

PERCEPTIVE AND ACOUSTIC CHARACTERISTICS OF EMOTIONS:
A Typological Research Based on the Material of Languages with Different Structures

L.V. ZLATOUSTOVA, G.Y. KEDROVA

Dept. of Philology
Moscow State University
Moscow, USSR, 119899

ABSTRACT

The present paper offers results of experimental phonetic research carried out with the purpose to solve some problems of describing different emotional states from the view of peculiarities of their phonetic expression in differently structured languages. With the help of complex techniques of auditory and acoustic analysis the main types of emotions relevant for speech communication and their phonetic parameters are determined.

INTRODUCTION

At present considerable interest is taken in the problem of phonetic expression of different emotional states, which is determined both by - the development of fundamental linguistics (mainly in the direction of communicative linguistics) and by the need to solve a number of most important applied tasks, such as diagnosing the emotional state of the speaker and listener by speech, normal and pathological emotions, identification of a person, as well as improving automatic speech synthesis and analysis. However, despite considerable success in the study of physiology and psychology of a person's emotional state, there has been produced no linguistic theory so far, which would describe and explain on all linguistic levels the similarities and variations in the phonetic expression of emotions in different languages in connection with the general type of a language structure. This task can be approached by first - revealing the typology of intonational patterns, corresponding to the types of emotional states relevant to speech communication; then - detailed description of their phonetic variants and variations; and finally by indexing emotional-phonetic "distinctive features" and their "weight" coefficients.

The paper presents the results of an experimental phonetic research undertaken with the view of solving some of the above-mentioned tasks seem to us very urgent, namely, revealing the types of emotions

which are most frequent in speech communication; establishing their phonetic parameters; preliminary observations on the typology of phonetic expression of emotions in different languages. The research has been conducted on the material of one-word utterance by complex techniques including auditory and acoustic analyses and statistical analysis of the obtained data. The research has been mainly oriented towards the Russian language.

1. MATERIAL AND TECHNIQUES OF OBTAINING RECORDS

The study of Russian plays and especially of stage directions which contain lexical designation of an emotion, has revealed the most frequent emotional states, seventeen of which have been selected as having the maximum frequency coefficients. Together with the neutral emotional state they have made up the initial list of emotional states to be analysed, i.e. neutral, pleasure, joy-delight-admiration, displeasure, indignation, anger, malice-hatred, irritation, contempt, rage, irony, menace, reproach, fear-fright, entreaty, despair, distress-bitterness, surprise. With these emotions by means of modelling an appropriate situation the word "ПОСАДКА" (LANDING) has been uttered by four native speakers and a professional actor (the age range - from 25 to 40). The words designating the emotional states as well as the cue word have been translated into English, French, German, Spanish, Polish, Czech and Bulgarian. The native speakers of these languages have been given the same task. To increase the degree of phonetic comparability of the material the words in all languages were selected so that they would be similar to the Russian word "ПОСАДКА" in their rhythmic and segmental composition. The recording has been made in a studio.

2. AUDITIVE AND PERCEPTIVE EXPERIMENTS

The records have been offered for auditory analysis to both native speakers of Russian (20 - 40 persons) and native speakers

of other languages (5 persons for each language). At the first stage it was necessary to identify the emotion as one from the list. At the second - to judge whether the produced utterance corresponds to the implied emotion or not. The group of Russian auditors (6 persons) were to assess the identity and difference of all possible combinations (pairs) of emotional realizations. Another group (16 persons) was to conclude as to the subjective similarity or dissimilarity of the emotions under analysis on the basis of the list of pairs of emotional states given to them.

3. ACOUSTIC ANALYSIS

Sonagrams (frequency range - up to 4 kHz) and oscillograms (intensity variations) have been made. The length of segments, changes of the voice pitch on the vowel segments and the maximum values of segments' intensity have been calculated.

4. EXPERT ANALYSIS OF SONAGRAMS

Realizations of all emotions produced by three Russian speakers and presented on sonagrams without indication of the emotion or the speaker have been offered to an expert in sonagrams with the task to class the multitude of these realizations according to certain features chosen by the expert himself. The features he chose and the classes they gave were registered in protocols.

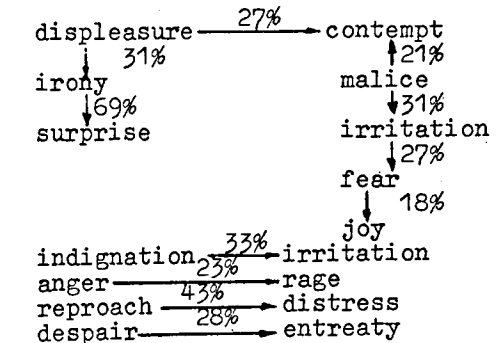
5. RESULTS

a) For the Russian language

The results of the acoustic and auditive analysis for the Russian language in the cases of all speakers show a rather high degree of concordance. Among the speakers under analysis it was possible to single out one whose emotional realizations were in auditors' opinion the easiest for identification. He also has the least disparity between the sounding and the implied emotion. His data will be used further on as an illustrative material. As a result, the following conclusions have been made: most of the emotions can be differentiated by ear rather easily. Anger, rage, irony, surprise, malice, irritation and entreaty are most distinctly opposed to all others; pleasure, contempt, distress, reproach and indignation are on the contrary the least distinguishable. Besides the assessment "similar stimuli", the assessment "identical stimuli" has been introduced. Alongside with the pairs made by the repetition of the same stimulus, certain realizations of irony and surprise have been recognized to be similar.

The identification of emotional states by the phonetic form of the stimulus in auditive experiments has yielded the following results: by degree of correct identification the emotions under analysis are ranged as follows: neutral (86% correct identification), surprise (88%), fear (68%), rage (75%), menace (75%), distress (69%), joy-delight-admiration (68%), entreaty (62%). Statistically relevant mistakes in the identification of emotions are given below in diagram 1 (it is noteworthy that these pairs of emotions' realizations were found similar in their phonetic form).

Diagram 1. Graph of confusions between emotional states in their diagnosis in speech

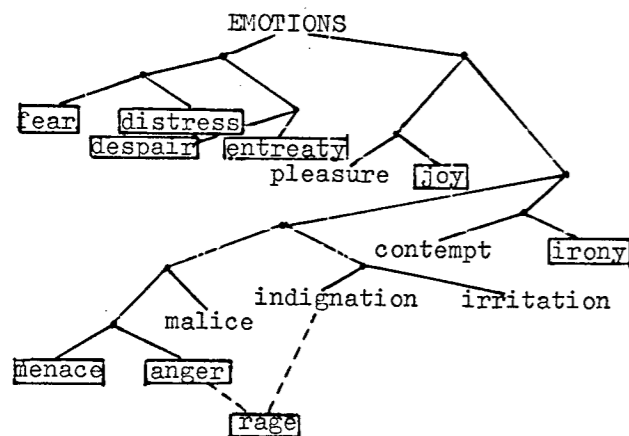


When estimating the acceptability of stimuli the majority of the auditors found all stimuli natural for the given speaker except the emotion of reproach. Thus, it is clearly seen that not all differences in phonetic form which can be perceived by ear are of equal importance for the emotions' identification. Therefore the question of relevant and irrelevant differences in the phonetic expression of emotions is closely connected with the subjective estimation of the distance between the emotional states under analysis. The subjective space of the emotions in question is as follows: surprise, reproach and neutral emotions do not form any groups and are opposed to all other emotions and to one another. The relationships among other emotions are given below in diagram 2.

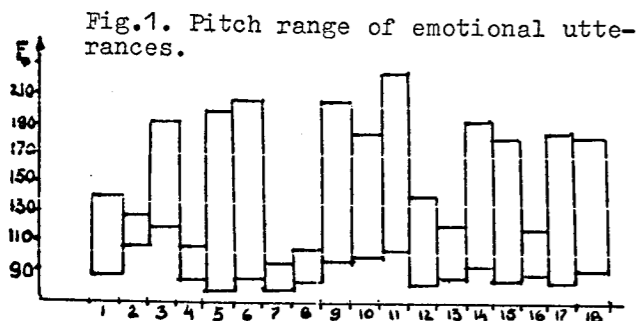
The comparison of the subjective estimation of the distance between the emotions with the data of auditive experiments shows that intense emotions as well as biologically vital ones (such as reaction to a danger, signal of aggression, etc.) have certain advantages in the process of identification. The data obtained as a result of perceptive analysis make it possible to range the results of the instrumental acoustic analysis as communicatively relevant parameters of sounding and peculiarities connected with the individual manner of the speaker and other characteristics.

nce factors.

Diagram 2. Similarities/dissimilarities between the emotional states.



Emotions that are intonationally marked and for this reason are easily and correctly identifiable are taken in a framework. Voice Pitch. It seems convenient to begin the presentation of the results of the instrumental acoustic analysis with the Voice Pitch. Pitch range and the average level of F₀ proved to be rather significant for the differentiation of emotions. Fig.1 shows the Pitch range of all emotions, 1 - being neutral utterance, 2 - pleasure, 3 - joy-delight-admiration, 4 - displeasure, 5 - indignation, 6 - anger, 7 - malice, 8 - irritation, 9 - contempt, 10 - rage, 11 - irony, 12 - menace, 13 - reproach, 14 - fear-fright, 15 - entreaty, 16 - despair, 17 - distress-bitterness, 18 - surprise.

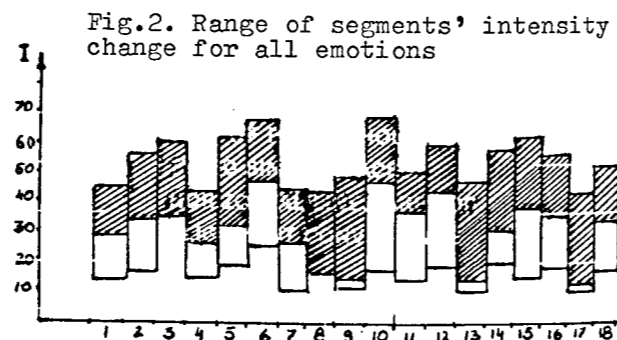


The emotional states with marked Pitch range are: pleasure, displeasure, malice, irritation, despair (minimum range of F₀), surprise and irony (maximum range of F₀). On the whole, the Pitch range of all emotions except those with marked minimum range, is larger than neutral. According to average F₀ value, displeasure, malice and irritation are lower than neutral, while joy, contempt and irony are higher than neutral.

According to the form of F₀, two types of the contour are important for the expression of different emotions: level contour

(level-falling) and rising (rise-fall) one. It is noteworthy that the form of F₀ contour influences also the tempo deformations of the rhythmical word structure. The tone contour is an important parameter for the emotions of menace (level), distress (level-falling), contempt (level-falling), malice (level-rising), entreaty (rise-fall). This parameter is especially significant for the realization of surprise and irony in one of its phonetic variants: it is rising and rising-falling in these cases.

Intensity. The overall range of intensity changes is not very informative: on the whole for all emotions it is larger than for a neutral utterance (except displeasure and irritation). See Fig.2.



More informative is the range of intensity change on vocal segments (the shaded area). It is minimal for displeasure, anger, rage, menace, irony and malice. The average intensity of vowels and the maximum values of intensity correlate with the differences in overall intensity of the stimulus. It becomes possible to differentiate between intense emotions, i.e. anger, rage, joy, menace and entreaty, and weak emotions, i.e. distress and contempt. For intense emotions the average peak values of intensity on one segment are 31.6 - 34.1 dB, for weak ones - 18.3-20.8 dB, for neutral utterance - 26.6 dB.

According to the relative intensity of syllables in the total intensity of stimulus we can make up a group of emotional realizations with strong emphasis on the end. These are: anger, fear, menace, irony and surprise (though in a lesser degree). Similarly we can form a group of emotions with weakened post-tonic part. They are: contempt, distress, irritation, and to a smaller extent - joy and malice. Besides, irritation, contempt, indignation, distress and fear are grouped as having an emphasis on the stressed syllable. The relative intensity of syllables within the total intensity of the word is as follows:

anger	.21	.44	.35
fear	.22	.46	.32

menace	.24	.42	.34
irony	.26	.38	.35
surprise	.26	.41	.32
indignation	.23	.48	.28

emotions with weakened post-tonic part			
contempt	.33	.48	.19
distress	.32	.47	.21
irritation	.28	.51	.21

Duration. The fluctuations of the total word duration from speaker to speaker are so substantial that it is impossible to rely on the absolute values when judging the changes of tempo. It can be pointed out in general that the maximum duration is typical of rage and menace, the minimum - of irritation and displeasure. According to the duration of unstressed syllables in relation to the stressed one the emotions under analysis can be ranged as follows:

for post-tonic part		for pretonic part	
distress	1.02	rage	.38
neutral	.98	irritation	.36
menace	.89	distress	.33
fear	.88	menace	.32
anger	.88	pleasure	.28
pleasure	.85	displeasure	.28
irritation	.82	irony	.28
rage	.81	fear	.26
malice	.80	malice	.26
displeasure	.78	joy	.25
entreaty	.78	indignation	.25
joy	.78	entreaty	.24
surprise	.77	anger	.23
irony	.75	surprise	.22
contempt	.74	despair	.20
indignation	.73	contempt	.18
despair	.58	reproach	.14
reproach	.51	neutral	.36

As it can be seen, distress and menace have the maximum duration of unstressed syllables, which leads to levelling-out of all the syllables. On the contrary, in despair and reproach we find considerable reduction of post-tonic and in reproach and contempt - of pretonic syllables. In the realizations of irritation and rage pretonic syllable is marked by unusually long duration. On the whole the ratio of longer and shorter syllables in emotional speech is different from that in neutral speech. The former has greater duration emphasis on the stressed syllable. The length of certain phonetic segments in respect to the length of the whole word makes it possible to reveal considerable increase of sibilants (in our case S) in the emotional states of malice, rage, displeasure as well as the decrease of voiceless plosives in post-tonic part in rage and anger.

The share of vowel segments in the total duration of the word shows the degree of the word's vocalization. According to this

parameter we distinguish between "voiceless" emotional states, namely despair, displeasure and contempt - and "voiced" emotions, i.e. anger and rage.

Spectrum. The results of the expert analysis of sonagrams have not enabled us to single out definite spectrum peculiarities connected with this or that emotional state. According to such features as the position and relative intensity of upper formants (F₃ and F₄), frequency area and localization of noise in plosive consonants and spirants, only the speaker's individuality can be determined with sufficient validity. Thus, it has to be stated that frequency range up to 4kHz might be not large enough to reveal the acoustic differences in the spectrum (to be presented in sonagrams) which could be correlated with those in timbre.

b) Results of the analysis for other European languages

The results of the perceptive and instrumental acoustic analyses of some other European languages have been used in the first place to confirm or refute the hypotheses concerning the relevance of these or those phonetic features of different emotions as well as to define the relative weight of different parameters. As it appeared, the phonetic system of a language and caused by it "phonetic background" of a listener influence greatly not only the correctness of identification but the very possibility of making some definite conclusion about the emotional state of the speaker. In general it is more difficult to judge emotions in a foreign language than in a native one.

For Russian auditors identifying the emotional state, the most weighty parameter is the form of tone contour if only it is not falling. Despite the peculiarities of segments' duration or their intensity, such utterances are perceived as surprise. Deformation of duration structure, like for example in languages with phonologically long vowels, also leads for Russian listeners to recurrent mistakes in emotions' perception, especially of those characterized by serious deformations of rhythmical structure typical of a neutral utterance, i.e. despair, entreaty and irony. Differences in greater or lesser tenseness of the articulatory basis of a language tell on the perception of the emotions based on contrasts in intensity (both between words and segments of words). The best identifiable in all languages are such emotions as surprise, distress, anger, rage, despair-entreaty, irony and also neutral utterance.