuli[b]	34.06[d]	50.45[d]	17.23[d]
Rater groups × stimuli[c]	2.48[d]	1.37	.78

[a]$df = 3/44$.
[b]$df = 7/308$.
[c]$df = 21/308$.
[d]$p < .001$.

Table III. *F* Values for Repeated-Measures ANOVA of Test-Retest Ratings of Eight Control Stimuli

	Rating scales		
Between effects	Pleasantness	Activity	Potency
Rater groups[a]	8.32[e]	.55	3.91[d]
Test-retest[b]	4.54[d]	.57	.12
Rater groups × test-retest[c]	0.66	1.28	.77

[a] $df = 3/44$.
[b] $df = 1/44$.
[c] $df = 3/44$.
[d] $p < .05$.
[e] $p < .001$.

tion attributions to the respective tone sequences. Following the procedure outlined by Cohen (1968), orthogonal weights were assigned to all main factors and all two-way and three-way interactions between factors. The dependent variables were independently regressed onto those weights by entering all main factors in fixed order and entering interactions into the regression equation on the basis of the highest partial correlation with the dependent variable for the variables not yet entered.[8] Since all factors and interactions are orthogonal, due to the factorial design, the unique variance contributed by each factor or interaction could be determined. The results, shown in Table IV, include the standardized regression coefficients for each factor or interaction (indicating predictive strength) which reached significance, significance levels, and the unique contribution to the variance due to particular effects. Only two-way interaction effects attaining the $p < .01$ significance level are given separately in the table. There are 19 two-way and 6 three-way interaction effects with beta coefficients significant at the $p < .05$ level. Since the unique variance contributed by these effects is generally below 1%, their contribution to the variance has been included in the column "Other effects," which lists the total contribution of all remaining effects entered into the regression equation. The rightmost columns indicate the multiple correlation (R^2) of the dependent variable with all factors and interactions listed in the table, the number of predictors on which R^2 is based, and the expected R^2. Highly significant R values for all dependent

[8]For disgust and surprise, all predictor variables had to be entered on the basis of the highest partial correlations of the remaining variables since the default values of the computer program would not allow the main effect factors to be entered into the regression equation in a fixed order.

Table IV. Standardized Beta Coefficients (b) and Unique Contribution to the Variance (uc) from a Regression Analysis of Type I Stimuli Ratings

	Amplitude variation (AV)		Pitch variation (PV)		Pitch contour (CO)		Pitch level (PL)		Tempo of sequence (TE)		Envelope (EN)		Filtration cutoff level (FI)		Significant interaction effects ($p < .01$)			Other effects[d]		No. of predictors and expected R^{2e}	
	b	uc	b	uc	b	uc	b	uc	b	uc	b	uc	b	uc		b	uc	uc	R^2		
Pleasant	$-.11^{a}$	.013	$.29^{c}$	.085	$-.17^{b}$	.028	$-.25^{c}$	.061	$.46^{c}$	.209	$.28^{c}$	.076	$-.38^{c}$	.141	PV/CO	$.15^{b}$	.023	.043	.679	11	.087
Activity	$-.10^{b}$	.011	$.16^{c}$	.025	.06	.004	$.23^{c}$	.052	$.83^{c}$	.684	$.15^{c}$	.023	$.20^{c}$	.041		–		.017	.857	10	.079
Potency	$-.08$	.006	.07	.004	$.09^{a}$	.009	$.28^{c}$	.078	$.35^{c}$	.119	$-.10^{a}$	.009	$.70^{c}$	.491	AV/PL	$.13^{b}$	.017	.016	.762	11	.087
															PL/FI	$.12^{b}$	.013				
Anger	$-.05$	.002	$-.23^{c}$	.053	$.12^{a}$	.015	$.26^{c}$	.070	$.27^{c}$	.071	$-.11$	.011	$.54^{c}$	.292	PV/CO	$-.24^{c}$	.057	.058	.629	12	.094
Boredom	.04	.002	$-.09^{a}$	.008	$-.23^{c}$	.053	$-.26^{c}$	.066	$-.75^{c}$	.565	$-.15^{c}$	.022	$-.25^{c}$	.061		–		.008	.785	8	.063
Disgust	$-.01$	–	$-.36^{c}$	.129	.05	.002	.07	.005	$-.17^{b}$	.028	$-.25^{c}$	.062	$.45^{c}$	.204	PV/CO	$-.29^{c}$	.085	.084	.599	11	.087
Fear	$.13^{a}$	.016	$-.13^{a}$	.018	$.33^{c}$	.112	$.23^{c}$	.052	$.32^{c}$	.103	$-.16^{a}$	.024	$.31^{c}$	.095	PV/CO	$-.18^{b}$	.031	.058	.536	12	.094
															TE/FI	$.17^{b}$	.027				
Happiness	$-.12^{a}$	.015	$.36^{c}$	.129	.07	.006	.08	.006	$.61^{c}$	.370	$.28^{c}$	.080	$-.21^{c}$	.046	PV/CO	$.20^{c}$	.040	.019	.730	11	.087
															TE/FI	$-.14^{b}$	.019				
Sadness	.08	.006	$-.06$	.003	$-.15^{c}$	.023	$-.20^{c}$	.041	$-.77^{c}$	.589	$-.18^{c}$	.032	$-.20^{c}$	.040	PL/TE	$.12^{b}$	.014	.008	.756	9	.071
Surprise	$-.02$	–	$.11^{a}$	.012	$.25^{c}$	.063	$.30^{c}$	.088	$.63^{c}$	.399	$.18^{c}$	.035	$.16^{b}$	.027	PL/TE	$.17^{c}$	.030	.022	.676	9	.071
Mean		.009		.047		.032		.052		.314		.037		.144			.036	.033	.701	10.4	.082

[a] $p < .05$.
[b] $p < .01$.
[c] $p < .001$. Degrees of freedom ranging between 1/120 and 1/118 for individual effects.
[d] Includes all interaction effects for which the t value for the regression coefficient reached $.01 < p < .05$.
[e] Expected $R^2 = k/n-1'$, k = number of predictor variables, n = number of data points (cf. Cohen & Cohen, 1975, p. 107).

Table V. Emotional Attributions Significantly Associated with Acoustic Parameters

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

variables show that about two-thirds to three-quarters of the variance in the emotion ratings can be accounted for by the manipulations of the acoustic parameters of the tone sequences. Tempo of the sequence seems to be the most powerful predictor. The variance explained is highest for those variables for which high regression weights for tempo have been found. This seems to be particularly true for emotions with extreme positions on the activity dimension, such as boredom, sadness, and happiness.

Averaged across all rating scales, tempo of the sequence explains 31% of the variance (mean unique variance), compared to 14% for filtration level and between 3 and 5% for the remaining acoustic parameters (except for amplitude variation: less than 1%). Filtration level, or number of harmonics, which is the second most powerful predictor overall, seems to be most relevant for attributions of the potency dimension. Tempo and filtration explain fairly similar portions of the variance in the pleasantness ratings.

In order to avoid tedious repetition in the text, verbal statements of the results are presented in two tables. Table V lists for each of the acoustic parameters the emotional attributions that are significantly associated with it (listed in decreasing order of predictive strength of the respective factor for this scale). This table shows which emotions are likely to be attributed to

Table VI. Acoustic Parameters of Tone Sequences Significantly Contributing to the Variance in Attributions of Emotional States

Rating scale	Single acoustic parameters (main effects) and configurations (interaction effects) listed in order of predictive strength
Pleasantness	Fast tempo, few harmonics, large pitch variation, sharp envelope, low pitch level, pitch contour down, small amplitude variation (salient configuration: large pitch variation plus pitch contour up)
Activity	Fast tempo, high pitch level, many harmonics, large pitch variation, sharp envelope, small amplitude variation
Potency	Many harmonics, fast tempo, high pitch level, round envelope, pitch contour up (salient configurations: large amplitude variation plus high pitch level, high pitch level plus many harmonics)
Anger	Many harmonics, fast tempo, high pitch level, small pitch variation, pitch contours up (salient configuration: small pitch variation plus pitch contour up)
Boredom	Slow tempo, low pitch level, few harmonics, pitch contour down, round envelope, small pitch variation
Disgust	Many harmonics, small pitch variation, round envelope, slow tempo (salient configuration: small pitch variation plus pitch contour up)
Fear	Pitch contour up, fast sequence, many harmonics, high pitch level, round envelope, small pitch variation (salient configurations: small pitch variation plus pitch contour up, fast tempo plus many harmonics)
Happiness	Fast tempo, large pitch variation, sharp envelope, few harmonics, moderate amplitude variation (salient configurations: large pitch variation plus pitch contour up, fast tempo plus few harmonics)
Sadness	Slow tempo, low pitch level, few harmonics, round envelope, pitch contour down (salient configuration: low pitch level plus slow tempo)
Surprise	Fast tempo, high pitch level, pitch contour up, sharp envelope, many harmonics, large pitch variation (salient configuration: high pitch level plus fast tempo)

different levels of each acoustic parameter with higher than chance probability.

Table VI contains the same information (plus interaction effects), but is organized by rating scales. For each semantic differential dimension and emotion attribution the acoustic characteristics significantly contributing to the variance in the attributions are listed in decreasing order of predictive strength. This table shows for each emotion which levels of which acoustic parameters contribute significantly to the likelihood that this emotion will be attributed to a tone sequence. This listing does not imply, of course, that only these combinations of parameter levels will lead to the attribution of a specific emotion. The table represents a summary of the individual statistical effects which have been found.

Analysis of Type II and Type III Stimuli Effects

Type II and Type III stimuli had been included in the present design as a preliminary attempt to extend this research approach into the area of the psychology of music by manipulating tonality and rhythm. In addition, three rather than two filtration levels were used to check for curvilinear effects. Analysis-of-variance of the results showed no evidence of the latter. As for Type I stimuli, sequences for which fewer harmonics were passed (between one and eight) were judged as significantly ($p < .05$) more pleasant, happy, and bored, whereas sequences for which more harmonics (16) were audible were judged as significantly ($p < .05$) more active, potent, angry, disgusted, and fearful.

The manipulation of tonality showed for both Type II and Type III stimuli that the major mode was seen as significantly ($p < .05$) more indicative of pleasantness and happiness, the minor mode of disgust and anger, than of other emotional states. The rhythm manipulation for Type III stimuli showed that rhythmic tone sequences were significantly ($p < .05$) more often labeled active, fearful, and surprised, whereas nonrhythmic stimuli were more often associated with boredom. Since the number of stimuli used in this pilot study was fairly small, the results need to be replicated in a more comprehensive study specifically designed for this purpose. However, the present findings show that the present research approach may be profitably used in studying the emotional impact of music.

DISCUSSION

The results reported above, which consistently replicate the earlier findings of the pilot study (Scherer, 1974), show rather strong systematic effects of the manipulation of acoustic parameters on the emotion attributions by naive judges. The regression analysis results indicate that a linear model of the judges' response system fares quite well in predicting the attributions of emotional meaning. Although there are rather large variations in the weights assigned to the individual acoustic cues in the judgmental strategy, the additive combination of main effects accounts for most of the variance that can be explained by the manipulation of the acoustic parameters.

The interaction effects found indicate that configurational relationships seem to consist mostly of particularly strong effects of combinations of cues which are also strongly weighted in isolation. However, in some instances, particular combinations of parameters may serve to differentiate attributions of subclasses of certain emotions. Thus, whereas the attribution of pleasantness is generally associated with a downward slope of the pitch contour, the combination of large pitch variation and an upward slope is also considered to indicate pleasantness. It is possible that different types of happiness, i.e., quiet, subdued bliss vs. buoyant gaiety, are communicated by specific configurations of acoustic cues. The analysis of such configurational effects would require more detailed specifications of the emotional states than were used in this study, either by providing more explicit labels or by content-analyzing freely elicited descriptions. Although an additive linear model of cue combinations can be used quite successfully in predicting the emotion attributions, it is conceivable that a curvilinear relationship exists between the range of possible values for acoustic parameters and the kind and degree of emotion inferred.

Response curvilinearity of this nature cannot be evaluated in the present study since only two levels of each acoustic parameter were utilized to keep the Type I stimuli down to a number that could be rated without producing fatigue and irritation on the part of the judges. For the Type II and Type III stimuli, three levels were used for the filtration parameter, but no clear-cut evidence for curvilinearity was found. However, there is some evidence for curvilinear effects in earlier research. Vitz (1966) demonstrated a curvilinear relationship (inverted U) between stimulus variation (pitch, loudness, and duration variation of tone sequences) and perceived pleasantness. Since his intermediate level of pitch variation (316-562 Hz) corresponds to the level of large pitch variation in this study (168-345 Hz and 296-617 Hz), only the lower portion of the curve may have been studied here. On the other hand, Vitz's extreme pitch variation extends far beyond the range of the fundamental frequency of the human voice which was used as a guideline in the present research. It is possible, then, that the downward slope of the curvilinear relationship found by Vitz may not be applicable to emotion attribution from human speech. The important issue of response curvilinearity in inferring emotion from auditory cues certainly deserves further study in experiments with a large number of levels manipulated for each acoustic cue.

The frequent references to the inference of emotion from speech in this paper may seem to be ill-advised, given the nature of the stimuli used. However, all of the acoustic cues that were manipulated in this study are major vocal cues. The amplitude, pitch, and duration parameters are directly comparable to similar acoustic parameters of speech. The envelope or the shape of the sounds used can be seen as related to speech articulation or enunciation with round envelope being comparable to drawl or slurred speech and sharp envelope to clipped speech. Filtration cut-off level is clearly related to voice timbre with few harmonics closer to resonant voice and many harmonics closer to blaring or shrill voice. It may thus not be unduly optimistic to assume that similar results will be found in studies using stimuli that more closely approximate real speech.

This optimism is supported by findings of encoding studies in which the ecological validity of a number of acoustic cues is assessed, usually on the basis of simulated emotional displays. The tabular summary of the patterns of results, reported by Scherer (in press) is largely congruent with the pattern of cue utilization shown in Table VI. It is possible, however, that since most of the encoding studies make use of simulated emotional expression, expressive behavior is determined by mimicking the inferential cue utilization rules, thus artificially increasing the apparent congruence between the encoding and decoding of emotion.

The present results also agree substantially with music research which has systematically studied the effects of tonality, melodic contour, pitch level, pitch variation, and tempo (cf. Farnsworth, 1958; Lundin, 1953). Research by Hevner (1936) found the major mode, "happy, merry, graceful, and playful," and the minor mode, "sad, dreamy, and sentimental," while "such qualities as excitement, vigor, dignity, serenity, etc., are not determined by either mode." Gundlach (1932), found that Europeans characterized high pitch level as sentimental, whimsical, animated, and glad, and low pitch level as mournful, somber, tranquil, dignified, and grotesque. Extreme pitch variation was found to be uneasy, animated, grotesque, brilliant, and glad; while moderate pitch variation was tranquil, dignified, delicate, mournful, awkward, and somber. Fast tempos were called brilliant, animated, uneasy, glad, whimsical, flippant, and grotesque; slow tempos were dignified, somber, tranquil, melancholy, mournful, delicate, and sentimental. These results bear notable resemblance to those shown in our Table V.

In the light of the agreement of the present results with both the paralinguistic and psychology of music literatures, it is not possible to determine whether the subjects in this study were responding to the tone sequences as speech or as music. The general agreement as to the effects of these cues across contexts suggests that a mechanism common to both speech and music perception is at work. Given the possible generality of these effects, it is interesting to speculate about the origin of affect recognition from acoustic cues. The recognition of emotion seems to have strong survival value since both encoding and decoding of emotion-related states are common in many animal species. Although it is questionable whether human emotions can be altogether reduced to behavior tendencies that can be universally found in most species (Plutchik, 1962), there seems to be little doubt that there is some phylogenetic continuity (cf. Strongman, 1973). Especially the strong involvement of the autonomous nervous system in producing vocalization (through respiration and phonation) suggests that many emotional vocalizations are physiologically based and phylogenetically continuous rather than learned cultural signaling patterns (cf. Scherer, in press). Moreover, the acoustic analysis of animal vocalizations

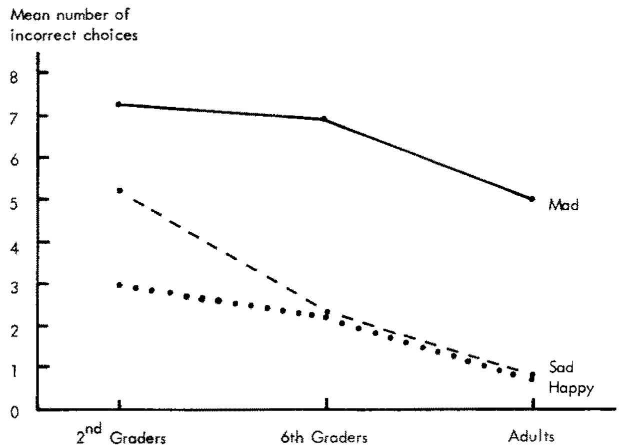

Fig. 1. Accuracy of tone sequence labeling for children and adults.

(cf. Marler, 1976; Tembrock, 1971, pp. 163-203) reveals some interesting parallels to the vocal expression of emotion in humans.

More thorough studies on the comparability of the acoustic structure of animal and human affect vocalizations as well as cross-cultural studies on the encoding and decoding of affect are called for to assess the important issue of phylogenetic continuity in this communication channel. If, similar to the mechanisms postulated for facial affect expression (Ekman, 1972), there are unlearned neuromuscular programs, modified by cultural display rules for the vocal expression of emotion in humans, it seems not unlikely that there are also unlearned auditory "feature detectors" that provide "automatic" decoding facilities. This would imply that it is not necessary to acquire cue utilization and inference rules at least for some major components of emotional expression during the course of the socialization process. Developmental decoding studies of this kind are scarce. The first author conducted a pilot study in which 10 second-graders, nine sixth-graders, and 10 adults were asked to choose between alternative labels (illustrated by stylized facial expressions) for 12 synthesized tone sequences that had been found consistently to receive particular emotion labels in the main study. The results, shown in Fig. 1, reveal both an affect effect ($F = 18.39$, $df = 2|27$, $p < .001$)—anger is more difficult to decode at all age levels—and a linear trend for age with decreasing numbers of errors for older subjects ($F = 7.46$, $df = 2|27$, $p < .01$). These results seem to indicate that there is an improvement in the mastery of inference rules related to the attribution of emotion from auditory stimuli during the course of socialization. It is possible, of course, that the synthesized tone sequences can be more easily judged after more extended exposure to music and that the differences would be less pronounced for natural speech samples. However, more comprehensive studies using sophisticated methodology need to be conducted in order to elucidate the role of maturation and learning in the decoding of affective expression.

Independent of the developmental aspect, research on the decoding of the vocal expression of affect needs to go beyond the undifferentiated approach of assessing the "accuracy" of recognition of simulated affect displays. The contribution of the present study may be seen in providing the possibility of checking emotion attributions based on natural speech samples against predictions based on acoustic analyses of these speech samples and the cue utilization rules found for synthetic stimuli.

REFERENCES

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

Brunswik, E., & Reiter, L. Eindrucks-Charaktere schematisierter Gesichter. *Zeitschrift für Psychologie,* 1937, *142,* 67-134.

Cohen, J. Multiple regression as a general data-analytic system. *Psychological Bulletin,* 1968, *70,* 426-443.

Cohen, J., & Cohen, P. *Applied multiple regression/correlation analysis for the behavioral sciences.* Hillsdale, N.J.: Lawrence Erlbaum, 1975.

Davitz, J. R. *The communication of emotional meaning.* New York: McGraw-Hill, 1964.

Ekman, P. Universal and cultural differences in facial expression of emotion. In J. Cole (Ed.), *Nebraska Symposium on Motivation, 1972.* Lincoln: University of Nebraska Press, 1972.

Farnsworth, P. *The social psychology of music.* New York: Dryden, 1958.

Gundlach, R. A quantitative analysis of Indian music. *American Journal of Psychology,* 1932, *44,* 133-145.

Hevner, K. Experimental studies of the elements of expression in music. *American Journal of Psychology,* 1936, *48,* 246-268.

Kramer, E. The judgment of personal characteristics and emotions from nonverbal properties of speech. *Psychological Bulletin,* 1963, *60,* 408-420.

Lundin, R. *An objective psychology of music.* New York: Ronald, 1953.

Marler, P. Social organization, communication and graded signals: The chimpanzee and the gorilla. In P. P. G. Bateson & R. A. Hinde (Eds.), *Growing points in ethology.* Cambridge: Cambridge University Press, 1976.

Meyer, L. *Emotion and meaning in music.* Chicago: University of Chicago Press, 1956.

Plutchik, R. *The emotions: Facts, theories, and a new model.* New York: Random House, 1962.

Rappoport, L., & Summers, D. A. (Eds.), *Human judgment and social interaction.* New York: Holt, Rinehart & Winston, 1973.

Rohracher, H. *Charakterkunde.* Wien: Urban and Schwarzenberg, 1975.

Rosenthal, R. Estimating effective reliabilities in studies that employ judges' ratings. *Journal of Clinical Psychology,* 1973, *29,* 342-345.

Scherer, K. R. *Nonverbale Kommunikation.* Hamburg: Buske (IPK-Forschungsbericht), 1970.

Scherer, K. R. Judging personality from voice: A cross-cultural approach to an old issue in interpersonal perception. *Journal of Personality,* 1972, *40,* 191-210.

Scherer, K. R. Acoustic concomitants of emotional dimensions: Judging affect from synthesized tone sequences. In S. Weitz (Ed.), *Nonverbal communication.* New York: Oxford University Press, 1974.

Scherer, K. R. *Personality inference from voice quality: The loud voice of extraversion.* Unpublished manuscript, 1976.

Scherer, K. R. Non-linguistic vocal indicators of emotion and psychopathology. In C. E. Izard (Ed.), *Emotions and psychopathology.* New York: Plenum Press, in press.

Scherer, K. R., Koivumaki, J., & Rosenthal, R. Minimal cues in the vocal communication of affect: Judging emotions from content-masked speech. *Journal of Psycholinguistic Research,* 1972, *1,* 269-285.

Slovic, P., & Lichtenstein, S. Comparison of Baysian and regression approaches to the study of information processing in judgment. In L. Rappoport & D. A. Summers (Eds.), *Human judgment and social interaction.* New York: Holt, Rinehart & Winston, 1973.

Strongman, K. T. *The psychology of emotion.* London: Wiley, 1973.

Tembrock, G. *Biokommunikation. Informationsübertragung im biologischen Bereich.* Berlin: Akademie-Verlag, 1971.

Vitz, P. C. Affect as a function of stimulus variation. *Journal of Experimental Psychology,* 1966, *71,* 74-79.

Zajonc, R. B. Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology,* 1968, *9,* 1-29.