The Relative Importance of Vocal Speech Parameters for the Discrimination of Emotions

R. van Bezooijen and L. Boves
Nijmegen, the Netherlands

1. Introduction

In the experimental literature on vocal expressions of emotion two quite independent mainstreams may be distinguished, namely research which focuses on the description of emotional expressions in order to establish the characteristic features of distinct emotional categories, and research which examines the extent to which subjects are able to distinguish emotions from one another when vocally expressed. In the present contribution an effort is made to link the two approaches by comparing the results of a multiple discriminant analysis based on auditory ratings of vocally expressed emotions to the results of a multidimensional scaling analysis based on the outcome of a recognition experiment in which the same emotional expressions were used as stimuli. The aim of the comparison was to gain insight into the relative importance of various vocal speech parameters for the discrimination of emotions by human subjects.

2. Method

2.1. Speech material

The data base comprised 160 emotional expressions, namely 8 (4 male and 4 female) speakers × 2 phrases × 10 emotions. The speakers were students, native speakers of Dutch, between 20 and 26 years of age. None of them had had any professional training in acting. The phrases were 'two months pregnant' (/tve: ma:ndə zvɒŋər/) and 'such a big American car' (/zo:n ɣro:tə ɑmerika:nsə o:to:/). The emotions were the nine emotions included in the emotion theory developed by Izard (1971), i.e., disgust, surprise, shame, interest, joy, fear, contempt, sadness, and anger, plus a neutral category.

2.2. Auditory ratings

The recordings were randomized per speaker and per phrase, and rated by six slightly trained judges on 13 vocal parameters, i.e., pitch level, pitch range, loudness/effort, tempo, precision of articulation, laryngeal tension, laryngeal laxness, lip rounding, lip spreading, creak, harshness, tremulousness, and whisper. To collect the ratings use was made of preprinted successive interval scales. Every utterance could be given only a single rating on each scale, which means that the scores represent some kind of perceptual average. All scales were considered as absolute except the pitch level scale, which was effectively split up into separate scales for male and female speakers. The scales of lip rounding and lip spreading were mutually exclusive, i.e., only one of the two scales could be rated at any one time. The same was the case for laryngeal tension and laryngeal laxness.

2.3. Recognition experiment

The same 160 stimuli which were auditorily described - now randomized separately for male and female speakers - were offered to 24 male and 24 female students attending a Teachers' Training College in Nijmegen. The mean age was 20 years and 6 months, ranging from 18 to 28 years. The subjects were seated in a language laboratory and listened to the recordings via headphones. They were asked to indicate on a rating sheet which out of the ten emotional labels fitted best each of the emotional portrayals they heard.

3. Results and discussion

First, we will present the outcomes of a multiple discriminant analysis based on the auditory ratings. Next, the outcome of a multidimensional scaling analysis based on the results of the recognition experiment will be given. Finally, the results of the two analyses will be compared.

Before the auditory scores were subjected to a multiple discriminant analysis, a number of statistical analyses were carried out in order to make sure that the ratings were reliable, and that the various parameters were relevant to the aim of the study and informative.

The reliability of the means of the scores was assessed by means of the so-called Ebel-coefficient (Winer, 1971, p. 286). It appeared that the coefficients ranged from .86 for precision of articulation to .95 for pitch level, values which may be considered to be satisfactorily high. Therefore, in the subsequent analyses use was made of the mean of the ratings of the six transcribers on each of the 13 parameter scales for each of the 160 stimuli. In order to assess whether the scores on the scales varied as a function of the emotions expressed, the scores on each of the scales were subjected to separate analyses of variance with three fixed factors each, namely sex of speaker, phrase, and emotion (level of significance = 5%). It appeared that the factor sex of speaker was significant for only one scale, and that the factor phrase was significant for only two scales. However, the effect of the factor emotion was significant for all scales, except for lip rounding. Since we were not interested in a parameter which apparently had not been systematically used to differentiate between emotions, lip rounding was excluded from the further analyses.

In addition, product-moment correlations were computed in order to examine how the different parameters were related. The highest correlation, that between loudness/effort and laryngeal tension, was .71, which means that only half of the variance in one parameter was accounted for by the other. The correlations among the rest of the variables were considerably lower. On the basis of these results we decided not to discard any more parameters.

The mean scores for each of the 160 emotional utterances on each of the 12 parameter scales retained were used as input for a multiple discriminant analysis. A stepwise procedure was chosen, i.e., a method in which the discriminating variables are selected for entry into the analysis on the basis of their discriminating power. The objective of the analysis was to attain an optimal separation of the ten groups of 16 utterances per emotion by constructing a restricted set of discriminant functions or dimensions which are linear combinations of the original variables.

With three dimensions, the first one accounting for 41% the second one for 22%, and the third one for 18% of the variance in the original variables, 62.5% of the 160 utterances were assigned to the right emotional category. This is only 4.5% less than the percentage of correct responses yielded by the recognition experiment with human subjects. With 10 categories to choose from, which corresponds with an accuracy to be expected by chance of 10%, this means that the accuracy of both the statistical and human classification are well beyond chance expectation. In Figure 1 the positions of the group centroids in the discriminant space spanned by the first and second dimensions are presented.

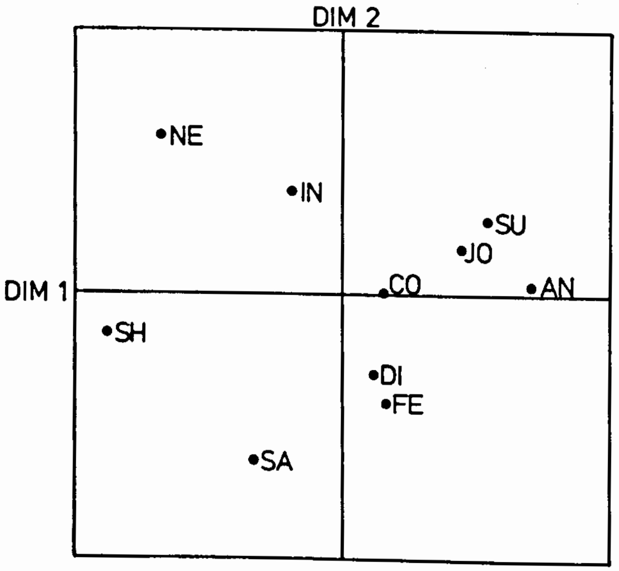

Figure 1. Positions of ten emotions in a two-dimensional space resulting from a multiple discriminant analysis on auditory ratings.

The positions of the different emotions along the first dimension presented in Figure 1 clearly suggest this dimension to be called a dimension of activation, shame, neutral, and sadness being the most passive emotions and joy, surprise, and anger being the most active ones. An interpretation in terms of level of activation is made even more plausible by the fact that the three parameters which correlate most highly with the first dimension are loudness/effort, laryngeal tension, and laryngeal laxness ($r = .73$, .66 and .54, respectively $df = 158$, $p < .001$). The only other correlation worth mentioning is that of pitch range with $r = .35$ ($df = 158$, $p < .001$).

The second dimension presented in Figure 1 is more difficult to name. If it were not for the extreme position of neutral and the neutral position of anger, it could be interpreted as an evaluative dimension, having unpleasant emotions like sadness and fear at its negative end and the pleasant emotions of joy and interest at the positive end. Another interpretation is suggested when we note that the positive end of the dimension is characterized by the least 'emotional' and at any rate the least impairing emotions, whereas the negative end shows a concentration of highly inhibiting emotions. The second dimension did not correlate very highly with any of the discriminating variables. The highest correlations were those with laryngeal laxness, pitch range, creak, and loudness ($r = .49, .44, -.36$, and .31, respectively, $df = 158$, $p = < .001$). The combination of much creak, high tension, low loudness, and narrow pitch range fits in with our previous interpretation of the second dimension in terms of emotional inhibition.

The third dimension extracted in the discriminant analysis is not considered here since we did not succeed in giving it a meaningful interpretation. It had disgust at the negative extreme, fear at the positive extreme, and neutral somewhere in the middle. It correlated most highly and almost exclusively with pitch level and tempo ($r = .74$ and .43, respectively, $df = 158$, $p < .001$).

Thus, from the distribution of the emotions along the three discriminant functions, and also from the correlations between these functions and the 12 vocal speech parameters, it appears that statistically speaking the most powerful variables for the separation of emotions are those which are related to level of activation. Are there any indications that the same would hold for the discrimination among emotions by human subjects?

In order to assess whether level of activation was an important criterion in the classification of emotional utterances by human subjects as well, the confusion data resulting from the recognition experiment were subjected to multidimensional scaling. Multidimensional scaling is a dataprocessing technique which has been especially designed to represent similarity between objects - and confusions among emotions could be interpreted as such - in terms of proximity in an n-dimensional space, thereby providing insight into the nature of the dimensions or stimulus characteristics that underlie the similarity judgments that subjects have emitted. In our case, the confusion data were processed by means of the MINISSA- program developed by E. Roskam and M. Raaijmakers from the University of Nijmegen, Holland, and

J. Lingoes from the University of Michigan, USA. Since asymmetrical input data were not permitted, the confusion data were first made symmetrical (Klein, Plomp, and Pols, 1970).

The combination of the Euclidean metric and two dimensions resulted in a stress of .11, which, according to Wagenaar and Padmos (1971), is significant at the 5% level. In order to enhance the comparability with the discriminant functions, the two axes resulting from the multidimensional scaling were rotated orthogonally in such a way as to maximally approach the positioning of the group centroids in the discriminant space as depicted in Figure 1. In Figure 2 the resulting configuration is presented.

4. Conclusion

Comparison of the two-dimensional spaces presented in Figures 1 and 2 shows that the projections of the ten emotions on the first dimension are quite similar: in both configurations neutral, shame, and sadness are situated towards the negative end; disgust, contempt, and fear towards the middle; and joy, surprise, and anger towards the positive end. The only notable level of activation has been one of the main stimulus characteristics on which the subjects have based their categorization decisions. In general the dimension of level of activation plays indeed a central role in the discrimination among emotional expressions. This also appears from the fact that in a recognition experiment in which groups of Taiwanese and Japanese adults judged the same emotional utterances, the same dimension emerged as well. On the basis of this outcome it could be hypothesized that the vocal parameters related to level of activation, i.e., loudness/effort, laryngeal tension, laryngeal laxness, and pitch range are not only important for the separation of emotional expressions in a statistical sense, but also in connection with the classificatory behavior of human subjects.

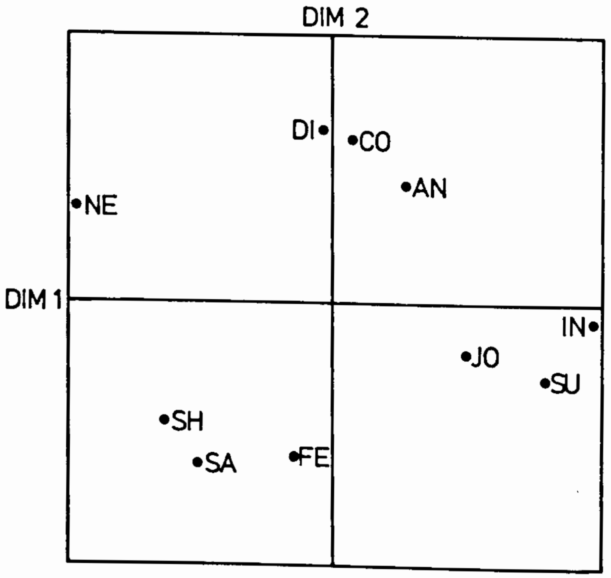

Figure 2. Positions of ten emotions in a two-dimensional space resulting from multidimensional scaling on recognition scores.

On the other hand, the projections of the ten emotions on the second dimensions presented in Figures 1 and 2 are very different from one another. Not much can be said, therefore, about the relative importance of the other vocal speech parameters for the discrimination amongst emotional expressions by human subjects.

References

Izard, C.E. (1971). *The face of emotion.* New York: Appleton Century Crofts.

Klein, W., Plomp, R., and Pols, L.C.W. (1970). Vowel spectra, vowel spaces, and vowel identification. *Journal of the Acoustical Society of America,* **48**, 999-1009.

Wagenaar, W.A. and Padmos, P. (1971). Quantitative interpretation of stress in Kruskal's multidimensional scaling technique. *British Journal of Mathematical and Statistical Psychology,* **24**, 101-110.

Winer, B.J. (1967). Statistical principles in experimental design. *Psychometrica,* **32**, 241-254.