Music and growl of a lion: anything in common? Measurement model optimized with the help of artificial intelligence will answer

K Sapozhnikova¹, S Hussein¹,², R Taymanov¹, Iu Baksheyeva³

¹ D.I. Mendeleyev Institute for Metrology, 19 Moskovsky pr., St.Petersburg, 190005, Russia
² Computer Engineering and Systems Department, Mansoura University, 60 Elgomhoria Street, City, 35516, Egypt
³ St. Petersburg State University of Aerospace Instrumentation, 67 Bolshaya Morskaya str., St. Petersburg, 190000, Russia

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

Abstract. Adaptive possibilities of population, to a significant extent, are conditioned by the capability to transform and receive information. Music has overcome information borders of bioacoustic signals but it represents their organic development. At present, automatic interpretation of information hidden in emotionally coloured bioacoustic signals of non-human animals not to mention music, has not been realized. The paper deals with topical tasks, the solution of which relies upon the opportunity to decode the flows of nonverbal information of such a kind. The prospects of developments in this field based on new measurement methods using artificial intelligence, are shown.

1. Introduction

N. Wiener and S. Lem showed that the development of technical systems and the evolution of biological systems lead to the appearance of many similar structural and functional solutions. However, these two types of systems differ significantly in system complexity and durability of the constituent processes that take place. Therefore, the study of the emergence and improvement of the methods and means of obtaining, transmitting and processing information, which were used in the course of biological evolution, may be of interest to professionals who deal with measurement technology.

This approach to solve the problem of increasing the reliability of measurement results has proved its efficiency at VNIIM in developing methods for organizing metrological self-check of various measuring instruments and measurement systems [1].

Application of such an approach to emotions, in particular, transmitted by acoustic signals, opens up new perspectives in measuring technique.

2. The evolution of communication between living creatures

In the course of evolution, nonverbal acoustic signals have originated long before the appearance of humans. These signals accompanied the estimation reactions of living creatures to the situations, which
were important for population vital activities, and influenced the behaviour of other members of the population. As evolution was progressing, the number of different signals increased. Life has originated in the ocean. When marine animals needed to transmit information about the occurrence of danger, an unexpected situation with unknown consequences, or the necessity to be ready for active actions, infrasound signals proved to be efficient. The need for estimation reactions to changing situations and transmission information that is important for the population to its other members has been preserved at the next stage of the evolution when living creatures left water to occupy dry land.

Terrestrial animals preserved the main neurophysiological reactions that determine the emotions of marine animals. Later, these reactions characterized by the infrasound frequency range developed. If one considers mammals, this range corresponds to delta- and theta-brain biorhythm frequencies.

However, the density of air is many times lower than that of water. This feature has not allowed animals to apply infrasound frequencies for transmitting information by oscillations of the medium around animals on land.

To transmit signals that were vital for population development, in the course of evolution, to create a corresponding mechanism has become necessary. Amphibians, reptiles, and mammals (much later) gained a respiratory organ with the function of reproducing sound signals, modulation of these signals by neurophysiologic reactions within the infrasound range being realized.

To receive these acoustic signals (hereinafter, they are called emotiogenic ones), it was necessary to have a special organ of hearing and realize a structure that demodulates the received signals by the part of a nervous system connected with this organ. These new features enabled animals to select infrasound frequency oscillations within the biorhythm range and stimulate the neurophysiologic reactions that are adequate to them.

Hereinafter, these neuro-physiological reactions are considered as emotions, regardless of whether they are formed, e.g., in crabs (as proto-emotions) or humans.

The beginnings of social relations inside animals’ population and the formation of their packs, caused the need to enrich transmitted information to coordinate their actions. This need resulted in the evolution of the sound apparatus aimed at improving the controllability of sound reproduction.

The maximum distance that enables an animal to hear the signal, depends mainly on its loudness. If an acoustic power is constant, the loudness grows with the increase of sound frequency. To transmit information to other animals being at different distances, the animals’ ability to change the frequency and power of audio signals became necessary.

The use of sound oscillations of a relatively high frequency as a carrier frequency contributes to the growth of the sound loudness to the level close to its limit, which facilitates the transmission of a bioacoustic signal of danger to a greater number of the population members. It is typical that the high frequency sounds are typical for small animals.

If a predator transmits bioacoustic signals using a relatively low-frequency carrier but the carrier is modulated by the signal with the same biorhythm frequency, this can be perceived as a threat (“anger”). As the simultaneous appearance of other predators is undesirable, the loudness of such a signal is wittingly limited.

At the lowest level of controllability of the sound apparatus, there were poliharmonic bioacoustic signals reflecting only the simplest basic emotions. At the next level, there appeared a possibility of initiating or slowing down such signals as well as using their combinations (the simplest emotional images). The highest level is characterized by the ability to control the acoustic signal parameters, which opened the way to onomatopoeia (sound imitation), and later to intonated speech.

The appearance of a human ability to transmit and receive verbal information led to the increase in the meaningfulness of communication, including the use of logical reasoning, but did not suppress the ability to transmit and perceive emotional information. Considering the special value of imaginative perception and associative thinking for adaptation to changing conditions, the possibilities of receiving and transmitting emotional information have been enriched by elements of musical and visual arts.
As for music coming to life, its development can be traced since the time of primitive man. The enrichment of verbal communication with variations of pitch and rhythm, using harmony and group singing, contributed to the success of human evolution [2].

Following a similar trend, at comparatively recent stages of human society development, a relative increase in the number of connections between members of certain groups of society and, accordingly, growth in the diversity of information that they exchange, also have resulted in the progress of these groups [3].

3. Opportunities for measuring emotions

The simplest emotions and simplest emotional images in total are encompassed in scientific literature by the term “basic emotions”. Basic emotions in animals can be quite clearly identified by the situations that contribute to their occurrence. At present, in the literature, there is no generally accepted conception of basic emotions, their names as well as the number, some authors often including feelings (comparatively long-term reactions) in the list of emotions (short-term reactions).

To study the relationship of the spectrum of bioacoustic signals that transmit emotions, and the neuro-physiological reactions giving rise to emotions when these signals are perceived, a concept of an appropriate mechanism was developed. Then, based on it, a model was formed for measuring the expected emotions arising from certain acoustic impacts [4], which was supplemented and developed later [5, 6]. The first step of the measurement model consisting of 3 steps, is shown in figure 1.

![Figure 1. The first step of the measurement model [6].](image)

To test the effectiveness of the measurement model, records of bioacoustic signals of animals are expedient to apply. The advantage of such a choice is that the emotions reflected in the signals of animals, in contrast to the emotions in the sound of a human’s voice, are not under non-human animals’ voluntary control, i.e., they are “honest” ones [7, 8]. In zoologists’ investigations, records of emotional bioacoustic signals are often accompanied by characterization of the situation that caused them.

In [9], there are given the results of the first experiments with recognizing emotions in the records of bioacoustic signals of 14 animals, which we based on the measurement model [6]. Artificial intelligence was used as a recognition means. Such emotions as “fear” and “happiness” were identified with 100 % confidence, while for other emotions, the level of confidence lied within (67-87) %.

These data have confirmed the link of certain modulating infrasound oscillations with emotions being transmitted by bioacoustic signals, but raised the issues of both a basic emotion definition and reason for a noticeable share of unspecified emotions.

The analysis enabled the assumption that mainly, the revealed difficulties were due to variations of basic emotions during measurement time.

To test this assumption, bioacoustic signals of mammals (growl, hiss, yelp, purr, etc.) used in [9] after nonlinear conversion and selection of infrasound oscillations were processed with the help of continuous wavelet transform (CWT). An example of visualization of the processing results is given in figure 2. The analysis has shown that for the considered cases, the first revealed components with the frequencies in the zones of delta-, theta- and alpha- brain biorhythms, distinguished by the maximum values, as a rule, have the duration up to 0.3 s. Sometimes, it is only one maximum, and sometimes, in some cases, are two or three (in different zones).
When CWT was applied to relatively long records of acoustic signals (longer than 1.5 s) after the first maximum, previous maximums could disappear, and new ones could arise. It shows that emotional colour of mammals’ bioacoustic signals can change noticeably even during a few seconds. It seems that this fact can explain the main reason of vagueness in terms of emotion recognition according to the procedure applied previously, which relied upon obtaining integral evaluations for a few second intervals [9]. Additional CWT calculations of signals after nonlinear conversion, using time delays 0.1 s, 0.2 s, and 0.3 s have confirmed that in all the cases, the basic emotion is identified most clearly for the time delay 0.2 s with the uncertainty 0.02 s.

The concept of emotion origination mechanism enabled considering the task of identification of basic and more complicated emotions using quantitative estimates. Let us suppose that for living creatures that breathe air, the signals corresponding to basic emotions include only signals modulated by oscillations with frequencies within delta- and theta- rhythm zones and their combinations. Let us also assume that if the level of modulating oscillations in the zone of any of two biorhythm types is significantly greater than both the level of noise and level of the signal in the zone of the second biorhythm type, the value of the first level is assigned “1” and that of the second one is assigned “0”. If these levels are close, depending on the signal-to-noise ratio, it is possible to attribute them “11” or “00”. Accordingly, provided the carrier frequency of bioacoustic signal $f > 500$ Hz, “1” can be assigned to it, while provided $f \leq 500$ Hz, “0” can be assigned. Then, for each emotogenic bioacoustic signal, decoding of its meaning can be proposed (see table 1). The total number of basic emotions in table 1 is limited to 7. (It is natural that marine animals breathing by means of gills are characterized by no more than 3 emotions). The table has been formed taking into consideration the correlation between activated biorhythms and the animal’s state.

Basic emotions causing the signals included in table 1 do not have generally accepted names. For higher non-human animals, including marine ones, and especially for humans, emotions corresponding to the signals from the list given in table 1, can have various nuances taking into consideration a greater variety of levels of signals with the frequencies of delta- and theta- rhythms. The nuances also depend on the levels of oscillations with the alfa- and beta-rhythm frequencies. The former characterizes visual sensibility and muscular activity, while the latter reflects brain stress related to a task being solved.

The meaning of signals based on the above-mentioned factors can correspond to different emotions. For example, signal “100” characterizes “anger”, “threat”, or “disgust”, while signal “010” reflects emotion “interest” or “surprise”, signal “110” concerns with various manifestations of tenderness while signal “111” testifies sadness or happiness.

Subsequent analysis of bioacoustic signals has shown that in records, to detect structured fragments with the durability less than 1 s is possible. In many cases, they are repeatable or have minimal variations. The sequence of such fragments seems to cause a complicated listener’s emotion. Taking
into consideration the previous experiments [9]. 81 typical fragments were chosen from the records of animals’ voices. Further, they were processed according to the measurement model [6]. Binary codes of signals (ratios) were experimentally checked with the help of these records. Not all 7 types of signals (see table 1) were found in them. Nonetheless, the concept of basic emotions taking into consideration the decoding given becomes more certain being characterized by numbers. Later it can be refined.

| Delta | Theta | Frequency | Signal meaning                                                                 |
|-------|-------|-----------|-------------------------------------------------------------------------------|
| 1     | 0     | 0         | Discontent, e.g., at meal or meeting with a specific animal                    |
| 0     | 1     | 0         | Warning! Here we have something important (to a partner that is nearby)       |
| 1     | 1     | 0         | Pleasure of comfort (to a partner that is nearby)                             |
| 0     | 0     | 1         | No grounds for anxiety (to a community).                                     |
| 1     | 0     | 1         | Threat for a community has originated. Call for help.                         |

4. Prospects for the development of instruments aimed at measuring (recognizing) emotions

Not only does cognitive interest cause the studies of the evolution of emotions and methods providing their transmission. If we limit ourselves to acoustic signal features, the following fields of research and development can be mentioned:

a) investigation of animals’ language (In particular, it is necessary to decode their communication, which would enable monitoring of ecological situation in certain places);

b) automatic interpretation of speech from any language to another one considering emotional meaning of intonation used while pronouncing words and sentences (Such a demand was repeatedly mentioned by philologists. One of the purposes is to facilitate studies of foreign languages);

c) development of the scientific foundations of the theory of music as emotion language based on numerical assessments (The prospects of this work are associated with the development of methods for enhancing the emotional impact of various musical forms typical for various national and social cultures);

d) development of the theory, methods, and means focused on the designing an optimal sound scale (Basing on this work it is planned to improve musical instruments, primarily, synthesizers and other keyboard instruments);

e) development of methods for designing technical instruments that will enable helping a composer and performer in advance to take into consideration the influence of sociocultural features of supposed listeners on their probable reaction (This field was formulated by professional musicians as one of the actual);

f) diagnostics of characteristics of the brain and psyche based on voice sounding features [10] (Diagnostics of infants’ brain development in the first months of life is a subject of particular interest to the authors of the present paper).

In the listed fields, the authors of the present paper are carrying out certain investigations together with specialists from other organizations. To provide progress in most of these fields, automatic recognition of emotion information contained in emotionally coloured soundings, is necessary.

The aforementioned model used for measuring expected emotions caused by acoustic signals [6], can be the basis for recognition of such information.

At the first stage of the development of such an automatic device, recognizing basic emotions with a rather high degree of confidence should be provided. To achieve this result, it will be necessary to define the limits of assumed “1” and “0” in each bit, receive more information on the influence of spectrum components within alfa- and beta- biorhythm frequencies as well as refine the border being used for dividing carrier frequency into two ranges.
Success at the first stage will provide a possibility to proceed to the study of human complex emotions and feelings as well as to that of the connection of emotional perception with national culture and culture of certain social strata of modern society.

For such studies, the sociological surveys will be required being carried out within the framework of restrictions [11, 12]. However, the research should be supplemented by an analysis of the correlation of survey results with records of sequences of expected basic emotions, obtained based on the aforementioned measurement model. This is a qualitatively new approach.

The solution to the applied tasks that concern measuring emotions, e.g., speech interpretation, usually requires automatic recognizing emotiogenic signals within tenths of a second. In some cases, simultaneous automatic analysis of the sequence of these signals, is necessary to allow recognizing a feeling being formed. It should be added that for a number of applications, while identifying emotions, emotional nuances should be considered, but they are determined by ratios that at present, are insufficiently known. Therefore, one of the most important tasks is recognizing emotions with the help of artificial intelligence [6, 9].

However, a continuation of studies of artificial intelligence possibilities [9] carried out in this field, have shown that it is critical to supplement emotion definitions. Using Radial Basis Function Neural Networks (RBFNNs), accompanied by k-means clustering method, proved insufficiently effective. From 35 fragments of bioacoustic signals chosen by an expert (each of the fragments represented one of 7 basic emotions: anger, disgust, fear, happiness, interest, sadness, and tenderness) only 29 were identified by RBFNN. After excluding wrong signals using k-means clustering, 28 fragments were identified but with 100 % confidence. Thus, for original bioacoustic signals, improving the method has resulted in the increase in the identification confidence to 80 % only, i.e., 20 % of signals were taken away. It is specific that only “fear” was identified in all the cases, while recognition of “interest” was the worst. To investigate complicated emotions and processes of feeling formation, to take away a part of signals is undesirable.

In our view, a future prospect could be related to clustering associated with determination of codes for any cluster and evaluation of the limits of assumed “1” and “0” in each bit for signal levels. Corresponding clustering seems to be a multistep process taking into consideration the necessity to preserve the binary codes and duration of bioacoustic signal fragments no longer that 1 c.

Certainly, for some clusters there will be no analogy in the available set of bioacoustic signals. However, if the number of binary bits influencing the emotional colour is known and permissible limits of variation of infrasound oscillation levels are established, an opportunity will open to form a full set of emotions based on binary codes.

5. Conclusions
The level reached by population in the course of evolution can be characterized by the level of methods and means for transmitting and receiving information inside the population, which enable it to adapt to the changing environment. Successful spreading of mammals on the Earth, to a significant extent, is due to the development of their vocal organs, qualitative leap in the development of primitive human being linked with origination of singing and musical instruments. During many centuries, the process of human society civilization was accompanied by the improvement of music theory and musical instruments, development of methods for verbal communication (speech and written language) as well as information transmission instruments.

At present, the devices that transmit and receive information are practically available to anyone. Automatic speech interpretation and written text translation from any language and to any other language is also possible.

However, nowadays, the requirement has originated for automatic interpretation of the emotional content of non-verbal information hidden in acoustic signals including music. Opportunities for such interpretations are provided by the modern level of measuring instruments and computer technologies.
References

[1] Taymanov R and Sapozhnikova K 2018 Smart Sensors and MEMS for Industrial Applications (2ed) ed S Nihtianov and A L Estepa (Woodhead Publishing, Elsevier Limited) chapter 1 pp 1-22
[2] Mithen S 2009 The Neurosciences and Music III — Disorders and Plasticity: Ann. N.Y. Acad. Sci. 1169 pp 3–12
[3] Taymanov R and Sapozhnikova K 2016 J. Phys.: Conf. Ser. 772 012040 6 p
[4] Taymanov R and Sapozhnikova K 2010 Meas. Sc. Rev. 10 3 pp 78-88
[5] Sapozhnikova K, Taymanov R, Baksheeva Iu, Kostromina S, Gnedykh D and Danilova I 2017 18th Int. Congr. of Metrology, EDP sciences, Web of Conf. (Paris, France, September 19th- 21) ed C Corleto 10006
[6] Sapozhnikova K, Hussein S, Baksheeva Iu and Taymanov R 2018 J. Phys.: Conf. Ser. 1065 072041
[7] Juslin P N and Laukka P 2003 Psychol. Bull. 129 5 pp 770–814
[8] Briefer E F 2012 J. Zool. pp 1-20
[9] Sapozhnikova K, Hussein S, Taymanov R and Baksheeva Iu 2018 Proc. 28th Int. Sc. Symp. Metrology and Metrology Assurance 2018 (September 10-14 2018, Sozopol, Bulgaria) pp 297-302
[10] Baksheeva Iu, Sapozhnikova K, Taymanov R and Danilova I Proc. 27th Int. Sc. Symp. Metrology and Metrology Assurance 2017 (Sept. 8-12, 2017, Sozopol, Bulgaria) pp 292-6
[11] Mari L and Wilson M 2014 Measurement 51 pp 315-27
[12] Pendrill L and Fisher W P Jr 2015 Measurement 71 pp 46-55