Application of neural networks in the instrument for measuring expected listener’s emotions

To cite this article: K Sapozhnikova et al 2018 J. Phys.: Conf. Ser. 1065 072041

View the article online for updates and enhancements.
Application of neural networks in the instrument for measuring expected listener’s emotions

K Sapo zhnikova1, S Hussein1, 2, Iu Baksheeva3 and R Taymanov1
1 D.I. Mendeleev Institute for Metrology, St. Petersburg, Russia
2 Computer Engineering and Systems Department, Mansoura University, Egypt
3 St.Petersburg State University of Aerospace Instrumentation, St. Petersburg, Russia

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

Abstract. In the society there is a need for measuring expected emotional reactions of certain listener’s category to a musical composition or excerpt performance, sounding of a musical instrument, speech intonation, and even advertisement. In the course of developing such an instrument that relies on the model of forming basic and complicated emotions, a method has been required that will enable recognition of emotions automatically. A Radial Basis Function Neural Network based on k-fold cross validation methodology can perform this function. Calculation-experimental investigations confirmed its efficiency and allowed clarifying the proposed transformation function of signals in the listener’s brain.

1. Introduction

Music arouses various emotional reactions of listeners. To have an instrument measuring expected emotions being caused by acoustic impacts, would be useful for composers and performers of musical compositions, organizers of various musical shows, advertisement designers, and even medical doctors. Such an instrument will enable quantitative evaluation of music perception by statistically similar group of listeners. Music has contributed to human survival in the course of evolution. Therefore, the link between music, on one hand, and emotions and feelings, on the other hand, which it causes, should rely on interrelations of a number of factors associated with both music and human. The authors of [1] gave more than 20 qualitative features characterizing music for each chosen emotion, at the same time, some of the features overlap and have noticeable uncertainty. To assess them, a rather long-time interval is required. Researchers have also tried to acquire the data linking perceived emotions and the music using subjective estimations: reports of listeners after listening to music (e.g., [2] and many others). Some researchers have suggested that verbal self-reports of musical emotions are unreliable [3]. This point of view can be explained by the fact that the results of psychologic tests express integral feelings that listeners feel after listening a music composition as a whole. The difference between emotions and feelings is that the emotions originated earlier within the evolution process and they arise much faster under external influences. Three-steps measurement model assigned for measuring expected emotions caused by music and other acoustic signals is justified in [4, 5]. Moreover, it can be considered as the model of the “mechanism” of emotion formation, since it links soundings heard with neurophysiologic emotiogenic reactions that trigger the processes of appearing the signs characterizing emotions being arisen. The first step of the model at the output of which basic emotions [6, 7] are formed, is shown in figure 1. Basic emotions are the most ancient ones that are the same for both animals and humans. Subsequent steps of the model show formation of complicated emotions based on sets of basic emotions.
Complicated emotions are associated with the cognitive human activity, upbringing, education, national history, as well as belonging to a certain socio-cultural group [4, 5]. The main model unit is the non-linear converter. Preliminary calculations [8, 9] have shown that after non-linear conversion of sounds of major tonic triads, unlike the minor ones, intermodulation products of a comparatively high-level within the infrasound frequency range are formed. The necessity to use non-linear converter in the model has been experimentally proved by neurophysiologic investigations [4] based on presenting tonic triads and modulated sounds. The results of studies related to changing the perception of acoustic signals with age, evaluation of expressiveness of music performance [5, 10, 11] and some others became an additional evidence. Assumptions on the non-linear conversion of acoustic signals in recent years have been also pointed out by neurophysiologists [12]. The results obtained give the basis to turn to developing the instrument for measuring expected emotions, which was mentioned above. However, the number of complicated emotions and time of their generation are significantly greater than those of the basic ones, which impedes the analysis.

The aim of the present investigation is the development of a method that will enable automatic recognition (classification) of basic emotions and, in the long run, complicated emotions and feelings.

2. Data processing for evaluation of basic emotion features
A composer and art-therapist developed a specific set of short musical excerpts (for 7-15 s) that predominantly expressed basic emotions such as interest, happiness, sadness, anger, fear, disgust, and surprise. These excerpts were presented to experiment participants who reported about their emotions. After data processing, those excerpts for which the emotions corresponding to the professional expert’s ideas were identical to the emotions of the experiment participants, were chosen from the original set. 14 excerpts characterized by the emotions of fear, happiness, sadness, and anger, gave the most convincing results. The proposed functional diagram of data processing shown in figure 2, starts with a stage of adding together original music signal with its delayed version by 0.2 s. That stage is followed by a non-linear transformation stage. It was shown in [5, 9-11] that if input signals are harmonic ones, in the course of forming emotiogenic reactions, intermodulation products of the 4th order obtained after this transformation play a significant role. Actually, as a rule, harmonics in the input acoustic signal usually exist, which can decrease the non-linearity power required. To evaluate how important this effect is and whether one-sided pressure on an eardrum is significant, three version of the non-linear transformation were applied:

\[
y(x) = x^3 \quad (1) \\
y(x) = x^4 \quad (2) \\
y(x) = (x + |x_{\min}|) \quad (3)
\]

where \( y \) is the signal value after non-linear transformation, \( x \) is the original signal value, \( |x_{\min}| \) is the module of the minimal value of the original signal.

The third stage is a STFT that transforms the signal into the frequency domain to get relative energies (8 features) for predefined frequency 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). Besides,
frequency zone of the maximum energy of the original signal is calculated according to 3 bands: less than 80 Hz, from 80 to 1000 Hz, higher than 1000 Hz, forming one extra feature. All in all, those 9 features form a feature vector [13].

The transformations applied to a music signal to extract the feature vector are illustrated by figure 3.

3. Classification of basic emotions
The fourth stage is based on the comparison of the feature vector for the music excerpts and predetermined feature vectors for the basic emotions. Automation of just this comparison and, in the long run, also comparison with feature vectors associated with complicated emotions, is necessary for the software of the measuring instrument under development.

A Radial Basis Function Neural Network (RBFNN) with k-fold cross validation is proposed to apply for the classification. It is a special type of Multilayer Perceptron (MLP) neural networks that is used in non-linear classifications. MLP is composed of many neurons and each of them takes the value of weighted sum of its input values. A single MLP neuron is considered a simple linear classifier, but non-linear classifiers can be built by combining these neurons into a network. The RBFNN approach is considered to be more versatile than the MLP. The RBFNN performs classification by determining the similarity between the inputs and examples from the training set. Each node in RBFNN can be tuned without destroying the knowledge gained by the remaining nodes. For a new input vector, each neuron calculates the Euclidean distance between input features and features characterized the examples. If the input more closely resembles class A examples than class B examples, it is classified as class A. In this research, four RBFNNs [14] have been designed to detect the emotions. Their input vector consists of 9 features calculated before. Each neural network is trained and tested using 10-fold cross-validation [15] that is applied here since quantitative information about feature vector is insufficient. The procedure starts by dividing the set of training examples into ten sets. Then one set is removed from the complete set of data and the others are used to train the network for minimizing an error in equation (4).

\[ E(\omega) = \frac{1}{2} \sum_{\mu,i} (\zeta^\mu_i - O^\mu_i)^2 \]  

where \( \omega \) is the weight vector, \( O \) is the actual output, \( \zeta \) is the desired output, \( \mu \) is the example number, and \( i \) is the node number.
The network was tested by calculating the generalization error and the procedure was repeated to cover all the data subsets. Then, the neural network with the minimum generalization error was selected [14]. To recognize four basic emotions, four classifiers were designed for each transformation in such a way that only one output is assigned “1” when the basic emotion matches its label and “0” otherwise. The total correct classification in the testing stage is then calculated for each group of classifiers that corresponds to each transformation. Results of classification for non-linear transformations according to equations (1) - (3), are summarized in table 1.

| Transformation | Happiness, % | Fear, % | Sadness, % | Anger, % | Total quality of classification, % |
|---------------|--------------|--------|------------|----------|-----------------------------------|
| Transformation 1 | 91           | 94     | 92         | 88       | 91                                |
| Transformation 2 | 90           | 86     | 91         | 85       | 88                                |
| Transformation 3 | 92           | 95     | 93         | 97       | 94                                |

4. Conclusions
The values characterizing total quality of classification for transformations 1 and 2 are close. This fact confirms that the influence of input signal harmonics on expected emotions being caused by acoustic signals, is important. The improved quality of classification for transformation 3, in which the input signal has an offset excluding “a backward sag of the eardrum”, testifies that one-sided pressure on this membrane is significant. Taking into account that the relationship of the harmonics in the input signal spectrum is uncertain, it is possible to propose that the non-linear converter in the measurement model should be characterized by the polynomial consisting $x^3$, $x^4$ as well as an absolute term that simulates the one-sided pressure mentioned above.

In general, neural networks used provide high quality of automatic recognition of the basic emotions. This result gives grounds to expect that neural networks will be efficient for classification of complicated emotions too and to apply them in the software of the measuring instrument under development.

References
[1] Juslin P N and Laukka P 2004 Journal of New Music Research 33(3) 217-38
[2] Juslin P 2013 Phys Life Rev. 10 (3) 235-66
[3] Schubert E 2013 Front Psychol. https://www.frontiersin.org/articles/10.3389/fpsyg.2013.00837/full
[4] Taymanov R, Baksheeva I, Sapozhnikova K and Chunovkina A 2016 J. Phys.: Conf. Ser. 772 012036
[5] Sapozhnikova K, Taymanov R, Baksheeva Iu, Kostromina S, Gnedykh D and Danilova I 2017 18th Int. Congr. of Metrology, EDP sciences, Web of Conferences (Paris, France, Sept. 19-21, 2017) C Corleto (ed.) 1006
[6] Izard C E 1991 The Psychology of Emotions (New York: Plenum)
[7] Eerola T and Vuoskoski J K 2013 Music Perception 30 (3) 307–40
[8] Sapozhnikova K and Taymanov R, Proc. XVII IMEKO World Congress (Dubrovnik, Croatia, June 22-27, 2003) pp. 2049-53
[9] Sapozhnikova K and Taymanov R 2010 Meas. Sc. Rev. 2010 10(3) 78-88
[10] Sapozhnikova K, Taymanov R and Baksheeva I 2017 Proc. DAKAM’S Int. Art Studies Meeting Conference MUSCUL’17 / IV Int. Music and Cultural Studies Conference, (Istanbul, Turkey, June 9-10, 2017) pp. 25-32
[11] Sapozhnikova K, Taymanov R, Baksheeva I, Kostromina S, Gnedykh D and Taronov A 2017 Proc. DAKAM’S Int. Art Studies Meeting Conference MUSCUL’17 / IV Int. Music and Cultural Studies Conference, (Istanbul, Turkey, June 9-10, 2017) pp. 15-23
[12] Lerud K D, Almonte F V, Kim J C and Large E W 2014 Hearing Research 308 41-9.
[13] Diana N and Sabiq A 2016 Int. J. Comput. Commun. Syst. Eng. 5(5) 350-7
[14] Yoo S, Oh S and Pedycz W 2015 Neural Networks 69 111-25.
[15] Wong T 2015 Pattern Recognition 2839-46