DOCUMENT RESUME

ED 067 587 CG 007 478

AUTHOR	Scherer, Klaus R.
TITLE	Acoustic Concomitants of Emotional Dimensions: Judging Affect from Synthesized Tone Sequences.
PUB DATE	72
NOTE	8p.; Paper presented at the Eastern Psychological Association Meeting, April 27-29, 1972, Boston, Massachusetts
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Affective Behavior; Affective Tests; *Auditory Perception; Behavioral Science Research; *Behavior Patterns; Communications; *Communication Skills; Information Theory; *Psychoacoustics; Psychological Patterns

ABSTRACT

The ability of naive listener-judges to recognize the affective state of a speaker on the basis of nonlinguistic auditory cues independent of the verbal content of an utterance has been well established by a large number of studies. This study used artificial stimuli produced by a Moog synthesizer to vary pitch level and variation, amplitude level and variation, and signal duration and speed (tempo) systematically in a factorial design. The stimuli used, raters employed, procedure, and results are presented for two studies which were conducted. The results supported the contention that the attribution of emotional meaning from auditory stimuli is based on characteristic patterns of accoustic cues. This study suggested a rapproachement between studies on emotional expression in speech and the psychological investigation of emotion in music, with interesting implications concerning speculations on the common origin of music and speech in primitive emotional displays of our prehistoric ancestors. (Author/BW)

E.P.A. MEETING 1972

Acoustic concomitants of emotional dimensions: Judging affect from synthesized tone sequences

Klaus R. Scherer

University of Pennsylvania

Abstract

Electronically synthesized tone sequences with systematic variations of pitch, amplitude, and tempo were rated on emotional expressiveness. The results support the contention that dimensions of emotional meaning are communicated by specific patterns of acoustic cues. Implications concerning unlearned neural programs of emotional expression in speech and music are discussed.

Acoustic concomitants of emotional dimensions: Judging affect from synthesized tone sequences

Problem: The ability of naive listener-judges to recognize the affective state of a speaker on the basis of nonlinguistic auditory cues independent of the verbal content of an utterance has been well established by a large number of studies, summarized by Kramer (1963), Davitz (1964), Vetter (1969), and Scherer (1970). Results of a recent study by Scherer, Rosenthal, and Koivumaki (1971), using content-masking by randomsplicing (Scherer, 1971), electronic content filtering (Rogers, Scherer, and Rosenthal, 1971) and their combinations, suggest that a minimal set of vocal cues consisting of pitch level and variation, amplitude level and variation, and rate of articulation or tempo may be sufficient to communicate the evaluation, potency, and activity dimensions of emotional meaning.

In order to assess more precisely the way in which inferences of emotional content are based on specific acoustic cues and their combinations, one would want to be able to manipulate these cues experimentally. Since, in spite of recent advances in the area of speech synthesis, this is rather difficult to achieve with actual speech signals, the present study has used artificial stimuli produced by a Moog synthesizer to vary pitch level and variation, amplitude level and variation, and signal duration and speed (tempo) systematically in a factorial design.

Study I

Stimuli: A simple tone sequence modeled after the intonation contour of a short sentence, consisting of eight sine wave tones of differential pitch and duration, were synthesized repeatedly on a Moog electronic synthesizer with sequencing unit. Five parameters of the sequence were varied independently in a 4x2x2x2x2 factorial design with the following levels on each parameter: pitch variation - moderate, extreme, up contour, down contour; amplitude variation - moderate, extreme; pitch level-high, low; amplitude level-low, high; tempo - slow, fast. The resulting 64 stimuli, rendered two times each, were edited in random order on to a demonstration tape.

Raters: Ten undergraduates, six male and four female, were used as raters. They were recruited by sign-up sheets and were paid.

Procedure: The raters heard the tape-recorded stimuli in random order and were asked to rate each sample on ten-point scales of pleasantness, evaluation, activity, and potency as well as to indicate whether the sample to be rated could or could not be an expression of the following emotions: interest, sadness, fear, happiness, disgust, anger, surprise, elation, boredom.

Results: Table 1 shows F-ratios, significance levels, and the direction of the effect for main effects and two-way interactions with $p < .01$ yielded by a five-way analysis of variance with repeated measures. The parameters that seem to have had the most influence on the judges' ratings are tempo and pitch variation. Moderate pitch variation leads to ratings of generally unpleasant emotions, like sadness, fear, disgust, and boredom, showing little activity or potency. Extreme pitch variation and up contours produce ratings of highly pleasant,[2] active, and potent emotions such as happiness, interest, surprise, and also fear. Down contours have similar effects but do not seem to contain elements of surprise or uncertainty. Fast tempo leads to an attribution of high activity and potency as in the emotions of interest, fear, happiness, anger, and surprise. Slow tempo is seen as indicative of sadness, disgust, and boredom.

Extreme amplitude variation is seen as active and potent, mostly indicative of the emotions of fear and anger, whereas moderate amplitude variation is seen as happiness or disgust. High pitch level yields happiness and surprise, low pitch level, on the other hand, leads to ratings of disgust and boredom. High amplitude level leads to ratings of potency.

There is some evidence for differential acoustic manifestations of different types of specific emotions. For example, whereas anger is generally characterized by extreme amplitude variation and fast tempo, which may represent "hot" anger, a significant interaction effect shows that moderate pitch variation and moderate amplitude variation interact to produce higher ratings on anger, possibly indicative of "cool" anger. Another interesting interaction effect, that leads to consistently higher ratings on activity and surprise, usually associated with up contours, occurs between down contour and high pitch level which may represent a special type of novel situation.

Study II

Stimuli: 16 of the 64 stimuli used in Study I were chosen to represent happiness, fear, anger, and sadness.

Raters: 166 undergraduates, 69 male and 97 female, rated the stimuli during a demonstration in class.

Procedure: The raters were asked to choose between a pair of alternative labels for each of the 16 stimuli. The "correct" label was determined by the highest mean rating of the respective stimulus in Study I.

Results: The frequency distribution of the raters over the number of correct choices in shown in the following table:

Number of correct choices	1 - 7	8	9	10	11	12	13	14	15	16	Total
Number of raters	0	3	8	15	21	35	39	36	8	1	166

There were no significant differences in accuracy between male and female raters. The degree of accuracy shown by the judges is far above of what may be expected by chance ($p < .001$)[3]. Furthermore, most of the errors made are due to inaccurate choices on 4 of the 16 stimuli[4], the error distribution being significantly different from chance ($p < .001$)[5].

Conclusion: These results support the contention that the attribution of emotional meaning from auditory stimuli is based on characteristic patterns of acoustic cues. Specifically, there is evidence for earlier suggestions (Scherer, 1971; Scherer, Rosenthal, and Koivumaki, 1971) that specific cues or cue combinations communicate the major dimensions of emotional meaning. Relationships have been found between amplitude level and the potency dimension, between variation of pitch and amplitude as well as tempo and the activity and potency dimensions, and between pitch level and variation and the evaluative dimension.

The present approach suggests a rapprochement between studies on emotional expression in speech and the psychological investigation of emotion in music, with interesting implications concerning speculations on the common origin of music and speech in primitive emotional displays of our prehistoric ancestors (Langer, 1942). Pertinent studies on the cross-cultural universality of the vocal expression of emotion as well as on the development of the ability to recognize emotions from vocal or musical material in young children seem promising and have yet to be done. Judging from recent evidence (Ekman and Friesen, 1971) supporting Darwin's theory of innate mechanisms in emotional expression (Darwin, 1887), one may be justified in speculating about the existence of unlearned neural programs for the vocal expression and recognition of emotion, especially given the strong correspondences between respiratory phenomena and physiological correlates of affective state. This line of reasoning might eventually lead to a comparative analysis of the vocal expression of emotion in humans and auditory signals found in primate communication.

References

Darwin, Ch. The expression of the emotions in man and animals. London, 1872.

Davitz, J.R. (Ed.) The communication of emotional meaning. New York, 1964.

Ekman, P. and Friesen, W.V. Constants across cultures in the face and emotion. Journal of Personality and Social Psychology, 1971, 17, 124-129.

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

Langer, S. Philosophy in a new key. Cambridge, Mass., 1942.

Rogers, P.L., Scherer, K.R., and Rosenthal, R. Content-filtering human speech. Behavioral Research Methods and Instrumentation, 1971, 3, 16-18.

Scherer, K.R. Non-verbale Kommunikation. Hamburg, 1970.

Scherer, K.R. Randomized-splicing: A note on a simple technique for masking speech content. Journal of Experimental Research in Personality, 1971, 5, 155-159.

Scherer, K.R., Rosenthal, R., and Koivumaki, J. Minimal cues in the vocal commu ation of affect: Judging emotions from content-masked speech. Unpublished manuscript, Harvard University, 1971.

Vetter, H.J. Language Behavior and Communication. Itasca, Ill., 1969.

Footnotes

[1]The author expresses his gratitude to Martin Yaffee and Paul Leiman for help in the preparation of the synthesized stimuli. The data analysis was partially supported by a research grant (GS-2654) to Robert Rosenthal (Harvard University) who has contributed helpful comments. The study has been supported by an NSF institutional grant to the author's institution.

[2]After the present study was completed, the author was made aware of an experiment showing that pleasantness ratings of tone sequences bear a curvilinear relationship to the amount of stimulus variation, with moderate variation being perceived as most pleasant. (P.C. Vitz. Affect as a function of stimulus variation. Journal of Experimental Psychology, 1966, 71, 74-79). It is likely that extreme pitch variation in the present study corresponds to moderate variation in the former.

[3]Chi square test of goodness of fit to normal distribution.

[4]The reason for the much more frequent errors on these stimuli can be found in the fact that the mean difference in the ratings for both alternatives in Study I are much lower than for the rest of the stimuli. A correlation between number of errors and mean difference between alternatives for each stimulus yielded $r = .40$, $p < .10$, $N = 16$, one-tailed.

[5]Kolmogorov-Smirnov test of goodness of fit.

Table 1

F-ratios, significance levels, and direction of means[a]

Emotion \ Acoustic Parameter	PV	AV	PL	AL	TE	Interaction
Pleasantness	5.33** Ex,Down	1.81	< 1	< 1	2.05	11.26** LoAL + HiPL HiAL + LoPL
Activity	9.94*** Ex,Up,Down	5.98** Ex	4.23	8.73* Hi	35.48*** Fast	9.21** MoPV + LoPL DoPV + HiPL
Potency	23.46*** Ex,Up,Down	22.03** Ex	1.14	10.44* Hi	5.48* Fast	-
Interest	4.72** Ex,Up,Down	< 1	2.45	4.95	23.63*** Fast	-
Sadness	4.27** Mo	2.82	3.19	3.49	115.20*** Slow	13.97** MoAV + HiPl ExAV + LoPL
Fear	3.71* Mo,Ex,Up	6.32* Ex	< 1	1.12	11.05** Fast	-
Happiness	8.26*** Ex,Up,Down	7.17* Mo	9.38* Hi	< 1	33.30*** Fast	5.12** ExUpPV + Fast TE
Disgust	5.62*** Mo	22.50** Mo	6.43* Lo	< 1	6.37* Slow	-
Anger	1.22	6.70* Ex	< 1	3.84	7.43* Fast	4.83** MoPV + MoAV
Surprise	9.81*** Ex,Up	1.77	12.62** Hi	2.72	45.20*** Fast	7.38*** ExDoPV + HiPl
Ealtion	2.49	1.87	2.60	< 1	3.16	-
Boredom	5.59** Mo	< 1	5.50* Lo	< 1	60.19*** Slow	-

[a]Higher ratings were found for the level of each parameter shown in the cell

Abbreviations:

PV = pitch variation, AV = amplitude variation, PL = pitch level, AL = amplitude level, TE = tempo, Mo = moderate, Ex = extreme

*p < .05, **p < .01, ***p < .001