RECOGNITION OF EXPECTED LISTENERS' EMOTIONS BASED ON THE ARTIFICIAL INTELLIGENCE METHODS

K. Sapozhnikova, S. Hussein, R. Taymanov, I. Baksheeva

Abstract: The paper deals with the methods for improving the accuracy of emotion recognition as applied to the development of the instrument measuring an expected emotional impact of music compositions. Taking into account the "sincerity" of non-human animals, the investigation used their bioacoustic signals that were classified by a neural network.

Keywords: emotions, bioacoustic signals of animals, model for measuring emotions, artificial intelligence, neural network

1. Introduction

For centuries, the mystery of the music influence on listeners has agitated scientific researchers. They have long understood that the origin of music is conditioned by the desire to expand the possibilities for the transmission of emotions. Since vocalizations of humans can be under their voluntary control [1], it is natural that there have been attempts to uncover the mystery of music by analyzing the bioacoustic signals of non-human animals. The animal's emotions are recognized in accordance with the situation that caused them.

The development of audio recording devices has led to an increase in the number of studies related to emotional bioacoustic signals [2-4 and many others].

An animal population collection requires organizing communication between the population members. Signals corresponding to proto-emotions, which are vital for the population, i.e., the signals motivating certain actions of population members, could be transmitted only by changing the pressure of the environment because the touch is inherent in all the living creatures.

The genesis of life and of the first populations in ocean, limited the frequencies of the transmitted signals corresponding to the basic emotions to the infrasonic range. The number of such emotions is determined by the compromise between the need not only to transmit and receive important information for the population, but also to quickly recognize it in order to have sufficient time for adequate reaction.

In the course of evolution, proto-emotions have been kept as emotiogenic neurophysiologic reactions (for mammals they correspond to the brain biorhythms).

When living creatures mastered the air environment and started forming populations in large territories, to expand the frequency range of emotionally colored signals into an audio range became necessary. The corresponding range should have been effective for transmitting information depending on the distance to the addressees: if they were far away, using its high-frequency part was necessary, if addressees were nearby, its low-frequency part was appropriate.

To transmit specific emotions, to modulate sound oscillations with the emotiogenic reactions, turned out to be sufficient. To demodulate these reactions while receiving signals is necessary.

The number of mammals' emotions has been increased. The emotions that are common for humans and non-human animals (basic ones) are determined by the simplest combinations of neurophysiologic reactions characterized by the frequencies of brain biorhythms (delta, theta, alpha, and beta biorhythms). At present, the generally accepted list of basic emotions does not exist [5].

Until recently, the scientific instrumental studies in the field of musicology, were usually limited to the analysis of sound spectrum of various musical instruments and specific characteristics of musician performances. Mainly, emotions caused by music were assessed by listener's verbal self-reports.

The application of metrological analysis has opened new prospects. In a number of works, including those carried out within the framework of the grant provided by the Russian Foundation for Basic Researches, a model was proposed and justified for measuring the expected emotions that arise while listening to acoustic impacts including music excerpts [6–10 and others].

It differs from other models in the fact that it reflects the structure of forming emotiogenic neurophysiological reactions in humans and non-human mammals, which cause emotions being recognized. Basic emotions are the output of the 1st model step. In the 2nd step, the basic emotional images that have a general human character, are generated, and in the 3rd step, the formation of culture dependent emotional images, takes place. They are associated with the cognitive human activity, upbringing, education, national history, as well as belonging to a certain socio-cultural group. The model provided an opportunity to understand the fundamentals of forming complex emotions as certain sets of the basic ones, to highlight the signs of expressiveness of music, outline ways to improve the music scale, etc.

It also enabled starting to develop a specialized measuring instrument designed for composers and music performers to facilitate the search of such music interpretations that would be perceived as expressive by certain categories of listeners, differing, for example, in age, national culture, etc.

However, using previously accepted data processing methods [6-10 and others] for the automatic recognition of emotions and then of feelings being caused by emotion collections in the course of music sounding, has resulted in certain difficulties. Methods of artificial intelligence, which provide an increased speed of the image recognition, proved to be more efficient.

The first stage of development is aimed at automatic recognizing some emotions reflected in non-human animal's voices. This task is more simple, but has an independent meaning for a number of applications.

2.Bioacoustic signals and emotion features applied for investigation

For analysis, files given in Table 1, which characterize emotions considered as the basic ones, were chosen. From each file, the fragments were selected with the durability of (1-2) s. To achieve better frequency resolution, they were modified by extending up to 5 s with the help of additional pauses.

In the process of analysis, formation of emotiogenic neurophysiologic reactions while listening to the modified fragments, was under consideration. The signals corresponding to these sounds enter an input of the aforementioned measurement model. On the first stage, the summation of the input signal with its delayed version (time delay is 0.2 s) in time domain takes place. That stage is followed by a non-linear transformation according to (1). In previous research [11], it was found that such a transformation type can acceptably model this part of the neurophysiologic process.

$$y(x) = \left(x + |x_{\min}|\right)^3 \qquad (1),$$

where y is the value of the signal after non-linear transformation, x is the sum of the input signal with its delayed version, $|x_{min}|$ is the module of the minimal value of x.

The role of the member $|x_{min}|$ is to exclude the minus sign of the transformed signal.

On the next stage, the Short-time Fourier transform is carried out. After that, the spectrum of oscillations within the range (0.6 – 20) Hz are singled out. They are distributed according to the bands (sub-bands) related to the brain biorhythm frequencies.

To facilitate the analysis, the total energy of oscillations within this range was accepted equal to 1, while the biorhythm energy was normalized to it. In the first version of the investigation, the energy was distributed within 8 sub-bands (delta 1 (0.6- 2) Hz, delta 2 (2-4) Hz, theta 1 (4-6) Hz, theta 2 (6-8) Hz, alpha 1 (8-10) Hz, alpha 2 (10-12), beta 1-1 (12-16) Hz, beta 1-2 (16-20) Hz), while in the second version, 4 bands were taken (delta, theta, alpha, beta 1). The sub-bans (bands) with a significant energy share (i.e., biorhythms being activated) were singled out.

Assessing the connection between the biorhythms and emotions, it was accepted that activation of delta-rhythms testifies to threat, theta-rhythm - to perception of something unexpected, alpha-rhythm - to arousal, beta 1 – to concentration of attention [14].

Besides, an approximate assessment of the sounding addressee was provided, which is determined by the pitch of the acoustic frequency zone where the greater part of the energy of the modified fragments of bioacoustic signals is concentrated: lower than 500 Hz or higher than this limit.

Combination of sub-bands (bands) of activated biorhythms with the above mentioned pitch represents a set of features that, in case of rather short soundings, is expedient to use for determination of emotions, including the basic ones. The approach based on the concept of multidimensional quantities differs from the most widely spread concept on the emotion scale including only two qualitative features (valence and the level of arousal) [3], opens perspectives for automatic recognition of emotions.

Here, the necessity arises to define an emotion definition more exactly and to show the difference

Table 1. Bioacoustic signals of mammals and their interpretation

Animal	Zoologists' comment	Emotion accepted as the basic one with comments
Speckled ground squirrel *Spermophilus suslicus* [12]	Alarm calls	Interest/ alarm, information for many receivers
Yellow ground squirrel *Spermophilus fulvus* [12]	Dramatic scream of the yellow ground squirrel, taken from the burrow	Fear/ information for many receivers
Black-backed jackal *Canis mesomelas* [12]	Black-backed jackal howling Black-backed jackal howling	Sadness/ call of male (female) animal, information for many receivers
Dhole *Cuon alpinus* [12]	Heart-freezing howl of a dhole pair	Sadness/ call of male (female) animal, information for many receivers
Cheetah *Acinonyx jubatus* [12]	Howl of female animal	Sadness/ call of male (female) animal, information for many receivers
Clouded leopard *Neofelis nebulosi* [12]	Howling of a leopard	Sadness/ call of male (female) animal, information for many receivers
Far-East leopard Panthera pardus orientalis [12]	The threatening hisses of a young female toward a human	Anger/ threat, information for a specific addressee
Snow leopard *Uncia uncia* [12]	The growls and deep breathes; the snow leopard is threatening	Anger/ threat, displeasure, information for a specific addressee
Felis_viverrinus *Felis viverrinus* [12]	Loud threatening hiss toward an approached human	Anger/ threat, information for a specific addressee
Sable *Martes zibellina* [12]	The threatening sounds of an adult animal	Anger/ threat, displeasure, information for a specific addressee
Cheetah *Acinonyx jubatus* [12]	Aggressive sounds of a cheetah: the hiss	Disgust/ displeasure, information for a specific addressee
Cheetah *Acinonyx jubatus* [12]	Aggressive sounds of a cheetah: the growl	Anger/ threat, information for a specific addressee
Yellow ground squirrel *Spermophilus fulvus* [12]	The threatening chirr	Anger/ threat, information for a specific addressee
Dhole *Cuon alpinus* [12]	Calls during friendly interactions in a pack	Happiness/ pleasure with a high level of arousal, information for many receivers
Chinchilla [13]	Playful animal	Happiness/ pleasure with a high level of arousal, information for many receivers
Cheetah *Acinonyx jubatus* [12]	The comfort purring of an adult female	Tenderness/ pleasure with a low level of arousal, information for a specific addressee

between emotion and feeling. Mainly, emotion is considered as a short reflectory reaction that is required to preserve and develop population. For developed mammals and human, reactions related to some ethical norms, forming families, and so on, e.g., concerning origination of tenderness to a little one. Such reactions can be of a prolonged character and can be considered as intermediate, between the emotion and feeling.

2. Artificial intelligence as an instrument for recognizing emotions

The analysis performed has enabled the transition to the next stage that is learning of an artificial intelligence to recognize the basic emotions. To perform such a procedure, the soundings were applied, for which the analysis of the animals' situation and behavior, structure of activated biorhythms and pitch gave rise to selecting by experts the emotions given in Table 1.

A Radial Basis Function Neural Network (RBFNN), which is a special type of a neural network, is used in this research for classification.

While designing the network it is accepted that possible deviations from an "ideal image" are subjected to the Gaussian function:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad (2),$$

where x is the input, μ is the mean value or median, and Ϭ is the standard deviation.

It follows from this fact that the efficiency of the classification depends on the accuracy of the Gaussian function determination in a hidden layer. At that, training of one neuron cannot affect the other neurons while modelling any type of non-linearity with the right selection of neurons number in the hidden layer.

The RBFNN architecture consists of three layers:

- the first layer contains an input vector that represents the feature vector;
- the second one contains a hidden layer of RBF neurons with a number of neurons with Gaussian functions that can increase or decrease in numbers to provide acceptable error of classification (recognition);
- the third (output) layer contains one node per each type of emotions being recognized.

The neural network performs classification by comparing the features of input examples and those from the training set. Each RBFN neuron stores a "prototype", which is just one of the examples from the training set. When a new bioacoustic input needs to be classified, each neuron computes the Euclidean distance between the features characterizing the bioacoustic signal and its prototype (the estimate of the difference between the input vector and vector accepted as a "standard" that characterizes an ideal notion of a corresponding basic emotion). In other words, if the input more closely resembles the class A prototypes than the class B prototypes, it is classified as class A.

Fig. 1 illustrates the operation of the neural network for 5 features and 7 emotions accepted as the basic ones. (However, the list of 7 emotions is not exhaustive.)

Since the input set was not big, for learning of a neural network and subsequent classification, a standard procedure of K-fold cross-validation was applied.

The results of investigations, carried out with the animals' soundings applied in Table 1 are given in Table 2.

As it follows from the results obtained, to provide the minimal error for 9 features (if the range is divided into 8 sub-bands) it was necessary to realize 24 neurons in the hidden layer. For 5 features, only 17 neurons were required, the probability of correct classification for threat even increasing in the second version.

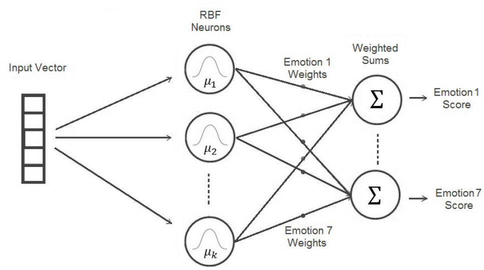

Fig. 1. Neural network for 5 features and 7 emotions accepted as the basic ones

4. Conclusions

In the 21st century, it became possible and desirable to develop instruments assigned for measuring multidimensional quantities that influence human or animal behavior. Such quantities include emotionality of music, "warmth" of musical instrument sounding, freshness of perfume aroma, evoking drink taste, "respectfulness" of addressing

Table 2 . Classification of emotions for versions 1 and 2 (In the table these results are shown with "/")

Emotion	Correct Classification	Wrong Classification	Total	Correct Classification %
Sadness	3/3	1/1	4/4	75/75
Disgust/ displeasure	5/5	1/1	6/6	83.33/83.33
Anger/ threat	5/7	3/1	8/8	62.5/87.5
Fear	4/4	0/0	4/4	100/100
Interest/ alarm	2/2	1/1	3/3	66.7/66.7
Happiness	6/6	0/0	6/6	100/100
Tenderness	3/3	1/1	4/4	75/75
Total	28/30	7/5	35/35	80/85.7

to an interlocutor, expressiveness of a picture or an architectural ensemble, advertisement efficiency, etc.

The experience gained at the first stage of developing the instrument measuring expected emotions of certain categories of listeners, enabled finding peculiarities that are characteristic for metrology investigations within the humanitarian field.

First, it is the necessity to organize an interdisciplinary team consisting of metrologists and engineers, mathematicians and programmers who have experience in using new software methods as well as specialists in the humanitarian field, e.g., biologists and psychologists.

Secondly, a model assigned for measuring a specific multidimensional quantity should be formed. The number of parameters that influences the quantity should be minimized, and the definition of the quantity should be refined in many cases. To improve the study efficiency, it is possible to model an input impact by its modification, e.g., by extension of the pauses between separate soundings.

Besides, the necessity to divide gradations of the quantity under investigation along a nominal scale reasonably, should be mentioned. The automatic classification of them is desirable.

In general, these peculiarities provide a qualitative difference between the developments in the traditional fields and the field considered, but they open a new way to more profound understanding of human and society.

References

[1] **P. N. Juslin, P. Laukka.** Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin,* No. 5, vol. 129, 2003, pp. 770–814.

[2] **I. A. Volodin1, E. V. Volodina, O. A. Filatova.** Strukturnye osobpennosti vstrechaemost' i funkcional'noe znachenie nelinejnyh fenomenov v zvukah nazemnyh zhivotnyh. *Zhurnal obshchej biologii*, tom 66, № 4, 2005, s. 346-362.

[3] **E. F. Briefer.** Vocal expression of emotions in mammals: mechanisms of production and evidence, *Jornal of Zoololgy,* 2012, pp. 1-20.

[4] **G. A. Bryant.** Animal signals and emotion in music: coordinating affect across groups, *Front Psychol.* No. 4, 2013, 990.

[5] **T. Eerola, J. K. Vuoskoski,** A review of music and emotion studies: approaches, emotion modeal, and stimuli, *Music Perception* No. 3, vol. 30, 2013, pp. 307–340.

[6] **K. Sapozhnikova, R. Taymanov.** Improvement of traceability of widely-defined measurements in the field of humanities. *Measurement Science Review,* No. 3, vol 10, 2010, pp. 78-88.

[7] **K. V. Sapozhnikova, R. E. Taymanov.** Ob obespechenii metrologicheskoj proslezhivaemosti pri «izmereniyah nevozmozhnogo», *Sbornik dokladi XX1 Nacioanlen Nauchen Simpozium s mezhdunarodno uchastie "Metrologiya i metrologichno osiguryavane",* 10-14 Septemvru 2011, Sozopol, Bolgariya, c. 13-18.

[8] **K. Sapozhnikova, R. Taymanov.** Role of measuring model in biological and musical acoustics, *Proceedings of the 10th International Symposium on Measurement Technology and Intelligent Instruments (ISMTII-2011),* Daejeon, Korea, 29 June- 2 July, 2011, D5-5, 6 p.

[9] **R. Taymanov R, Iu. Baksheeva, K. Sapozhnikova, A. Chunovkina.** Measurement model as a means for studying the process of emotion origination J. Phys.: Conf. Ser., 2016, 772, Art. 012036.

[10] **K. Sapozhnikova, R. Taymanov, Iu. Baksheeva, S. Kostromina, D. Gnedykh, I. Danilova.** Metrological approach to measurements of emotions being expected in response to acoustic impacts. *18th Int. Congr. of Metrology,* Paris, France, Sept. 19-21, 2017, *EDP sciences, Web of Conferences* 2017, C. Corleto (ed.), 1006.

[11] **K. Sapozhnikova, S. Hussein, Iu. Baksheeva, R. Taymanov.** Application of neural networks in the instrument for measuring expected listener's emotions. *Proceedings of the XXII IMEKO World Congress,* Belfast, UK, 03-06 Sept., 2018 (in print).

[12] http://www.bioacoustica.org/gallery/gallery_rus.html/

[13] http://www.audiomicro.com/free-sound-effects/free-animal-sound-effects/

[14] **R. E. Taymanov, S. N. Kostromina, K. V. Sapozhnikova.** Formirovanie shkaly ehmocij, vyzyvaemyh muzykoj, XVIII *Mezhdunarodnaya nauchno-tekhnicheskaya konferenciya "Nejroinformatika-2016",* Moskva, 25 – 29 aprelya 2016 g., s.232-242

Information about the Authors:

Kseniia Sapozhnikova: deputy head of laboratory (D.I.Mendeleyev Institute for Metrology (VNIIM), St.Petersburg, Russia), member of the International Committee on Measuring Instruments (ICMI), author of about 250 scientific publications and patents. Scientific interests: measurements under harsh environment, metrological self-check of sensors and multichannel measuring systems, measurements in the humanitarian field.

E-mail: k.v.s@vniim.ru

Sherif Hussein: associate professor (Computer Engineering and Systems Department, Mansoura University, Egypt), computer technical consultant for the Egyptian Ministry of Higher Education; author of about 60 publications. S. Hussein is currently a postdoctoral researcher at the VNIIM. He is member of the ICMI, IEEE, International Federation for Artificial Organs IFAO, and Institute of Engineering and Technology IET. Scientific interests: cloud computing, IoT, image and signal processing, modeling, and artificial intelligence.

E-mail: Sherif_hussein@mans.edu.eg

Roald Taymanov: head of laboratory (VNIIM), full member of the Metrological Academy, member of the ICMI, author of more than 360 scientific publications and patents. Scientific leader of many projects intended for science-based industry. Scientific interests: measurements of non-electric quantities, metrological self-check of sensors and multichannel measuring systems, measurements in the humanitarian field.

E-mail: taymanov@vniim.ru

Iuliia Baksheeva: associate professor (Saint Petersburg State University of Aerospace Instrumentation; Institute of Radio Engineering, Electronics and Communications; Department of Radio-Engineering Systems), author of about 50 scientific publications and patens. Scientific interests: system analysis and computer simulation, ultrasonic diagnostics, metrological self-check of sensors and multichannel measuring systems, measurements in the humanitarian field.

E-mail: baksheyeva@rambler.ru.