Karel Sedláček; Antonín Sychra

The method of psychoacoustic transformation applied to the investigation of expression in speech and music

Kybernetika, Vol. 5 (1969), No. 1, (1)--28

Persistent URL: http://dml.cz/dmlcz/125253

KYBERNETIKA ČÍSLO 1, ROČNÍK 5/1969

The Method of Psychoacoustic Transformation Applied to the Investigation of Expression in Speech and Music

KAREL SEDLÁČEK, ANTONÍN SYCHRA

The present study treats the comparison of sentences pronounced before and after listening to the vocal and dramatic music with the aim to find out the objective acoustic and physiologic carriers of the emotional expression in speech and music.

1. INTRODUCTION

1.1 Subject of Investigation

Many specialists have serious objections to the possibility of the exact study of expression in speech and music. Expression as an immediate manifestation of life can be, in their opinion, only experienced, but it is unanalysible and undocumentable because it cannot be measured. Only the products of intentional activity based on thinking are measurable and therefore documentable, as for instance Ringbom [18] in accordance with the philosophy of Palágyi [16] concludes. This point of view expresses the consensus of opinion, which is typical especially in the artistic sphere, that emotionality including artistic creation is merely a matter of intuition, inspiration and other activities held as irrational. The conception of any analysis meets with opposition. It is necessary to say that this doubt is not without importance and must not be overlooked. The phenomena we have to deal with are very complicated and the method of their study must also be worked out gradually. On the other hand it is not possible to omit these questions as present-day technology and science create the elementary conditions for it not only for putting the questions but also for finding successful answers [1], [2], [6], [11], [13], [25].

In our work we were not concentrating on the study of *emotion* itself, but of *expression, i.e. the communicative means transmitting the emotions.* Objectively it is easiest to study the physical communicative codes i.e. sounds, as they are the carriers of all information. By studying we can also learn much about the emotions themselves.

1.2 Process of the Perception of Speech

Some basic conclusions about the reception of emotional information may be obtained from the organization of the process of perception [21], [20]. The minimal schema of the stages of the hearing analysis are given in fig. 1.

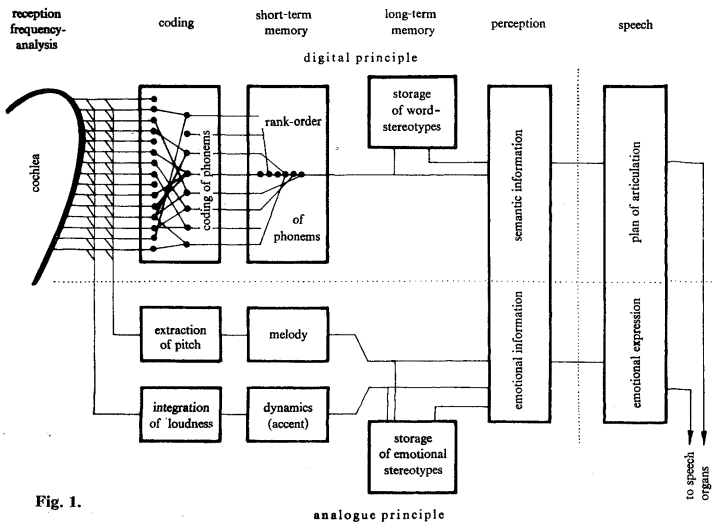


Fig. 1.

The perceived sound of speech is displaid in the cochlea into different frequency bands representing the formant regions. The coding of vowels is traced in the next step as the combination of formants, so that (at least) two formants represent the characteristic timbre of the vowel. At the same time the pitch is realized on nerve impulses originating from all the stimulated places in the cochlea. The third type of coding is the integration of the number of impulses conditioning the sensation of the loudness of the sounds.

A further step is the ranging of different sound pictures into certain time patterns where the successive order of speech sounds has the main significance. This is the function of the so-called *short-term-memory* in which the phonological picture of word is given. The same process runs through the other two channels. The ranging of the pitch of the syllables forms the melody of the word and of the sentence. The time-pattern of loudness is shown as the dynamics, i.e. as word accent. The three main components described above give rise to the auditory picture of the word which is then compared in the next step with the storage of word-stereotypes, saved in the *long-term-memory*. From this comparision we have to decide a) whether the sound heard belongs to the sounds of speech and b) what is its meaning. The semantic field of communication is thus given. The same process runs through the other two channels. The melodic and dynamic pattern is included in the system of emotional stereotypes by confrontation with those already received

either inherited as unconditioned reflexes or acquired by experience, by registering behaviour in different emotional situations. This emotional information complements the semantic information and is sometimes very important.

Between the two kinds of analysis, i.e. between the *semantic* and the *emotional*, there exists a fundamental difference. The changes of pitch are to a great extent imitated in the nervous system by similar changes in the frequency of the nerve impulses. In the same manner the loudness pattern is in direct dependence on the intensity of the sound. The perception of melody and dynamics has thus a character of *immitation* on the basis of the *analogue principle.* On the contrary during the coding and identification of vowels and consonants the *digital principle* comes into play; the whole continuum of speech sounds is divided up into a limited number of *discrete values* which are qualitatively different and which represent the phonemes as phonological units. For the intelligibility of speech, the information is reduced to the discrimination of the phonemes which represents only a small number of bits per one letter (cca 5 bits per one letter) (see also Charvát [9]). The semantic information based on the digital principle can be transformed very easily and without loss into another code, for instance into writing. Emotional information on the other hand can be transmitted only by sound, i.e. by speech; by transforming the report into writing it is partially lost.*

In accordance with what has just been said, Roman Jakobson [10] and his school rightly distinguish *semantic* and *physiognomic* expression. In this study we were concerned above all with physiognomic expression. Only the code which transmits physiognomic expression can be compared with music; in contrast, musical semantics, insofar as it exists at all, is governed by quite another system. This is not to say that we may overlook semantic expression as very frequently it influences the physiognomic expression.

2. PREVIOUS METHODS OF EXAMINING EMOTIONS

There are several ways of studying expressions: in the subjective method of *introspection* the test subject observes his psychic activity and announces the result of his observation to the experimenters. Here it is necessary for the participants of the experiments to make their announcement verbally, thus they make the semantic analysis. It is possible to make the experiment in two ways: either in the form of prepared questions, i.e. by using a *questionnaire*, or in the form of *free verbal statement*. The second way has the advantage that the listeners are not influenced by the experimenter, but the first is more suitable for statistical processing. This is very important as the results of the statistic processing of a group of listeners can be considered as a proof, the probability of which can be exactly computed. The introspective method however has many disadvantages because it is impossible to establish: a) which acoustic means are relevant for the

* We say "partially", not "altogether". Modern versologists basing themselves on phonology have proved that the delivery of a text is to a degree fixed by its phonological structure (see Mukařovský [14]). On the other hand the influence of the phonetic pattern upon the melodic intervals chosen by different composers cannot be proved [17].

expression, b) which acoustic means are carriers of other information, c) which acoustic means are not used at all even though they are present. By this way we cannot know to what extent the expression is carried directly by the psycho-physiological effect of the sound or what part of the expression belongs to the semantics of the context, or in music to conventional musical symbols. The same difficulties will probably very soon meet also **Knobloch, Poštolka** and Srnec in their study of interpersonal process [12].

We can denote the second trend in the study of expression as the objective methods, methods of *introspection* which were much used in the first half of this century. Above all it is a question of the registration of the *vegetative phenomena* accompanying the emotions, or a matter of the study of the *behaviour* of animals and human beings. The vegetative signs of emotion, such as changes in the galvanic skin resistance (the so-called galvanic skin reflex), changes in blood pressure, skin temperature, reactions of the pupils etc., can be considered with certain reservations as indicators of the presence or of the intensity of the psychical activity, but it is not possible to discriminate qualitatively the different emotional contents [30]. It happens very frequently that contrary emotional expressions produce the same reactions.

3. THE METHOD OF PSYCHOACOUSTIC TRANSFORMATION

3.1 The Purpose and the Principle of the Method

We shall now describe a method which we used for the study of the emotional expression in music and speech and which, as we assume, enables us to overcome some difficulties connected with the direct assessment of expressive content. The purpose of this method is:

a) To determine whether the expression which we suppose to be present in a muscial or verbal recitation is *objectively detectable* with mathematically defined probability.

b) To try to separate the expression transmitted by the verbal context from the expression which is carried by pure *acoustic means.*

c) To ascertain the acoustic means used for the communication of the expression and to differentiate them from the acoustic phenomena which are irrelevant for the expression.

d) Beside the expression in speech to investigate also the emotional content of music but without forcing the observers to perform the direct semantic analysis which, as is known, fails very often in the study of musical content.

The principle of our method is in the statement of *changes arising in the speech of a test subject under the influence of a certain musical fragment.* We suppose that the music influences the psychological activity of the test subject; it causes changes in his behaviour which also means in the expression of his speech utterances.

If this were not the case, then the recitations of the same context before and after listening to the music would not differ significantly or, if at all, then only accidentally. We can also say that the changes in the interpretation of a word context, before and after the hearing of the music, reflect directly the influence of the music upon the

individual. *A transformation of the music expression into a speech expression takes place.* Therefore we have designated this method as the *method of psychoacoustic transformation.*

3.2 The Arrangement of the Experiments

We chose a certain number (18) of short musical fragments accompanied by the composers' texts. For the most part these were the fragments of vocal and dramatic music. We also included in this test some instrumental parts of Dvořák's symphonic poems, to be exact, parts which in the composer's sketches had been connected with certain concrete verses of the poet K. J. Erben [26].

We chose very short fragments in order not to be influenced by other random factors and also for the reason that a longer fragments could contain more emotional reports. These musical samples we recorded on a tape with intervals of 20 seconds. We were playing these tests to actors from Prague's theatres and we gave them the task of reproducing the respective text with the dramatical expression which they heard in the music. These actors' recitations recorded on a tape were compared with the recitation of the same text as realized before the subjects had listened to the music. The actors did not know anything about the context of the given composition and for the most part they did not recognize the work from which the sample was taken.

3.3 The Analysis of the Sound Material

All records were analysed in detail from different objective and subjective points of view.

The *objective analysis* was made with the aid of electroacoustic devices, giving the possibility of a detailed study of several physical parameters. The sound intensity was registered by the level recorder (Bruell & Kjaer), the melody of speech by an apparatus developed by O. Janota. The spectral composition was analysed by sonagraph enabling spectral analysis in the broad frequency range as well as the precise registration of all transition phenomena. By the sonagraphic recording, using the narrow filter, we complemented and corrected the record of the melody [22]. Further the integrated long-interval spectra were recorded by using third-octave filters. These spectra represent the general tendencies of timbre. All apparatuses permitted the study of time factors, tempo and rhythm. The most exact in this respect was the sonagraph, enabling the sure identification of the boundaries between the syllables and phonemes in so far as such sharp boundaries existed.

The *subjective* analysis was directed towards the discovery of what emotions were communicated in the single recitations. This analysis was made by some groups of observers to whom the recitations were played. In the statistical processing of the answers we always expressed the significance of results by the theory of probability. We were aiming mainly at the determination of differencies between corresponding recitations.

Great attention had to be paid to the composition of the questionnaire. A well composed *questionnaire* ought in our opinion to observe the following principles: a) the frequencies of word characteristics defining the individual types of emotion

ought to correspond to the actual variety of the emotions and their distribution according to probability, b) individual emotional groups ought, for the sake of the application of statistical tests, to be logically systematised. A difficulty lies in the fact that none of the classifications of emotional experiences have hitherto corresponded to strict logical requirements. Difficulties further arise out of the fact that no language in its denotation of expression is a closed system; all languages are somewhere lexically inadequate and somewhere contain redundancies. For these reasons, in spite of the experience of earlier scholars [7], [8], we do not consider even the last version of our questionnaire to be satisfactory. Nevertheless we think that, under the condition of a suitable statistical test, this questionnaire of ours well serves the purpose of illuminating the basic problems of the material. If we choose a representative statistical sample then we can say that the results which we obtain have, to a certain extent, general (intersubjective) validity. We expect that we shall be able to make up a more suitable questionnaire on completing the system of emotions on the basis of experiments.

3.4 Discussion on Methodology

In the way described in 3.2 (The Arrangement of the Experiments) we obtained *pairs of taped recordings*, one before listening to the music and the other after it. We supposed that the expression of the recitation *before listening* would be determined by the sense of the text, or in very short fragments would be uncertain. As for the interpretation *after listening* to the music, we wanted to ascertain if the music was able to influence the expression of the actor's recitation.

The main advantage of this method lies, in our opinion, in the fact that it is not necessary to compare directly the physical, physiological and psychological phenomena, but the *differential method* is used on principle. *The direct comparison is performed only on the same level which is the only case where it can be exact.* This means that acoustic factors are compared with acoustic, physiological with physiological etc. and the differencies stated. The method of differences has also been used by Čistovičová in her studies of deformations of language and understanding of its sense [4]. Then the statistical correlations are ascertained ex post between these two series of differencies. At the same time the individual personal differences of voice, differences in the phonetic structure of the sentences are cancelled out, because it is always the same sentence pronounced by the same person which is evaluated. We prefer this comparative method to that proposed to us by Fónagy [5] to investigate the artificially produced sounds immitating speech, because we want to keep the natural conditions of the speech communications.

The conclusions at which we have in this way arrived allow of a certain *generalisation*, because we *are comparing music and speech*, i.e. *two different systems, in both of which expression is carried by acoustic factors.*

4.1 The Statistical Evaluation of the Subjective Judgements of the Observers in the Method of Psychoacoustic Transformation

The subjective evaluation of each listener concerning expression is influenced not only by the actual content of the message but also by a number of further random factors which modify the expression, although in defiance of their existence the expressive intention is conclusively detectable in the acoustic structure of the sentence; we mean hereby for example the character of the individual, his life experiences, his mood of the moment, stage-fright, his attentivness, and last but not least his individual sensitivity. Therefore for establishing the expression we must not rely on the judgement of a single subject. We must find a *representative sample* of listeners and hereby we can *eliminate random factors* and, to a degree, obtain *objective* or rather *intersubjective* results. In view of the newness of the subject of investigation it was not easy to select a statistically representative sample. The reasons for this lie in the difficulties arising from differences of nationality, social background, individual psychology, education and age, and from the variety of cultural traditions, ethnic conditions etc.

In the first series of experiments we solved this problem by confronting the basic sample with control groups from particularly distant societies [27]. In concrete terms: for the questionnaire test we used post-graduate students of esthetics from the Philosophical Faculty of Charles University (musicians, directors, art theoreticians, journalists etc.); for checking purposes we repeated the experiment with German teachers and students from Humboldt University, and finally with newly arrived students of Charles University from Asia, Africa and Latin America, who did not yet speak Czech. The numerical information, published in an earlier work [27], speaks very clearly of the *generality* and *objectivity of physiognomic expression* in speech, and of its *relative independance on the knowledge of the language.* Other scholars have already arrived at similar conclusions [29]. This of course does not mean that we should overlook the influence on expression of social conditions, national traditions and linguistic conventions, not does it mean that we should overestimate anthropological factors.

There is at present in press our study linking on to the radio investigation of Karbusický and Kasan [24]. In it we study eight groups of the public having different work and regional origins, with respect to their ability to recognize the emotional content of the eighteen musical samples. The results obtained from 816 people showed that *differences of ability to determine the expression are not fundamental.* At the same time we discovered that a group of from 60 to 100 persons may be considered representative [24]; even with a group of around 30 persons we begin to find a certain stabilisation of the results.

As far as concerns the statistical treatment of the results we were originally satisfied

with mere observation of statistical data, i.e. with the comparision of relative frequencies, without endeavouring to state mathematically the level of significance [23], [27]. In later work however this became absolutely essential.

In spite of having the cooperation of mathematicians* we were unable to decide which of the tests employed was the most appropriate.

We tried out a number of different methods, in particular comparison of means, dispersion variances and standard deviations, medians, determination of correlation coefficients and rank correlations. With time we concentrated on methods which permit mathematical evaluation of the differences between two or more statistical samples (tests of goodness of fit), both overall and in individual parameters.

Among the regular tests the best known are the *method* χ^2 and the *Kolmogorov-Smirnov distribution test* [19]. The former has the disadvantage that it cannot be applied if frequencies of 5 or less (sometimes two or less with obligatory correction) occur in one of the samples. However from the test material precisely such low and indeed frequently zero frequencies in some parameters are symptomatic as a proof of expressive precision: if for example a fragment is clearly joyful we cannot assume that a number of test subjects will detect sadness in it etc. For this reason we preferred the Kolmogorov-Smirnov test against the use of which there are in this case no serious theoretical objections. This test was particular suitable for the comparison of two fragments in all parameters simultaneously. To discover whether the difference is statistically significant in a given concrete parameter we used the simplest test, the χ^2 *four-fold table.* With low frequencies (5 or 2) we applied binomial evolution to state exact possibilities.

In sofar as we work with a multi-parameter test in which we can on the basis of extended material state the expected frequencies, we apply the *Poisson theorem.* This theorem approximates very well the binomial evolution because in a multi-parameter system of emotions the probability of the individual parameters does not as a rule exceed 0.1 (10%). Let us demonstrate a few examples.

An example of the evaluation of a judgement of a series of sentences before listening and after listening to the music is to be seen in Table I. This is the judgement of the sentence "jako by sem smrt načuhovala" ("as if Death were peeping in") taken from Janáček's opera Jenufa and interpretated by three actresses. Two interpretations were realized before listening to the music (ex. 1, 2) and three after it (ex. 5, 13, 14). At first we stated whether the recitations before listening to the music differ significantly or not. By means of the Kolmogorov-Smirnov test a very low value of the characteristic z was obtained ($z = 0{\cdot}9785$), proving that no significant difference exists. Similarly the recitations after listening (ex. 13 corresponding with ex. 2 and ex. 14 corresponding with ex. 1) did not differ significantly the characteristic z being 0·98012. Therefore it was possible to add the values in both groups. The difference between the judgement of recitation before listening and after listening was found at a very significant level 0·1% of probability ($z = 2{\cdot}4692$).

The next task was to discover in which expressive characteristics occur significant

* We are indebted to the Institute of Mathematics of the Czechoslovak Academy of Sciences, the department led by Academician Novák and in particular Dr. Nosál for giving us consultations and aiding us with calculations.

Table I.

Jako by sem smrt načuhovala (As if Death were peeping in)					
Sample No	1	2	5	13	14
	before listening		after listening of music		
1. Neutral, indicative sentence	15	23	0	4	5
2. Shame, coyness	1	0	0	0	0
3. Pride, superciliousness	0	0	0	0	0
4. Love, tenderness, sensuality	0	0	0	0	0
5. Joy, happiness	1	0	0	0	0
6. Playfulness	0	0	0	0	0
7. Celebration, glory, pathos	0	0	0	3	0
8. Humorousness, grotesqueness	1	0	0	0	0
9. Surprise, wonder	3	0	0	3	6
10. Excitement	0	0	1	0	1
11. Longing, desire, yearn	0	0	0	0	0
12. Grief, pain	2	3	4	12	8
13. Resignation, bitterness, pity	5	9	0	13	1
14. Fear, fright	2	1	30	6	11
15. Feeling of outrage, hurt	3	0	0	0	0
16. Reproach	2	0	0	0	1
17. Irony, derision	1	1	0	0	1
18. Refusal, awfulness, dreadfulness	2	0	0	0	0
19. Anger, hate, threat	0	0	0	0	0

differencies. To this end we used the χ^2 test as mentioned above, or the calculation of probabilities on the basis of binomial evolution. We discovered that above all the characteristic *announcing sentence* recedes significantly into the background after the listening to the musical fragments, at a level of probability of $2{\cdot}278 \,.\, 10^{-4}$ (0·0278%). In contrast there was a significant growth in the value of characteristics 13 – i.e. *sadness, tragic pathos, pain* $P = 0{\cdot}0973\%$ and 15 – i.e. *fright, fear* $P = 0{\cdot}07067\%$.

If we take a look at the actual figures in Table I we discover that the results of the two actresses are not quite identical in this experiment. In both we find an increase in the frequency of judgements in the above mentioned parameters but one of them gave much more emphasis to the parameter of *pain* and *resignation*, while the other stressed that of *fear*.

These results correspond quite unambiguously to the dramatic context of Janáček's music in the given scene. It was not by chance that when we gave this musical fragment to 816 listeners (in the above mentioned work [24]) for judgement, we arrived at the analogous conclusions by a direct questionnaire analysis. In spite

of the conviction carried by these results we discover that in some cases the interpretation of the listening comes out incomparably more pregnantly. In support of this statement we have included fragment no 5. where the Kolmogorov-Smirnov characteristic, as against the sum of fragments 1 and 2, came out as 3·6010 which is a highly significant value with negligible probability of a chance result; at the same time the parameter of *fear* was determined by as many as 30 listeners out of a total of 35, which under a choice of 20 possible parameters is an extraordinarily significant value.

The reader will certainly wonder how it is that this emotional content was not given unambiguously by the semantic structure of the sentence "as if Death were peeping in". The explanation of this fact is much less complicated than at first sight. In fact this phrase is not so semantically unambiguous for someone who does not know Janáček's opera, as the word „načuhovala" (peeping in) is dialectal and even allows of humorous interpretation.

We proceeded in a similar way with the treatment of the subjective questionnaire evaluation of the sentence "the bed is made already" from Janáček's "The Diary of One who vanished". The results are recorded in Table II. Before listening fragments 17 and 18 were produced, after hearing, fragments 22, 23 and 28; no. 17 relates to 22 and 23, 18 to 28.

In this case also we first of all endeavoured to discover whether the subjective evaluations of the fragments before the listening differ significantly from each other in any way. With the application of the Kolmogorov-Smirnov test this turned out not to be the case; the value obtained, $z = 1{\cdot}195$, is quite insignificant.

It is interesting that the pair of sentences "the bed is made already" do not differ in the slightest before the listening from the pair of sentences "as if Death were peeping in" before the listening. The characteristic $z = 0.6196$ is even in this case an insignificant value. From this it follows that before the listening, irrespective of the semantic content, the sentence comes out very often as a neutral utterance.

The first look at the frequencies already clearly shows that in the individual parameters after the listening the results here are not unambiguous as opposed to the case of the preceding sentence. This however we might have presupposed since the expressive content of the musical extract "the bed is made already" is more complex. From the context of Janáček's "The Diary of One who vanished" which of course the test subjects did not have at their disposal – it is apparent that this place catches the *amorous enticement* of Janíček by the gipsy Zefka; but simultaneously both the poetic text and Janáček's music in these scenes at the climax of the *love* tie between Janíček and Zefka are undercoloured with tones of *sadness and resignation* (immediately before it comes Zefka's lament about the hard life of the gipsies) [27].

First of all we established whether the variance of the characterisations of the expressions after the listening differs significantly from the corresponding sentences before the listening. We found very significant deviations in the case of sentences 17 as opposed to 22 and 18 as opposed to 28. In the first case we found $z = 2{\cdot}762$

Table II.

Tož už mám ustlané (The bed is made already)

Sample No.	17	18	22	23	28
	before listening		after listening of music		
1. Neutral indicative sentence	22	12	0	14	1
2. Shame, coyness	2	0	0	0	0
3. Pride, superciliousness	0	2	1	1	0
4. Love, tenderness, sensuality	0	0	13	0	0
5. Joy, happiness	5	7	13	5	5
6. Playfulness	0	0	11	0	0
7. Celebration, glory, pathos	0	0	0	0	0
8. Humorousness, grotesqueness	0	0	1	0	0
9. Surprise, wonder	1	0	0	0	0
10. Excitement	0	0	0	0	0
11. Longing, desire, yearn	0	0	2	0	0
12. Grief, pain	0	5	0	1	9
13. Resignation, bitterness, pity	3	6	0	8	20
14. Fear, fright	0	0	0	0	0
15. Feeling of outrage, hurt	2	1	0	0	0
16. Reproach	3	1	0	6	0
17. Irony, derision	2	3	0	2	1
18. Refusal, awfulness, dreadfulness	0	1	0	0	0
19. Anger, hate, threat	1	0	0	0	0

(significant at 0·1%), in the second $z = 1{\cdot}757$ (0·5%). As concerns the qualitative characterisations of the expressions it is interesting that each of the two actresses correctly captured one of the two presumed basic emotional tendencies. In fragment 22 there is a highly significant difference in the no. 4 parameters i.e. *love, tenderness, sensuality* ($P = 0{\cdot}0000404$) and parameter no. 6 i.e. *playfulness* ($P = 0{\cdot}0002256$). By this means the *amorous enticement* is well detected as we expected in this scene. The third significant difference was naturally given by the suppression of the *neutral announcing sentence* where the P-value is of the order of 10^{-9}. In comparing fragment 28 with fragment 18, which corresponds to the sentence before the listening, parameter 13 came out highly significant (*resignation, embitterment*) which we determined on the χ^2 four-fold table as 14·66. This is a very significant value at the 1% level. In this case too the announcing sentence dropped significantly albeil not so noticeably. *Thus these fragments amply verified the presupposed emotional content of the given extract.*

Experiment 23 ended in a failure. Though it was a sentence after listening to the music its distribution did not differ significantly from the announcing sentence no. 17. This happened

because the actress was not satisfied with the first realization, in which, though she caught the emotional content, she unwittingly half-sang this sentence under the influence of the music (this is the only case in all our extensive material where anything of the sort occurred). She immediately tried in great haste, annoyance and a touch of stage-fright to put it right, but failed. The actress expressed her feelings as described above by the words on the tape — „to je průšvih" ("that's a nuisance").

As a conclusion to these experiments we may say that the method of psychoacoustic transformation is quite demonstrably appliable in the study of the emotional expression *since the probability of the chance emergence of these results is negligible.* Moreover there is the advantage that demonstrable results can be obtained with quite a small number of experiments using only a few interpreters.

We must of course always be prepared for a certain variability [22]. This stems from the different experiences which the test subjects have lived through, from their moods of the moment and from their varying capacities for the reception of music. Thus there is the subjective factor at play here. We must emphasize that everything we have introduced here has only validity as a probe and by no means a final objective. Further control work must be carried out on wider material and we are currently engaged on a thorough and systematic verification which has been unthinkable of course without an all-round testing of the methodology.

4.2 Differences in the Objective Sound Patterns Related to Changes in the Expression

4.2.1 *Demonstration of Differences in Sonagrams*

We are of course not at present able to present a system of changes in speech or music production evoked by different emotional states. But it is quite possible to demonstrate on the basis of many examples that differences in the expression which can be statistically proved with a high significance are well marked in the acoustic picture of sentences even without precise measurement. The *acoustic differences* arising in this way can be considered as a direct and immediate result of psychoacoustic transformation, or de facto as a *model of the corresponding emotional experience.* As we have mentioned above we do not try to transfer the psychological and physiological realities into physical ones; we are simply tracing their mutual connections.

At the top of figure 2 (2a) we find the sonagram of the sentence "jako by sem smrt načuhovala" ("as if Death were peeping in") which was characterized by the observers as an *indifferent announcing sentence.* The sonagram at the bottom (2b) represents the same sentence pronounced by the same actress but with a strong expression of *fear* (68% of observes). The Kolmogorov-Smirnov distribution test showed in the comparison of the variance between the two sentences a very significant difference, the z being 3.282 which is a significant difference at the level of probability 0.1%.)

At first glance both sonagrams are quite dissimilar in their duration. The time for the expressive

sentence is much longer and also the rhythmic division is more rich, having a long expressive break in the middle, followed by a heavy, loud inhalation.

The upper sentence is pronounced with a clear voice (this is shown by the horizontally well defined formant bars). In the expressive sentence the voice is clear in the first half but in high components the noisiness is very much increased. In the second half of the sentence a sudden break occurs in the rythm, melody and timbre of the voice; the regular striation of formants disappears and the voice breaks into a wheezy whisper. The sound energy is shifted into much higher frequency regions. The tempo is slower and there is greater distinctness in the pronunciation. The example naturally cannot serve as a concrete proof, but only for demonstration of methods.

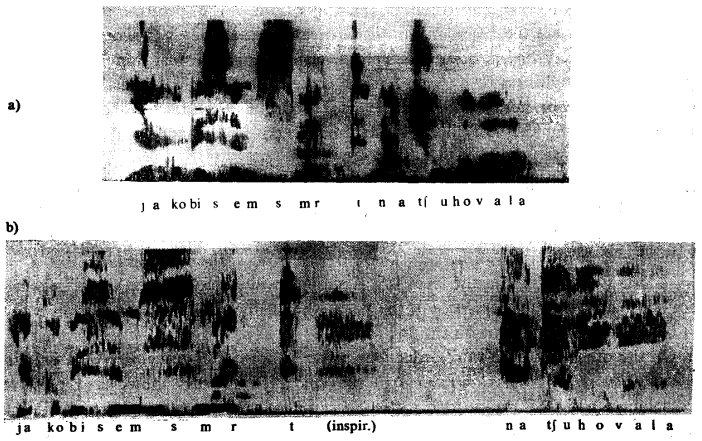


Fig. 2.

We see that in objective parameters so many factors and phenomena appear that we cannot connect them to precise emotional experiences nor assess the degree of relevance mechanically, but only on the basis of more extensive material. Some of the procedures in this documentation will be spoken about in the following paragraphs.

4.3 Relations between the Acoustic, Physiological and Psychological Parameters of the Emotional Expression in Speech

In our study published in Folia Phoniatrica in 1963 [23] we tried to demonstrate that the expression of joy or grief is closely related to the mean pitch of the phrase. Our supplementary revision of the same material has proved that the correlation index according to the test of Kendall is +0·66, which is a value significant on the level of 1%.

We must remark that in accordance with our earlier investigations, the most important factor is the mean pitch of the voice. To a certain extent nearly all the articulatory and phonatory tendencies which we have studied are somewhat related to this parameter. Such definite correlations do not occur very often. Constant melodic or melodico-rhythmical patterns are much less frequent. Complicated vectors of acoustic parameters are usually at play.

When studying these vectors we have to be aware of the fact that the speaker communicating emotions is not able to control immediately such acoustic parameters as the pitch and intensity of formants, the upper limit of the spectrum etc. But what he does control are the *basic articulatory*, or as the case may be, *phonatory intentions* or let us say *instructions* of which in our opinion the most important for expression are: brightness or darkness of the voice, softness or harshness of the articulation, purity or hoarseness of the voice, further liveliness or monotony of the melodic or rhythmical movement, smilling or weeping voice etc.

At first the correlations between these instructions and the expression of joy or grief and love or anger were established on the basis of several groups of observers who registered their subjective feelings which were evoked by 23 different recitations of two sentences. There exists a positive correlation of the rank-order *joy-grief* with the rank-order *brightness-darkness* of the voice. The correlation index $T = +0{\cdot}65$ is significant on the level 1%. A similar positive correlation exists between *joy-grief* and *purity-hoarseness* of the voice, the coefficient of the correlation being $+0{\cdot}45$ (equally significant on the 1% level).

On the other hand the rank-order *love-anger* is not correlated either with *brightness-darkness* or *purity-hoarseness* of the voice. There exists however a positive correlation with the rank-order *softness-harshness* ($T = -0{\cdot}49$ significant on the 1% level). No further correlations of this rank-order with the physiological instruction have been found.

Our second task was to try to define the articulatory and phonatory instructions in terms of acoustic parameters. We publish here only some of the positive findings of our rather extensive investigations.

A very close correlation exists between the mean pitch and the rank-order *brightness-darkness* ($T = +0{\cdot}676$ which is significant on the 1% level). Among other acoustic factors three parameters characterizing the spectrum of the voice are important, namely: the relative intensity of the higher components of the voice (that is the 3rd and 4th formant), the upper limit of the spectrum, and the factor we call the upper limit of the periodicity, that is, the highest harmonic (periodic) tone.

The material we have chosen for the present publication enables us to differentiate between the instructions of brightness-darkness and softness-harshness. The respective results are demonstrated on the Table III. The rank-order *brightness-darkness* shows the positive correlation with the pitch, with the upper limit of the spectrum and the upper limit of the periodicity. No correlation has been found in the intensity of high components. On the other hand, the rank-order *softness-harsh-*

Table III.

Object. / Subject.	Pitch	Intensity of the 3. 4. formant	Upper limit of spectrum	Upper limit of periodicity
Brightness-darkness	+ 0·676 1%	×	+ 0·368 5%	+ 0·359 5%
Softness-harshness	×	− 0·359 5%	×	− 0·312 5%
Purity-hoarseness	+ 0·445 1%	×	+ 0·32 5%	+ 0·32 5%

ness is characterized by the negative correlation with the intensity of the higher sound components and the negative correlation of the upper limit of the periodicity, whereas the remaining acoustic parameters are without any correlation.

The results obtained are in accordance with our suppositions as to the phonatory mechanism. We assumed that we would find a higher positive correlation between the rank-order *purity-hoarseness* and the upper limit of the periodicity which we could not demonstrate on the basis of this material because of the absence of hoarse voices. But we have discovered this dependence in another group of experiments containing phrases with the expression of fear.

These conclusions are naturally valid for the present material only; generalization will be possible only on the basis of more extensive experiments connected with statistical factor analysis. Nevertheless, in our opinion the possibility of defining the emotional expression of speech by means of the acoustic parameters has been thus demonstrated.

4.4 Evaluation of the Long-Interval Spectrum of the Sentence in Relation to the Expression

When studying expression it is necessary in some cases to ascertain, in addition to detailed analysis of individual components, the *general tendencies* which correlate with physiological instructions. We discovered these general tendencies either by statistical processing of detailed results or by direct *electroacoustical integration.* The general changes in the speech spectrum are easy to ascertain by sound analysis over a longer time interval, in our case by the analysis of a whole sentence.

This is the metod used by Ostwald [15] who evaluated the speech of psychiatric patients using half-octave filters. Recently Broeckx and Landrieu tried to apply a similar method to the investigation of instrumental music [3]. We prefer to use third-octave filters which in the formant regions correspond approximately to the critical band width of the cochlea. We used the Bruell &

16 Kjaer analyser with level recorder, and analysed the recitations made on an infinite magnetic loop. The writing speed was chosen so as to ascertain the average value of energy in the individual frequency bands.

For illustration we present examples of the same sentence spoken by the same actress with a different expressive content, and then a sentence from another actress.

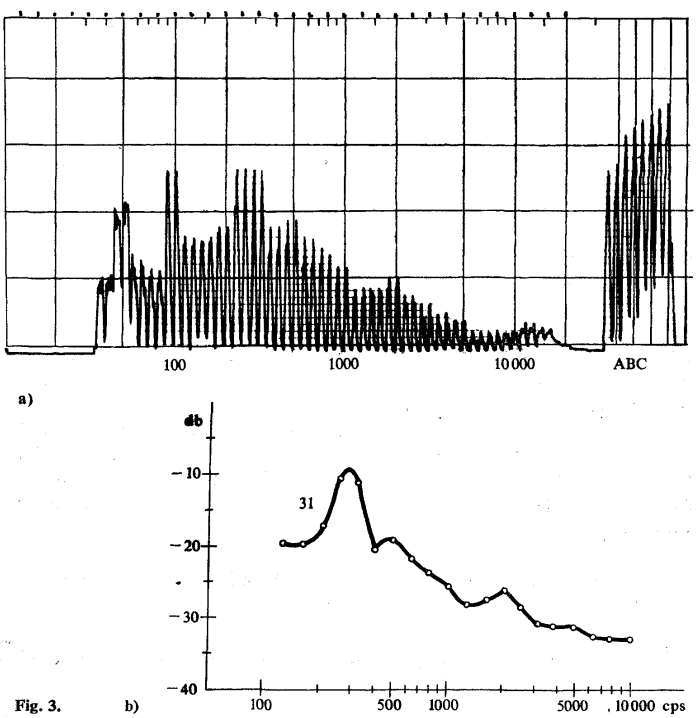


Fig. 3.

The first two examples, no. 31 (fig. 3; 3a – long-interval spectrum integrated by level-recorder, 3b – respective curve corresponding with 3a) and no. 33, (fig. 4), represent the sentence "tož už mám ustlané" (approximate translation: "the bed is made already") from Janáček's "Zápisník zmizelého" ("The Diary of One who vanished"), where the actress produced an intentional expression according to the instructions of the experimenters. The expression prescribed for 31 was *amorous*

enticement with a subtext of *joy*. The emotional content of the sentence was confirmed by a group of observers who evaluated very significantly (on the level 0·1%, according to Poisson theorem – see above) sample no. 31: in the characteristics of both *sensuality* and *joy*. From sample no. 33 in accordance with the instructions came out the characteristics of *enthusiasm*, *joy* and feeling of *force* and *freshness*.

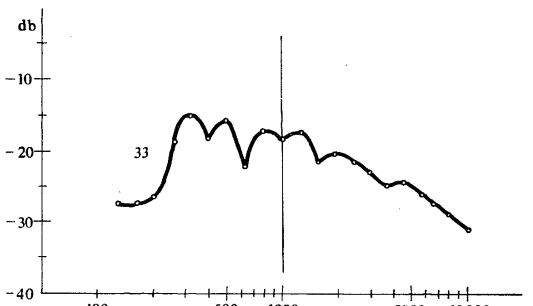


Fig. 4.

As can be seen, both examples are emotionally similar. In both *joy* is expressed to which in example 31 is joined an *erotic character*, whereas example 33 represents *joy* with *sthenic background*.

In the long-interval spectra there are several typical peaks. The first of them belongs to the fundamental tone, then follows a small peak of the second harmonic tone, and then the regions of the two first formants with a small drop between them. In example no. 31 (fig. 3) the second formant has its peak at 2000 cps, the high tones are very slightly indicated, and the spectrum ends at a frequency of practically 3500 cps. It is possible to characterize this spectrum as dropping very steeply with a large loss of high tones.

The spectrum of example no. 33 (fig. 4) shows quite another pattern. The peaks of the formants are only slightly indicated and form a convex curve as opposed to the preceding sample. The high tone content is much larger and also the upper limit of the spectrum is increased to 6000 cps.

The same samples were applied in the experiment which was evaluated from the point of view of articulatory and phonatory instructions, i.e. in the dimensions *brightness-darkness*, *softness-harshness*, *purity-hoarseness* of the voice. Sample no. 31 was defined as significantly *soft*; under the same conditions sample no. 33 came out on the borderline of *harshness*. In the dimension *brightness-darkness*, the timbre of both samples was registered as significantly *bright*.

It would seem rather bold to speak, on the basis of these examples, about difference in expressive content. But we must not forget that firstly both realizations were

made under the same circumstances, differing only in the condition of expression. Secondly the result of this experiment is the same as the result of a large statistical investigation, in which the correlations between the objective parameters of the sound seen on the sonagrams, and the acoustic and phonatory instructions were stated. As can be seen, the rank-order *softness-harshness* correlates negatively with the intensity of the higher components of the spectrum and with the upper limit

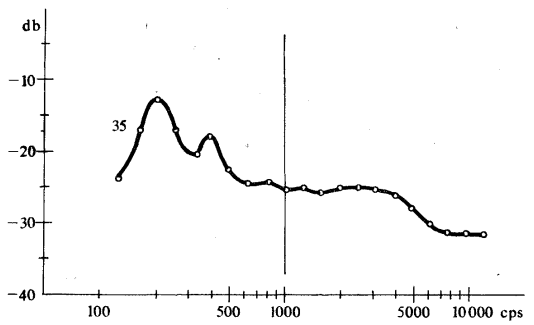


Fig. 5.

of the spectrum. Therefore a hypothesis may be pronounced that the expression of *love-lure* in example no. 31 is shown (as opposed to the *pure joy* in example no. 33) by softening to voice, the acoustic manifestation of which is the loss of intensity in the high tones and the lowering of the upper limit of the spectrum.

Another kind of long-interval spectrum is shown by example no. 35 (fig. 5) representing the same sentence but with a very different expression. The subjective evaluation gave very significantly the characteristics of *irony* and *derision* and also of *scorn*. In the test of phonatory and articulatory instructions this interpretation was characterized as a significantly hard pronunciation. The spectrum shows a very deep fundamental tone, and then the curve runs in the region of both formants almost horizontally at a level between the levels of the curves in the two preceding examples.

We suppose that for an evaluation of the psychoacoustic effect of the sound, the sound structure of the whole sentence is decisive. This is due to the dispersal of energy over the basilar membrane. This factor can be demonstrated if we set all examples to the same level of the peak of the fundamental tone as shown in figure 6. Example no. 31 (*softness, merriness* and *love*) shows a very steep slope with a restriction of high tones; the *merry but sthenic* sample no. 33 has a convex form with a gentle slope, and sample no. 35 expresses *displeasure* and *irony*: the conspicuously horizontal running of the curve should be noted.

For interest's sake only we include sample no. 19 (fig. 7) which was pronounced by another person but with the same expression of *joy* and *merriness* as no. 33

(according to the listeners' judgement). We see that the long-interval spectrum evoked by the same expression though spoken by a different person is of almost the same convex form.

Naturally we cannot consider the facts shown here as a proof of our hypothesis; we merely present them here as a directive for further more extensive investigations.

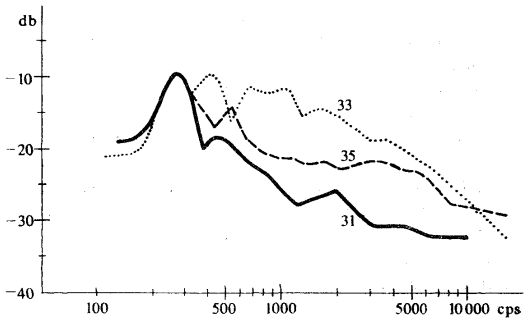


Fig. 6.

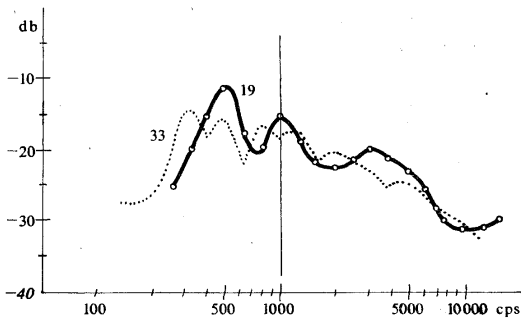


Fig. 7.

5. METHODOLOGY OF THE MUSICAL FACTORS OF SPEECH

When explaining the objective (acoustic and physiological) parameters we have been dealing up to now mostly with the *spectrum of the voice.* We have been paying particular attention to this question on account of the extreme difficulty of finding the regularities in this domain. Altogether we are dealing here with complicated vectors. Over a certain period — after countless measurements of formants and vain

seeking of correlations with the emotional expressions – we were even of the opinion that the formant structure of speech will probably be essentially irrelevant to expression, because full use is made of these parameters for other functions, in particular for the differentiation of speech sounds and of the individual timbre of the speech act. It was not until we turned our attention to the general expressive tendencies, phonatory and articulatory instructions as we designated them in the preceding paragraphs, that we began to discover positive results.

We found less serious obstacles in our way when we where investigating the other objective parameters of expression, i.e. the *melodic, rhythmical* and *dynamic factors* which correspond to the basic acoustic qualities of sound, frequency, duration and intensity. Scheman most suitably calls these factors *musical factors of speech* as these are basic elements exploited above all by music. In investigations of these factors in connection with expression it was until recently supposed that it would be possible directly to ascribe specific *constant melodic, rhythmical and dynamic patterns* to specific emotions. Trojan [29] and other scholars have come to a sceptical standpoint for they have discovered that the same emotions are frequently expressed by different melodic patterns and on the contrary different emotions by the same pattern. For these reasons we have assumed beforehand that *general tendencies* such as liveliness (as far as marked differentiation) or monotony of the melody, dynamics or rhythm, predominant rising or falling movement, crescendo or decrescendo, accelerando or ritardando, etc., will be more important than constant melodic and rhythmical patterns. To put it more precisely, we considered that the *constant patterns will be applied during expression though in the background of these tendencies.*

The experiences which we have so far acquired during the concrete analyses largely confirm these presuppositions. Thus for example the expression of fright in the sentence "As if Death were peeping in" was in a number of cases expressed by a specific melodic contour, that is to say a rise, a drop then a gradual fall. However the leap alluded to earlier, i.e. a melodic, dynamic and rhythmical break in the voice, is more important than this contour.

These basic general tendencies have the advantage that they can be expressed, by and large, by *quantitatively mathematically simple formulae.* Therefore we have tried to define these most elementary tendencies thus they are the following quantities.

A. Quantities from the Pitch of the Syllables

1. *Register* is the average voice pitch in the segment under consideration. In the first approximation we can define it as the average of the pitch of the individual syllables if there are no great differences in their duration

$$\bar{v} = \frac{\sum_{1}^{n} v_i}{n},$$

where: v_i means the pitch of the i^{th} syllable, n = the number of syllables.

2. *Compass* is the range of the melody and is given by the difference between the highest and lowest tones

$$a = v_{\max} - v_{\min} .$$

3. *Standard deviation* expresses the vacillation of pitch within the given compars around the average voice register irrespective of the sequence of the tones

$$\sigma_v = \frac{\sqrt{[\sum (v_i - \bar{v})^2]}}{n - 1} .$$

B. Quantities Defined by the Differences between Adjacent Syllables (measured by the number of semitones)

4. *First order difference*

$$d_{1i} = v_{i+1} - v_i .$$

5. *General tendency* (rising, falling) of the melody is defined as the sum of positive and negative differences

$$S = \sum_1^n d_{1i} .$$

Practically we fix this value simply as the difference between the first and last tone of the melody

$$S = v_n - v_1 .$$

6. *Steepness* (upwards and downwards) per one step is given by the value of the general tendency divided by the number of steps

$$s = \frac{v_n - v_1}{n - 1} = \frac{S}{n - 1} .$$

7. *Greatest difference* (greatest leap)

$$d_{1i\,\max} .$$

8. *The melodic movement per one step* is the statistic evaluation of all the intervals occuring in the melody; both, positive and negative differences are equally evaluated. For this purpose the quadratic mean was chosen. In this way is measured the vacillation of the melody from syllable to syllable in their sequence (in the sense of digrams) irrespective of changes in direction. The formula is the analogy of the effective value

$$d_{1\text{eff}} = \frac{\sqrt{(\sum_1^{n-1} d_{1i}^2)}}{n - 1} .$$

9. *Second order differences* are basic elements. They are derived from first order differences (analogically as the first order differences are derived from syllable pitch). The number of second order differences is then too less than the number of syllables

$$d_{2i} = d_{1(i+1)} - d_{1i}\,.$$

We consider it useful to define within this system only one quantity which is formed similarly as the melodic movement per one step in first order differences. We have called it:

10. *Liveliness including marked differentiation of the melody.* It is the statistic evaluation of the changes in the interval steps in the trigram sequence. Here are evaluated changes from small intervals to large and vice versa, and especially changes in the direction of the melody (this is the analogy of the second derivation)

$$d_{2\text{eff}} = \frac{\sqrt{(\sum_{1}^{n-2} d_{2i}^2)}}{n-2}\,.$$

This value, unlike the preceding one, ignores the size of the movement; thus for example with smooth chromatic upward trend it would equal zero whereas it reacts very sensitively to any directional rippless in the melody.

We consider that these relations, derived from the pitch of the individual syllables, are quite adequate for the description of the basic tendencies occuring in sentence intonation. By way of illustration we now present some of the characteristic types.

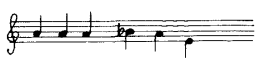
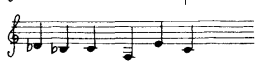
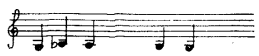
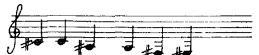
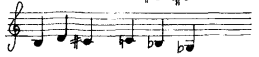
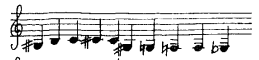
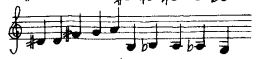
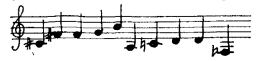
A *monotonous melody* is marked by a specific voice register whereat all the other characteristics converge towards zero or are in fact zero. In Table IV no. 34 corresponds to this type. It was evaluated by the listeners as a neutral *announcing sentence* with a precise subtext of *resignation* and *embitterment*. In a comparison with example 17 which is a simple *announcing sentence* we see that in 34 the monotony is highly graduated which in itself is a sign of emotional intent. On the other hand 17 nevertheless shows in spite of its small compass and dispersion, certain values tending towards liveliness which, however, must be considered as expressively irrelevant.

A *smooth rising* or *falling melody* will be expressed by the compass and the standard deviation; mainly however it will be expressed by the value of general direction, negative with a falling and positive with a rising melody, and by marked upward or downward steepness per one step. The melodic movement per one step will be close to the value of the steepness per one step. This means that the whole melodic movement is exhausted by the upward or downward steepness, and what is important, the liveliness of the melodies tends towards zero. Sentence 28 in our table is a sample of a falling melody, though of course it is not a smooth movement,

therefore liveliness per one step reaches more or less the same value as in the announcing sentence.

One distinct leap in the melody is manifested in the following way: though the values of movement and liveliness are not very high, upward and downward steepness is clearly expressed. An important characteristic of this melodic type is the fact that with a maximal leap (D_{max}), values of compass and upward or downward steepness are virtually exhausted. Example 19 in the table represents this type.

Table IV.

		$\bar{v}$	a	σ_v	S	s	$d_{1\text{eff}}$	$d_{2\text{eff}}$	$d_{\max}$
19		20·2	6·0	2·1	−5·0	−0·8	2·3	2·3	− 5·0
31		11·2	11·0	3·2	−1·0	−0·1	6·1	12·5	10·8
34		7·3	1·0	0·6	0	0	0·4	0·5	0·7
28		9·4	8·0	2·9	−8·0	−1·3	2·1	2·6	− 3·4
17		11·2	6·0	2·1	−4·0	−0·6	2·2	2·2	− 3·3
1		8·8	8·0	3·1	−3·0	−0·3	2·2	2·5	− 4·6
14		13·4	14·0	4·7	−8·0	−0·8	3·7	5·6	−10·0
11		14·4	17·6	4·8	−8·0	−0·8	5·9	9·2	−14·0

Marked differentiation of the melody is always characterized by a high value of liveliness $d_{2\text{eff}}$ and as a rule of movement, though liveliness is highly predominant. In this case also the compass may be roughly equal to maximum, but if there are more leaps than one in the melody this is shown by excessive liveliness. In the table this type of melody is represented by example 31.

In some cases these tendencies themselves may very clearly distinguish the expressive sentences after listening to the music from the announcing sentences before the music. If we compare sentence 1 of the Table IV (before the music) with

no. 14 (after the music) and produced by the same actress, both of them have already been classified as to expression in paragraph 4.1 on page 7, we discover that in the sentence after hearing the music certain fundamental changes have taken place. The voice register is raised, the compass expanded from a minor sixth to a major ninth, and the leap, which originally took approximately a fourth, now reaches a minor seventh. Accordingly the value of marked differentiation of the melody is also increased, though not so markedly as in example 31 because it is given only by a single leap. Thus we may say that in this case too the melodic break is relevant to the expression. A further gradation of these tendencies is to be seen in example 11 which was subjectively evaluated as substantially more expressive. The leap is here increased to a major ninth and the compass to an octave and a half. The values of the melodic movement and marked differentiation of the melody are also increased, though even here they do not reach the level of example 31. This is because in no. 11 the single melodic break is decisive whereas no. 31 has the marked differentiation of the melody as the main characteristic.

We shall be a in position to offer a final evaluation of this system when and only when the *statistical norms* of the individual quantities become known. Till now we have been merely concerned with the description of examples. However, in support of the system is the fact that the *interpretation of the differences discovered did not create any great difficulty*. In our opinion it will be possible to proceed in an analogical way in the analysis of *dynamics and rhythm*, for the dynamic and metrico-rhythmical patterns are not unlike the melodic patterns.

6. CONCLUSIONS

After an analysis of the methods used for the study of expression, i.e. the method of direct introspective analysis, behaviourist methods and finally the method of the objective investigation of biological reactions, the authors suggest a new, combined method which they call the method of psychoacoustic transformation and which they consider to be a useful complement to approaches employed hitherto.

The principle of the method lies in the comparison of a dramatic interpretation delivered under the influence of music with the same text spoken before the musical extract was heard. The interpretations are evaluated both by subjective tests (groups of listeners) and objectively (by acoustic analysis). By comparison of the two interpretations, psychical and acoustic differences are stated, enabling us to establish correlations between acoustic and psychical phenomena.

The advantage of this approach is that it enables us:

a) to investigate a spontaneous emotional reaction without needing to carry out a semantic analysis;

b) to exclude individual phonological and phonetic differences.

c) to make the comparisons always at the same level, i.e. acoustic, physiological, psychical etc. (a differential method);

d) to make the confrontation of the 2 systems, music and speech, and study their common regularities.

From among the numerous statistical methods we found best suited to our investigation the Kolmogorov-Smirnov test of goodness of fit and the test of the χ^2 four-fold table, supplemented according to requirement by calculations of exact probability with low frequencies. In the analysis of multiparameter questionnaires with low expected frequencies it was found most appropriate to approximate the binomical distribution by the Poisson theorem. In this way we discovered that the recitations after listening to the music differed significantly from those before listening. It was thus also possible to state statistically which emotional characteristics are typical for the individual fragments. In the great majority of cases it was demonstrated that the results of the statistical investigation of musical fragments agree significantly with the composer's intention.

Further we weighed the value of certain objective methods used to investigate the acoustic properties of expressive speech signals. From among the basic acoustic parameters it is necessary to register and study the pitch of the sound, its intensity, time factors (rhythm and tempo) and the spectrum structure both summarily and in detail. Among the apparatus currently available only the visible speech type of spectrograph gives usable results for the detailed analysis.

It is possible to discover the connection between the distant levels of acoustic signal and psychical actions if attention is paid to the basic physiological, articulatory and phonatory instructions (such as brightness-darkness, softness-harshness, purity-hoarseness, calmless-liveliness, monotony-marked differentiation of the melody, etc.).

By attributing acoustic parameters to types of expression the overall phonatory and articulatory tendencies have greater validity that the constant melodic, rhythmical and dynamic patterns.

The overall tendencies can in some cases be even defined mathematically which the authors in fact did especially in the domain of musical factors of speech namely the concepts register, compass, standard deviation, overall tendency (rising and falling) of the melody, melodic movement (per one step), liveliness (or differentiation) of the melody, and the greatest leap. These quantities can be applied, in addition to melody, to dynamics and rhythm.

On the basis of the physiology of hearing the authors have attempted to define the relationship between the physiognomical and semantic expression, where the former has a predominantly analogical mode of coding and transmission and the latter predominantly digital.

(Received August 7th, 1968.)

REFERENCES

[1] Амосов, Н. М.: Моделирование мышления и психики. Киев 1965.
[2] Ashby, W. Ross: Kybernetika. (Czech Translation) Praha 1961.

[3] Broeckx J. L. and Landrieu W.: Electrotechnical Analysis of the Structure of the Tessitura in Beethoven's Symphonies. Publications of the Seminar of Musicology of Ghent State University, No. 2, I.P.E.M. 1967.
[4] Čistovičová, L. A.: Průběžné rozpoznávání řeči člověkem. (Czech Translation) See [28].
[5] Fónagy, I.: Karel Sedláček and Antonín Sychra, Hudba a slovo z experimentálního hlediska. (Book-Review). Studia Musicologica Academiae Scientiarum Hungaricae *8* (1966), 413.
[6] Frank, Helmar: Théorie informationnelle de la réalisation et perception dans l'art des mimes. Cahiers d'Etudes de Radio-Télévision *24* (1959), 377—387.
[7] Hevner, Kate: Expression in music: a discussion of experimental studies and theories. Psychol. Rev. *42* (1935), 186—204.
[8] Hevner, Kate: Experimental studies of the elements of expression in music. Amer. J. Psychol. *48* (1936), 246—268.
[9] Charvát, Josef: Zur Frage Stress-bedingter somatischer Schäden. Endokrinologie *45* (1963), 1—2, 40—52.
[10] Cherry, E. C. and Halle, M. and Jakobson, R.: K logickému popisu jazyků v jejich fonologickém aspektu. (Czech Translation) See [28].
[11] Klaus, Georg: Semiotik und Erkenntnistheorie. Berlin 1963.
[12] Knobloch, Ferdinand and Poštolka, Milan and Srnec, Jan: Musical Experience as Interpersonal Process. A Contribution to the Semantics of Music. Psychiatry Journal for the Study of Interpersonal Processes *27* (1964), 259—265.
[13] Kolmogorov, A. N.: Automaty a život. In Rovenskij, Ujomov, Ujomovová: Stroj a myšlení. (Czech Translation) Praha 1962.
[14] Mukařovský Jan: Kapitoly z české poetiky. Svoboda, Praha 1948.
[15] Ostwald, Peter, F.: Sound Making: The Acoustic Communication of Emotion. American Lec. Communication, Springfield, Illinois 1963.
[16] Palágyi, Melchior: Naturphilosophische Vorlesungen. II. Ausgabe, Leipzig 1924.
[17] Riessauw, A. M.: Verbe et Mélodie. Etude sur les rapports entre le matériel phonétique et mélodique dans la composition musicale sur un poème de Verlaine. Publications of the Seminar of Musicology of Ghent State University, No. 2, I.P.E.M. 1967.
[18] Ringbom, Nils-Eric: Über die Deutbarkeit der Tonkunst. Helsinki 1955.
[19] Roth, Josífko, Malý, Trčka: Statistické metody v experimentální medicině. Praha 1962.
[20] Sedláček, Karel: Základy audiologie. Praha, 1956.
[21] Sedláček, Karel: Informationsinhalt der Sprachaudiometrie. Zeitschrift für Laryngologie-Rhinologie-Otologie und ihre Grenzgebiete *46* (1967), 577—583.
[22] Sedláček, K., Sychra, A.: Hudba a slovo z experimentálního hlediska. Praha 1962.
[23] Sedláček, K., Sychra, A.: Die Melodie als Faktor des emotionellen Ausdrucks. Folia phoniatrica *15* (1963), 89—98.
[24] Sedláček, K., Sychra, A.: Studie poznatelnosti emocionálního výrazu v hudbě u různých skupin obyvatelstva podle zaměstnání (In press).
[25] Schaff, Adam: Wstęp do semantyki. Warszawa 1960.
[26] Sychra, Antonín: Estetika Dvořákovy symfonické tvorby. Praha 1959.
[27] Sychra, Antonín: Kapitola z experimentálního výzkumu výrazu v hudbě a řeči. Acta Universitatis Carolinae, Philosophica et Historica *2* (1962), 49—87.
[28] Teorie informace a jazykověda. Ed. Doc. Lubomír Doležel, ČSAV, Praha 1964.
[29] Trojan, Felix: Der Ausdruck von Stimme und Sprache. (Eine phonetische Lautstylistik.) Wiener Beiträge zur Hals-, Nasen- und Ohrheilkunde, Band I, (Wien), 1948.
[30] Woodworth, Robert, S., Schlosberg, Harold: Experimentálna psychológia. (Slovak Translation) Kap. Emócie I—III, Bratislava 1960.

Metoda psychoakustické transformace použitá ke studiu exprese v řeči a hudbě

KAREL SEDLÁČEK, ANTONÍN SYCHRA

Autoři po rozboru metod, užívaných ke studiu exprese, a to metody přímé introspektivní analýzy, dále metod behavioristických, a konečně metody objektivního zkoumání biologických reakcí, navrhují novou vlastní kombinovanou metodu, kterou nazývají metodou psychoakustické transformace, a kterou považují za užitečný doplněk k dosavadním postupům.

Princip metody spočívá v tom, že jsou srovnávány herecké interpretace přednesené pod vlivem hudby s týmiž texty, přednesenými a vyslovenými před expozicí hudebního fragmentu. Interpretace jsou hodnoceny jednak subjektivními testy (soubory posluchačů), jednak objektivně. Srovnáváním rozdílů obou způsobů hodnocení jsou nalezeny korelace mezi akustickými a psychickými fenomeny. Přednost tohoto postupu je v tom, že umožňuje:

a) zkoumat spontánní emocionální reakci, aniž je třeba sémantické analýzy;

b) vyloučit individuální, fonologické a fonetické diference;

c) provádět srovnání vždy na téže rovině, tj. akustické, fyziologické, psychické atd. (metoda diferenční);

d) konfrontovat oba systémy, hudby a řeči, a studovat jejich společné zákonitosti.

Z četných statistických metod se v tomto šetření osvědčil test dobré shody Kolmogorova-Smirnova a test čtyřpólní tabulky χ^2, doplněný podle potřeby výpočtem přesných pravděpodobností při malých četnostech; při analýze multiparametrických dotazníků s nízkou očekávanou pravděpodobností je nejvhodnější aproximovat bionomické rozdělení Poissonovým rozdělením. Tímto způsobem autoři zjistili, že se ukázky po poslechu hudby signifikantně liší od ukázek před poslechem, a otevřela se také možnost statisticky dokumentovat, které emocionální charakteristiky jsou příznačné pro jednotlivé ukázky. V převážné většině případů bylo prokázáno, že výsledky statistického zkoumání hudebních ukázek souhlasí signifikantně se záměrem skladatele.

Autoři zvážili dále metodu některých objektivních metod, zkoumajících akustické vlastnosti expresivních řečových signálů. Ze základních akustických parametrů je nutno registrovat a studovat výšku zvuku, intenzitu, časové faktory (rytmus a tempo) a spektrální stavbu, a to jednak sumární, jednak detailní. Při detailní analýze dává z přístrojů, které jsou dnes dostupné, použitelné výsledky jedině spektrograf typu „visible speech".

Vzdálené úrovně mezi akustickým signálem a psychickými ději lze účelně překlenout, věnuje-li se pozornost základním instrukcím fyziologickým, artikulačním a fonačním, jako je jasnost-temnost, měkkost-tvrdost, čistota-zastřenost, monotonie-rozvrásněnost atd., a psychofyziologickým rozměrům, jako je libost-nelibost, jistota-nejistota, kladné-záporné hodnocení aj. Tyto psychofyziologické rozměry považují autoři za základ nového systému emocí, na němž pracují.

Při přiřazování akustických parametrů k expresivním typům se více uplatňují celkové fonační a artikulační tendence než konstantní melodické, rytmické a dynamické modely.

Celkové tendence je možné v některých případech definovat i matematicky. Bylo tak učiněno autory zejména v oblasti musických faktorů, kde byly definovány pojmy poloha, ambitus, standardní odchylka, celková směrová tendence melodie, melodický pohyb, oživenost resp. rozvrásněnost a největší skok. Tyto veličiny lze aplikovat vedle melodie i na dynamiku a rytmus.

Na základě teorie slyšení se autoři pokusili vyjádřit vztah mezi fyziognomickou a sémantickou expresí, a to tak, že ve fyziognomické expresi převládá analogový způsob kódování a přenosu, kdežto v expresi sémantické hraje rozhodující úlohu digitalizace.

Prof. MUDr Karel Sedláček, ORL klinika FVL UK, U nemocnice 2, Praha 2;
Prof. Antonín Sychra, DrSc., Katedra estetiky filosofické fakulty UK, Celetná 20, Praha 1.