

Music Stimuli Recognition in Electroencephalogram Signal

Jozsef Suto¹, Stefan Oniga^{1,2}

¹*Department of IT Systems and Networks, University of Debrecen,
Kassai St. 26, 4028 Debrecen, Hungary*

²*Department of Electronic and Computer Engineering, Technical University of Cluj-Napoca,
Dr Victor Babes St. 62A, 430083 Baia Mare, Romania
suto.jozsef@inf.unideb.hu*

Abstract—When humans are listening to music they perceive beats, rhythms and melodies. Music stimuli induce motor system activities and it has a powerful emotion trigger effect. Since music is a potential stimulus in electroencephalogram based emotion research we supposed that different kinds of songs are recognizable from electroencephalogram signal. In this study we try to recognize music-induced electroencephalogram responses with the popular Neurosky Mindwave device. This paper describes the test conditions and the efficiency of an artificial neural network in combination with different data pre-processing techniques. The final outcomes show the negative effect of frequency decomposition and that the meditation level has more significant effect on the recognition than a particular song.

Index Terms—Artificial neural networks; Brain-computer interfaces; Digital filtering; Electroencephalography.

I. INTRODUCTION

Electroencephalogram (EEG) signal is voltage fluctuation in the human brain which comes from the ionic current flow within neurons. EEG signals have been used mainly for neurological disorder detection and evaluation in clinical environment. In the early stage of EEG research the diagnosis is usually done by a clinical expert who is familiar with brain rhythms. The vast majority of EEG related papers deal with the application possibilities of EEG readers in medical applications such as epileptic seizure and mental disorder detection [1]. The advantages of modern microelectromechanical technology made the EEG sensors available for the whole research community and take the measurements outside the special clinical environment. Currently the *Neurosky Mindwave* headset is one of the most popular and cheapest EEG readers on the market [2].

Nowadays the new machine learning (ML) methods replaced the circumstantial visual inspection based EEG analysis. Some different automated solutions already have been suggested for medical and non-medical purposes. According to the EEG processing task the ML algorithm type can be different. For instance, the authors of [3] used support vector machine (SVM) to emotional state

classification. Amarasinghe *et al.* worked with artificial neural network (ANN) and tried to recognize two patterns to robot control [4]. Lin *et al.* used ANN to the classification of four emotion categories which come from music stimuli [5]. In this work we also applied ANN because the results in [6] demonstrated that a well-constructed ANN can outperform other ML methods.

Music has a well-known impact on emotional states. Already several neuroscientists investigated the connection between music and EEG signals in different contexts [7]. The authors of [8] developed an automatic music rating system from brainwaves where the output is like or dislike. In [9] the researchers also described another music recommendation program which utilizes the user's emotional states. In lots of works scientists tried to recognize the six elementary emotions (disgust, fear, sadness, anger, surprise and happiness) via brainwaves. Almost in all earlier emotion recognition articles the model training and testing were performed with the subject's self-reported emotions. However this feedback can be unreliable in many cases because emotions are not exact and change slowly [10]. Unlike other ML problems such as voice or image recognition, in EEG classification there is uncertainty between labels because it is rather difficult to decide that the emotion was correctly determined and labeled. In addition the wide overlap between emotion states causes complications in classification. Therefore in this work we tried to recognize the concrete songs from EEG signal instead of emotional states. Our results also reveal the capabilities of the Neurosky headset which can be useful in later works.

II. BRAIN-COMPUTER INTERFACE

Currently the brain-compute interface (BCI) is an outstanding research area where the goal is to create a communication link between computer and human brain. It provides a direct way to transform brainwaves into physical effects without using muscles.

A widely used EEG reader is the single channel Neurosky Mindwave Mobile headset. It consists of two dry-electrodes. One of them is positioned on the forehead in the *Fp1* position while the second one (an ear clip) ensures reference point on the *A1* position according to the 10-20 electrode

system. An illustration about the sensor and its electrodes locations can be seen on Fig. 1. The device provides non-invasive and painless measurement up to 8 hours with a fully charged AAA battery. Its sampling frequency is 512 Hz. The output data contain 12-bit raw brainwaves (between 3 Hz–100 Hz) and pre-determined attention and meditation levels from the build-in *eSense* meter (discrete values between 0–100). The communication between the computer and the sensor takes place via Bluetooth connection with 57600 baud rate. This EEG reader already has been used for control tasks where the attention level was the input of controllers. For instance, in [11] and [12] the authors controlled a toy car and a robot where their speed depended on the average or a threshold attention level. To this study we developed an own EEG signal acquisition framework in the *Processing* programming environment. The software establishes Bluetooth connection with the sensor and saves the raw data into text files. The data have been collected from 5 volunteers aged between 14 and 52 years. They have been informed about the purpose of the experiment and the operation of the brainwave recording device. Each subject sat comfortably in a chair in front of the computer in a silent room. During data acquisition, the participants kept their eyes closed and remained as motionless as possible. The music pieces were 30 seconds long with 10 seconds wide silence interval between each song without any stimulus for relaxing and avoiding emotional contamination from previous songs. The complete data set was divided into 20 % test, 20 % validation and 60 % training data. The data acquisition software, the collected data, list of songs and other materials are freely available from the <http://irh.inf.unideb.hu/user/sutoj/eeg.php> webpage after an official request.

III. METHOD

Traditionally, brainwaves can be decomposed into particular frequency bands. Each band relates to special feelings. A summary about frequency bands can be seen in Table I. The frequency bands are slightly distinct in different articles [13], [14].

TABLE I. EEG FREQUENCY BANDS.

Band	Frequency interval	Related feelings
Delta (δ)	0.5 Hz–4 Hz	Deep sleep
Theta (θ)	4 Hz–8 Hz	Calm and relaxed mood
Alfa (α)	8 Hz–13 Hz	Awake relaxation
Beta (β)	13 Hz–30 Hz	Excitement and boredom
Gamma (γ)	30 Hz–100 Hz	Visual stimuli

Frequency decomposition is a generally accepted technique in music stimuli analysis independently of the application type. However, the influence of music on the individual bands is not clear. For example, Fachner and Stegemann demonstrated that an increasing tendency can be observed in the frontal θ power during pleasant music listening [15]. The authors of [8] showed that an increase can be seen in γ and in frontal θ band power during music perception which depends on pleasant and unpleasant sounds while Sun *et al.* claimed that different music styles influence EEG bands differently because the energy magnitude and

intensity are varying due to the music type [7]. Therefore we separated the frequency bands and investigated them independently of each other.

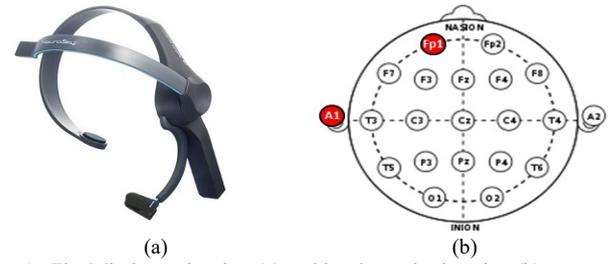


Fig. 1. The Mindwave headset (a) and its electrodes location (b).

Every change in the EEG signal which is not directly influenced by the brain potential difference is called *artifact*. The real signal which comes from the brain activity is significantly distorted by several known artifacts such as power line frequency, electrostatic interference, pulse, breathing, lid, eye and muscle movements, etc. [2]. The interfering noises can be more than 10 times stronger than the real brain activity. In addition eye movement is one of the major noise sources in recordings.

Probably the most obvious strategy for dealing with noisy data is the digital filtering. The Neurosky sensor already has an embedded band pass filter (3 Hz–100 Hz band). At first we used the whole band without frequency interval subdivision. Later the frequency bands from Table I have been separated. Since the δ band is suppressed by the device's built-in band pass filter, it was ignored. To the band separation finite impulse response (FIR) filters have been used with *Hamming* window. FIR filters do not cause phase shift and they are predominantly used in EEG signal processing [15]. The filter size has been determined by (1) where f_s is the sampling frequency, f_l and f_u are the lower and upper cut off frequencies and N is the window size (nearest odd number). According to it the filter size for θ , α , β and γ filters are 423, 339, 101 and 25 respectively. The frequency response of filters can be seen on Fig. 2. Additional information about FIR filter design can be found in [16]. In order to reduce muscle artifacts, participants had to remain in one place

$$N_{\text{odd}} \geq \frac{3.3f_s}{|f_u - f_l|}. \quad (1)$$

Most ML solutions follow a general machine learning chain. It consists of data acquisition, segmentation, feature extraction, classifier training and classification stages. In our case the signal is a single channel, discretized data flow. The data is divided into small pieces or in other words into *windows*. The window is the base of the training and classification phases because the entire information content from the window feeds the ML algorithm. In this work the window size was 3 seconds, 5 seconds and 10 seconds long with approximately 99 % overlap between adjacent windows in training phase and without overlap in the classification phase. By the wide overlap we achieved a richer training data set.

With feature extraction we try to take out the useful

characteristic of raw signal. With an appropriate feature set the classifier model will be simpler and its performance will be better [6]. In this study the features have been extracted from the time and frequency domains as in [17] where the authors collected the most relevant feature extraction methods to the human activity recognition problem. Although activity and music stimuli recognition have two different objectives, many similarities exist between them. The used features can be found in Table II.

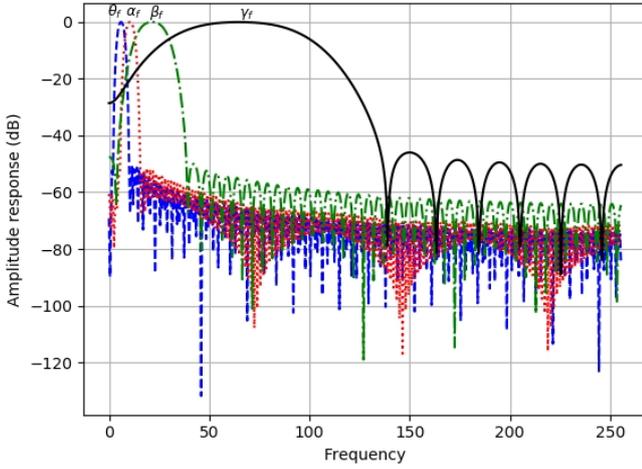


Fig. 2. Frequency responses of filters.

TABLE II. USED FEATURES.

Domain	Feature
Time	Mean
	Standard deviation
	Mean absolute deviation
	Root mean square
	Interquartile range
	75'th percentile
	Kurtosis
	Signal magnitude area
	Max-min difference
	Zero crossing rate
Frequency	Spectral energy
	Spectral entropy
	Spectral centroid
	Principal frequency

Each feature was normalized with the standard scaler (2) where F_u and F_n indicate the initial and normalized feature matrices while σ_j and μ_j are the standard deviation and mean of the j^{th} feature type. Normalization makes features equally important

$$F_n(i, j) = \frac{(F_u(i, j) - \mu_j)}{\sigma_j}. \quad (2)$$

The ANN design is based on [6] where the authors compared three ANN structures with different activation and error functions. Their result showed that an ANN with mean square error function, L2 regularization and two layers with tangent and linear activation functions can be a good choice in several ML problems. The main *hyper-parameters* of the network such as the *regularization strength* (ν), *initial learning rate* (ϕ_0) and the *learning decay tendency* (τ), were

randomly drawn from the (3) exponential scale

$$\phi_0, \tau, \nu \in 10^{U(-4, -1)}, \quad (3)$$

where U refers to the uniform distribution.

The additional network parameters were the following:

- Neurons on the hidden layer: 50;
- Training algorithm: gradient descent with 0.15 momentum and 10 samples mini batch;
- Initial bias and weights: bias values were 0 while weights came from the (4) distribution

$$W^i \in N\left(0, \frac{1}{\sqrt{\eta}}\right), \quad (4)$$

where W^i indicates the weight matrix and η denotes the number of inputs on the i^{th} layer;

- Learning decay tendency: exponential according to (5)

$$\phi = \phi_0 e^{-\tau \epsilon}, \quad (5)$$

where τ is the decay factor, ϕ_0 is the initial learning rate and ϵ refers to the epoch counter.

IV. RESULTS

At the beginning of the investigation the effect of frequency decomposition has been tested with 5 seconds wide window size. At first features from Table II have been extracted from the original 3 Hz–100 Hz signal and they were the input of the ANN. Thereafter the 3 Hz–100 Hz range has been divided into its bands with the above described FIR filters and the features from the individual bands were the new input of the network. As an example, a segment of the original signal and its frequency decomposed versions can be seen on Fig. 3. Finally, the features from the individual bands have been concatenated (θ - γ bands) and the extended feature vectors (with 56 elements) were the information source of the classifier. The results can be seen in Table III where each value is the best result after 100 trials.

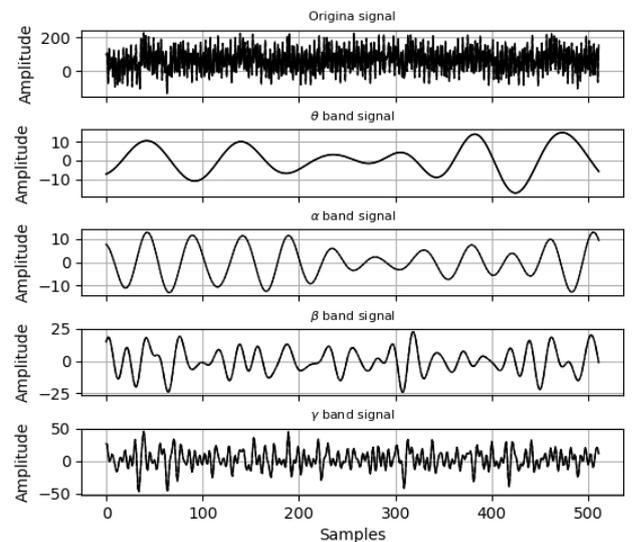


Fig. 3. The original signal and its filtered versions.

TABLE III. HIGHEST RECOGNITION RATES AFTER 100 TRIALS.

Test	1. Sub.	2. Sub.	3. Sub.	4. Sub.	5. Sub.
5 s: θ band	24.5 %	24.5 %	24.6 %	25.1 %	21.7 %
5 s: α band	26.5 %	23.2 %	23.3 %	24.9 %	21.4 %
5 s: β band	24.7 %	25.3 %	26.2 %	26.8 %	25.8 %
5 s: γ band	30.7 %	34.0 %	28.5 %	28.0 %	29.7 %
5 s: θ - γ bands	35.9 %	36.0 %	34.4 %	34.2 %	32.2 %
5 s: 3 Hz–100 Hz	35.5 %	35.2 %	31.6 %	32.8 %	33.9 %
3 s: 3 Hz–100 Hz	31.3 %	33.6 %	31.0 %	30.3 %	31.6 %
10 s: 3 Hz–100 Hz	36.9 %	34.6 %	38.0 %	36.4 %	35.1 %

TABLE IV. CONFUSION MATRIX.

Song	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1.	4	3	12	14	10	9	14	4	19	15
2.	5	17	8	8	7	18	4	7	9	20
3.	1	10	17	2	8	7	6	15	17	12
4.	0	2	9	20	4	4	9	1	11	9
5.	3	4	6	1	39	10	3	5	8	19
6.	0	2	3	2	3	49	4	9	14	11
7.	1	3	4	7	6	8	43	7	10	10
8.	1	3	4	4	2	13	5	40	12	14
9.	0	6	1	1	6	15	8	5	56	17
10.	0	2	2	3	4	4	11	1	17	68

Surprisingly the frequency band decomposition caused performance degradation because neither features from the individual bands nor joint features from all bands were more efficient than features from the entire 3 Hz–100 Hz band. Usually the narrow θ and α bands produced small recognition rates while the widest γ band contains the most useful information from the four bands. The joint features from $\theta - \gamma$ improved the accuracy but this extension was not as efficient as we expected.

In the following step the effect of window size expansion and reduction have been examined without frequency decomposition. In the first case the window size was 3 seconds while in the second case the size was 10 seconds. The results also can be seen in Table III. The narrower window decreased the recognition rate but the wider window caused significant improvement. Since the highest recognition rate was more than 35 % which is significantly higher than the 10 % chance probability, we supposed that some kind of pattern might exist in the EEG data. To find patterns, the confusion matrices to each subject also have been examined. As an example Table IV contains the confusion matrix of the first subject.

Confusion matrices show a very interesting correlation between elapsed time and recognition rate improvement. Actually we can conclude that the recognition accuracy is related to the relaxation level instead of the stimulation effect of a concrete music.

V. CONCLUSIONS

This paper described a novel music stimuli recognition approach which does not require any feedback from users. The results showed that the recognition of a concrete song from the EEG signal is very difficult. In spite of the weak accuracy, two interesting conclusions can be drawn from the measurements. Firstly, the popular frequency decomposition approach which has been used in several earlier articles such as in [7] and [10] caused accuracy loss. Secondly, the most important conclusion comes from the confusion matrices.

They show that the recognition rate is discontinuously increasing with the elapsed time. Based on this observation we suppose that recognition depends on the relaxation state rather than the real effect of music. We did not find any mention about it in previous papers. Perhaps the Neurosky Mindwave headset causes this phenomenon. In order to investigate this hypothesis the next step toward this research direction is to test multi-channel EEG readers such as the Emotive EPOC headset which measures EEG activity from 14 points on the head.

REFERENCES

- [1] U. R. Acharya, S. V. Sree, G. Swapna, R. J. Martis, J. S. Suri, "Automated EEG analysis of epilepsy: a review", *Knowl-Based Syst.*, vol. 45, pp. 147–165, 2013. DOI: 10.1016/j.knosys.2013.02.014.
- [2] R. Maskeliunas, R. Damasevicius, I. Martisius, M. Vasiljevas, "Consumer-grade EEG devices: are they usable for control tasks?", *PeerJ*, vol. 4, 2016. DOI: 10.7717/peerj.1746.
- [3] Y. P. Lin, C. H. Wang, T. P. Jung, T. L. Wu, S. K. Jeng, J. R. Duann, J. H. Chen, "EEG based emotion recognition in music listening", *IEEE T. Bio-Med Eng.*, vol. 57, pp. 1798–1806, 2010. DOI: 10.1109/TBME.2010.2048568.
- [4] K. Amarasinghe, D. Wijayasekara, M. Manic, "EEG based brain activity monitoring using artificial neural networks", in *7th Int. Conf. Human System Interactions*, Costa da Caparica, 2014, pp. 61–66. DOI: 10.1109/HSI.2014.6860449.
- [5] Y. P. Lin, C. H. Wang, T. L. Wu, S. K. Jeng, J. H. Chen, "Multilayer perceptron for EEG signal classification during listening to emotional music", in *IEEE Region 10 Conf. (TENCON 2007)*, Taipei, 2007, pp. 1–3. DOI: 10.1109/TENCON.2007.4428831.
- [6] J. Suto, S. Oniga, "Efficiency investigation of artificial neural networks in human activity recognition", *J. Ambient Intell. Human Comput.*, 2017. DOI: 10.1007/s12652-017-0513-5.
- [7] C. Sun et al., "The effect of different types of music on electroencephalogram", in *Proc. IEEE Int. Conf. Bioinformatics and Biomedicine*, Shanghai, pp. 31–37, 2013. DOI: 10.1109/BIBM.2013.6732586.
- [8] F. Kalaganis, D. A. Adamos, N. Laskaris, "A consumer BCI for automated music evaluation within a popular on-demand music streaming service "Taking listener's brainwaves to extremes", in *Artificial Intelligence Applications and Innovations. IFIP Advances in Information and Communication Technology*, vol. 475, pp. 429–440, 2016. DOI: 10.1007/978-3-319-44944-9_37.
- [9] Y. Gou, C. Wu, D. Peteiro-Barral, "An EEG based brain informatics application for enhancing music experience", *Brain Informatics. Lecture Notes in Computer Science*, vol. 7670, pp. 265–276, 2012. DOI: 10.1007/978-3-642-35139-6_25.
- [10] T. Zhang, R. Fruchter, M. Frank, "Are they paying attention? A model based method to identify individuals' mental states", *Computer*, vol. 50, pp. 40–49, 2017. DOI: 10.1109/MC.2017.84.
- [11] J. Katona, T. Ujbanyi, G. Sziladi, A. Kovari, "Speed control of Festo Robotino mobile robot using Neurosky Mindwave EEG headset based brain-computer interface", in *7th IEEE Int. Conf. Cognitive Infocommunication*, Wroclaw, pp. 251–256, 2016. DOI: 10.1109/CogInfoCom.2016.7804557.
- [12] E. Mathe, E. Spyrou, "Assessment of user experience with a commercial BCI device", in *Proc. Hellenic Conf. Electrical and Computer Engineering Students*, Petras, 2015.
- [13] N. N. Y. Chu, "Brain-computer interface technology and development: the emergence of imprecise brainwave headsets in the commercial word", *IEEE Consumer Electronics Magazin*, vol. 4, pp. 34–41, 2015. DOI: 10.1109/MCE.2015.2421551.
- [14] M. Teplan, "Fundamentals of EEG measurement", *Meas. Sci. Rev.*, vol. 2, pp. 1–11, 2002.
- [15] J. Frachner, T. Stegemann, "Electroencephalography (EEG) and music therapy: On the same wavelength?", *Music Med.*, vol. 5, pp. 217–222, 2013. DOI: 10.1177/1943862113495062.
- [16] L. Tan, J. Jiang, *Digital Signal Processing Fundamentals and Applications*. Elsevier, 2013, pp. 161–288.
- [17] J. Suto, S. Oniga, P. Pop-Sitar, "Feature analysis to human activity recognition", *Int. J. Comp. Commun.*, vol. 12, pp. 116–130, 2017. DOI: 10.15837/ijccc.2017.1.2787.