

Auditory Saliency Classification Using EEG Data

Silvia Corchs and Francesca Gasparini

Department of Informatics, Systems and Communication, University of Milano-Bicocca, Milan, Italy
Email: {silvia.corchs, gasparini}@disco.unimib.it

Abstract We investigate if starting from EEG signals it is possible to classify salient audio events. We set up an experiment and collected EEG data with proper audio stimuli. We trained several kNN classifiers. As the EEG signals under different frequency bands are related to different brain activities, we analyzed EEG data from each single electrode and for each single frequency band, as well as EEG data from combined electrodes and frequency bands. From our preliminary analysis we found that for each frequency band there is a set of electrodes that permits to achieve better classification results and these electrodes are related to specific brain regions. We propose a cross-subject approach, where different classifiers are trained on each single individual, and tested on data collected from all the others. This initial investigation suggests several hints to improve this classification task.

Keywords: Auditory saliency, EEG data, Emotiv EPOC

1 Introduction

Attention plays a very important role by separating relevant signals from irrelevant ones. Understanding brain behavior in presence of salient stimuli can be useful in the field of Brain Computer Interface (BCI) and Human Computer Interaction (HCI). Human machine interactions have become frequent in modern life. In order to make this collaboration more efficient, machines should understand when we are alert, and attentive or engaged in focused mental activity. Spatial saliency has been investigated in a very large number of behavioral, functional brain-imaging and theoretical studies [1,2], but less attention has been paid to auditory saliency [3,4,5]. However, when interacting with reality or observing a video, humans are not only driven by visually salient stimuli but also by auditory salient ones. While many computational models are available for the case of visual saliency maps [6,7,8], fewer models exist for acoustic attention [9]. Within this context, the aim of this work is to investigate if it is possible to classify salient audio events starting from EEG data. Different authors have shown that auditory attention can modulate EEG responses sufficiently to be used as a control mechanism in BCIs [10,11,12]. For example, auditory-event related potential (P300) BCI spelling system for locked-in patients is widely used [13,14]. A BCI could be constructed that determines how the locked-in patient is focusing attention, using such information to navigate a command menu or control a device. Using EEG responses to better understand the mechanism of auditory attention can also be very useful to improve hearing aids technology.

In this article we present an analysis and investigation that aims to help in the understanding of auditory attention and measurable brain signals correlates. The works in the state of the art focus on different and complementary aspects of the study here presented. For example, temporal and spatial attention has been studied using EEG, considering auditory binaural cues and unilateral ones [15]. Several works instead have been recently done on emotional state classification from EEG data both in case of visual and audio stimuli. These works have provided a reference for our work [16,17]. One of the challenges of studies of auditory salience is the open interpretation of what auditory salience refers to. Some works in the literature address the study of simple beep tones as salient stimuli [18]. Instead, in the present work we focus on more complex salient stimuli like for example animal sounds (horse, rooster), the sound of a telephone or a knock, superimposed with background stimuli like music or rain, among others. In the results here presented we used EEG data recorded by using a 16 electrodes Emotiv EPOC device. To generate these data, we have setup an experiment in our laboratory with proper stimuli composed by a background track where a salient sound is superimposed. Data acquisition and dataset generation are described in Section 2. In the preliminary analysis here reported we consider kNN as binary classifier. The features adopted are evaluated in both time and frequency domains on the EEG data. Details on

this classification approach are reported in Section 4. We have trained several kNN classifiers, different for what concerns training and test data to investigate from different point of views how brain activity is related to auditory attention. We here also propose a cross-subject approach, where different classifiers are trained on each single individual, and tested on data collected from all the others. The performance of these different classifiers are reported and compared in Section 5. Finally in the conclusions several suggestions for future works are proposed to improve the classification performance.

2 EEG Data Acquisition

In our study we have performed two experimental sessions involving a total of 20 subjects, mean age 30, std 8, 6 female, 14 male. All subjects involved were volunteers, recruited among students, and researchers of our department at the University of Milano Bicocca. All of them had normal hearing and none had a known history of neurological disorders. No participants under the age of 18 were involved in our study. Informed consent was given by all subjects and the study was conducted in conformity with the declaration of Helsinki.

Each experimental session consists of 10 trials (audio stimuli) of 14 seconds each, alternated with 4 seconds of silence, as depicted in Figure 1. Each trial corresponds to an audio stimulus where a salient sound is superimposed to an audio background. We have generated a database of 18 audio stimuli (trials), combining 9 different audio backgrounds with 12 salient sounds, listed in the first and second columns of Table 1, respectively. Finally, the 10 trials adopted in the two experimental sessions are listed in the last two columns of the same Table.

Table 1. Audio stimuli (trials) adopted in our two experimental sessions starting from a set of background audios and salient sounds

background	salient sound	Trials of experiment 1	Trials of experiment 2
S=Sea	horse	R-telephone	NAM1-horse
P=Pianoforte	rooster	P-shrill	S-rooster
R=Rain	scream	S-scream	H-gun
H=Helicopter	notification	NAM2-horse	P-shrill
IM=Instrumental Music	gun	RM2-notification	S-scream
NAM1=New Age Music 1	telephone	IM-synthetic1	R-telephone
NAM2=New Age Music 2	knock	P-synthetic2	RM2-notification
RM1=Relaxing Music 1	shrill	R-honk	IM-wobble
RM2=Relaxing Music 2	honk	NAM1-rooster	RM2-knock
	wobble	NAM2-shrill	R-gun
	synthetic1		
	syntetic2		

EEG data were recorded using an Emotiv EPOC+ 16-channel EEG wireless recording headset (Emotiv Systems Inc., San Francisco, CA, USA). The electrode scheme was arranged according to the international 10 – 20 system and included 14 active electrodes at AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4 positions, and two electrodes used as reference (CMS and DRL) (see Figure 2).

EEG data are acquired with an internal sampling frequency of 2048 Hz. These data are proper bandpass filtered to avoid aliasing using hardware filters and down-sampled to 128 Hz with a precision of 16 bit. Special attention was taken to put the headset in an appropriate manner on each subject's head. All subjects were instructed to remain as still as possible, with closed eyes in order to minimize muscular and occipital artifacts during the EEG recordings.

3 Processing of EEG Raw Data

For each subject we have recorded 14 raw mono dimensional signals corresponding to the 14 active electrodes. We have preprocessed these data using EEGLab, an open source environment for electrophysi-

Table 2. Features adopted

simple statistics	mean, standard deviation, skewness and kurtosis
δ : mean of the absolute value of the first differences	$\delta = 1/(N - 1) \sum_{i=1}^{N-1} x(i + 1) - x(i) $
δ_2 : mean of the absolute value of the second differences	$\delta_2 = 1/(N - 2) \sum_{i=1}^{N-2} x(i + 2) - x(i) $
Signal Energy (E): is defined as the area under the squared signal	$E = \sum_{i=0}^{N-1} x(i) ^2$
Sub-Band Energy (EB): is the signal energy in a given sub-band B, where M_B is the highest frequency index in that sub-band	$EB = \sum_{n=0}^{M_B} X(n) ^2$,
Frequency Centroid (FC): represents the average point of the spectral power distribution	$FC = \frac{\sum_{n=0}^M n X(n) ^2}{\sum_{n=0}^M X(n) ^2}$
Frequency Bandwidth (FB): is defined as the size of frequency interval assigned to a signal	$FB = \sqrt{\frac{\sum_{n=0}^M (n-FC)^2 X(n) ^2}{\sum_{n=0}^M X(n) ^2}}$

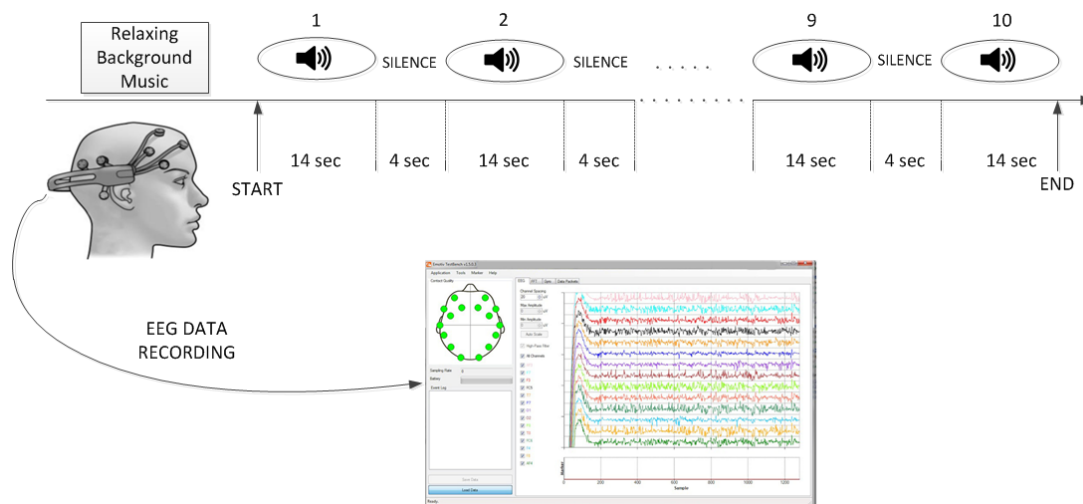


Figure 1. EEG data acquisition protocol

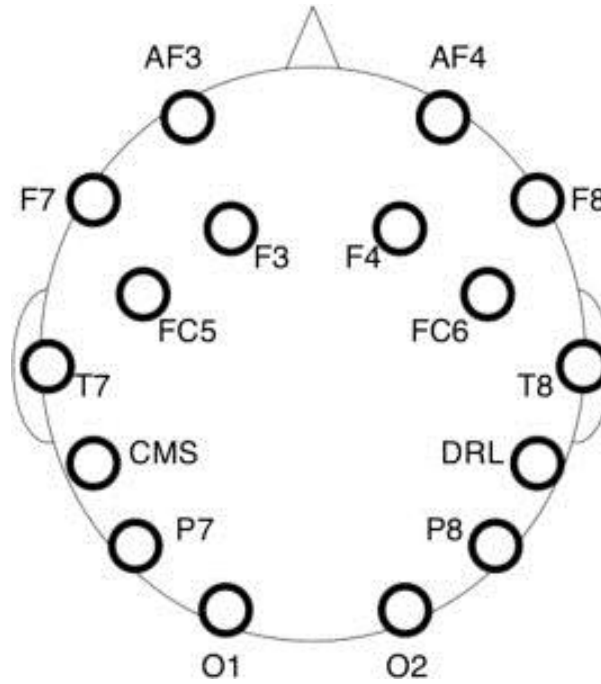


Figure 2. Emotiv Epoc electrode scheme

ological signal processing [19], to obtain signals in the following three frequency bands: Alpha 8 – 12 Hz (α), Beta 12 – 30 Hz (β), and Gamma 30 – 80 Hz (γ). Alpha waves are mainly found in the occipital region of the brain and they are related to memory brain functions and mental efforts. Increasing mental effort causes a suppression of alpha activity [20]. Beta brain waves are recorded in the frontal and central regions of the brain and they dominate when attention is directed towards cognitive tasks. They are present when we are alert, attentive, engaged in problem solving, judgment, decision making, and engaged in focused mental activity [21]. Finally, Gamma activity seems to play a significant role in the perception of both visual and auditory stimuli [22]. We have discarded Delta (below 4 Hz) and Theta (4 – 8 Hz) frequency bands as they seem to be not so relevant in healthy and awake adults [23]. Thus we denote the EEG data adopted for our classification task as: $EEG_{wave,e}(t,s)$, where $wave = \alpha, \beta, \gamma$, e refers to the 14 active electrodes listed in the previous section and depicted in Figure 2, t refers to the 18 trials (see Table 1), and s is the considered subject, with $s = 1, \dots, 20$.

4 Saliency Classification

In this work we have considered a kNN classifier, with euclidean distance and $k = 10$ neighbors to classify auditory salient events.

To generate the ground-truth data, each audio stimulus (trial) has been segmented into non overlapping frames of 0.125 seconds, obtaining (14 seconds/0.125 seconds =) 112 frames each. These frames have been manually labeled as salient (1) or no-salient (0).

Each data signal $EEG_{wave,e}(t,s)$ has also been segmented to obtain $F = 112$ frames of $\Delta t = 0.125$ seconds, corresponding to $N=16$ samples, and on each frame, features in both time and frequency domains have been evaluated. The list of the ten considered features is reported in Table 2 [24,25]. Given subject s , trial t , sensor e and brain wave $wave$, we adopt the simplified notation $x(i)$ to denote the i -th amplitude sample of each of the F frames of $EEG_{wave,e}(t,s)$. In the frequency domain, we use $X(n)$ to denote the DFT of each frame of the signal in time domain.

Thus the data available to be used in the classification task is done by: 2 experimental sessions \times 10 subjects \times 10 trials \times 112 frames \times 14 electrodes \times 3 brain waves \times 10 features.

5 Results and Discussion

We have evaluated several kNN classifiers, with different intents, varying the training data, the test data and the adopted features. In the 18 trials here considered, the number of salient and no-salient frames is strongly unbalanced. For this reason in the training data of all the considered classifiers we have randomly chosen the no-salient frames among those available to be balanced with respect to the salient ones. For each classifier we have adopted a Z-score normalization of the features so that they all have the properties of a standard normal distribution with zero mean and $\sigma = 1$.

5.1 Best Predictor among Subjects

In the first analysis we are interested in evaluating the performance of each of the 10 subjects involved in the first experimental session in predicting the 10 subjects involved in the second experimental session, with the aim of finding the one that can be considered as the best ‘predictor’. To this end we have trained 10 classifiers $CS(s)$ on the data collected for each subject ($s = 1, \dots, 10$) involved in the first experimental session. We have considered the EEG responses to all the 14 electrodes across all EEG frequency bands. For each of these classifiers, we have 252 training frames labeled as salient (126), and no-salient (126) and for each frame a feature vector of 140 elements (10 features \times 14 electrodes). We have tested each of these classifiers on the data collected from each of the subjects of the second experiment ($s = 11, \dots, 20$). The performance of these classifiers are reported in terms of accuracy, recall, precision and F-measure averaged with respect to the 10 subjects of the test sets ($s = 11, \dots, 20$). These values are shown in Table 3. From this analysis it emerges that $CS(7)$ can be chosen as the best predictor, considering all the brain waves, and all the sensors together.

Table 3. Performance of each of the $CS(s)$ classifiers built on each of the subjects involved in the first experimental session ($s = 1, \dots, 10$), and tested on the data obtained from the second experimental session ($s = 11, \dots, 20$)

	$CS(1)$	$CS(2)$	$CS(3)$	$CS(4)$	$CS(5)$	$CS(6)$	$CS(7)$	$CS(8)$	$CS(9)$	$CS(10)$
Accuracy	0.49	0.50	0.50	0.50	0.50	0.47	0.50	0.50	0.49	0.51
Recall	0.13	0.34	0.64	0.35	0.40	0.18	0.80	0.37	0.26	0.51
Precision	0.47	0.51	0.51	0.51	0.52	0.45	0.51	0.52	0.51	0.52
F-measure	0.20	0.41	0.57	0.41	0.45	0.25	0.62	0.43	0.34	0.52

Table 4. Performance of $C_{final}(7)$ classifier built on subject 7 of the first experimental session and tested on the data obtained from all the 10 subjects of the second experimental session ($s = 11, \dots, 20$), and corresponding average values

	Subj 11	Subj 12	Subj 13	Subj 14	Subj 15	Subj 16	Subj 17	Subj 18	Subj 19	Subj 20	average
Accuracy	0.59	0.52	0.53	0.57	0.55	0.53	0.53	0.63	0.54	0.54	55
Recall	0.69	0.65	0.66	0.59	0.61	0.56	0.60	0.63	0.60	0.62	62
Precision	0.58	0.53	0.53	0.58	0.55	0.52	0.53	0.62	0.53	0.53	55
F-measure	0.63	0.58	0.59	0.58	0.58	0.54	0.56	0.63	0.57	0.57	58

5.2 Optimal Electrodes with Respect to the Brain Wave

In this analysis we are interested in finding the electrodes that permit the best classification accuracy with respect to the three different frequency bands here considered. To this end we have trained 420

classifiers $C_{all}(s, wave, e)$, one for each subject ($s = 1, \dots, 10$) of the first experimental session and for each brain wave ($wave = \{\alpha, \beta, \gamma\}$), and considering all the 14 electrodes e . For each of these classifiers, we have 252 training frames labeled as salient (126), and no-salient (126) and for each frame a feature vector of 10 elements. We have tested each of these classifiers on the data collected from each of the subjects of the second experiment, with respect to the corresponding brain wave and electrode. Thus each classifier ($C_{all}(s, wave, e)$), has been tested on 10 test sets $TS(s, wave, e)$, obtained considering the $s=11, \dots, 20$ subjects involved in the second experimental session. For each brain (wave $wave=\alpha, \beta$, and γ), and sensor e , we have averaged the classification accuracies over the observers both in the training and in the test sets. From this analysis it emerges that in case of the α brain waves the three sensors that show the highest performances are P7, P8 and O1. These sensors are located in the occipital region of the brain which is actually the region where alpha waves are more localized. For what concerns β and γ waves the electrodes that better perform in our classification are AF3 (β) and AF4 together with F8 and F3. All these electrodes are located in the pre-frontal cortex which is related to control of attention [26] add to top-down mechanisms.

5.3 Best Waves and Electrodes Combination

Finally we have considered a classifier $C_{final}(7)$ where we have adopted as training data the EEG signals collected from observer 7, which was identified from the first analysis as the best predictor. For what concerns the feature vector, we have here combined the best electrodes found for each brain wave. In Table 4 the performance of this classifier with respect to data collected from all the subjects of the second experimental session is reported. Comparing this Table with Table 3 we observe an overall improvements in the performance measures.

6 Conclusions

Our preliminary investigation on EEG data reveals that a proper combination of electrodes and brain waves permits a better classification of auditory saliency, than each of them considered singularly. Even if the classification performance achieved is still not adequate, we are confident that there is room for a significant improvement. First of all a proper regularization of the classification output that takes into account temporal correlation should be applied. Moreover, other features can be considered: for instance Wavelet-Based features, adopted by Murugappan et al. in the field of emotion discrimination from EEG data [27], or fractal dimension feature in combination with statistical features [28], or cepstral coefficient features as adopted by [29]. Furthermore we plan to test several different classifiers, such as Bayes classifiers, or Support Vector Machine. Finally, as the performance obtained is significantly different with respect to subjects used for training and testing, we plan to train a classifier on data from a group of subjects to predict the mental state of the other individuals.

References

1. T. Gruber, M. M. Müller, A. Keil, and T. Elbert, "Selective visual-spatial attention alters induced gamma band responses in the human eeg," *Clinical neurophysiology*, vol. 110, no. 12, pp. 2074–2085, 1999.
2. M. M. Müller, T. Gruber, and A. Keil, "Modulation of induced gamma band activity in the human eeg by attention and visual information processing," *International Journal of Psychophysiology*, vol. 38, no. 3, pp. 283–299, 2000.
3. A. P. Souza, L. B. Felix, A. M. M. de Sá, and E. M. Mendes, "Vision-free brain-computer interface using auditory selective attention: evaluation of training effect," in *XIV Mediterranean Conference on Medical and Biological Engineering and Computing 2016*. Springer, 2016, pp. 196–199.
4. H. Tang, S. Crain, and B. W. Johnson, "Dual temporal encoding mechanisms in human auditory cortex: Evidence from meg and eeg," *NeuroImage*, vol. 128, pp. 32–43, 2016.
5. H. Higashi, T. M. Rutkowski, Y. Washizawa, A. Cichocki, and T. Tanaka, "Eeg auditory steady state responses classification for the novel bci," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*. IEEE, 2011, pp. 4576–4579.
6. L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, p. 1254–1259, 1998.

7. S. Corchs, G. Ciocca, and R. Schettini, "Video summarization using a neurodynamical model of visual attention," in *Multimedia Signal Processing, 2004 IEEE 6th Workshop on*. IEEE, 2004, pp. 71–74.
8. T. Yubing, F. A. Cheikh, F. F. E. Guraya, H. Konik, and A. Trémeau, "A spatiotemporal saliency model for video surveillance," *Cognitive Computation*, vol. 3, no. 1, pp. 241–263, 2011.
9. C. Kayser, C. I. Petkov, M. Lippert, and N. K. Logothetis, "Mechanisms for allocating auditory attention: an auditory saliency map," *Current Biology*, vol. 15, no. 21, pp. 1943–1947, 2005.
10. I. Choi, S. Rajaram, L. A. Varghese, and B. G. Shinn-Cunningham, "Quantifying attentional modulation of auditory-evoked cortical responses from single-trial electroencephalography," *Frontiers in human neuroscience*, vol. 7, pp. 115, 2013.
11. N. Hill and B. Schölkopf, "An online brain–computer interface based on shifting attention to concurrent streams of auditory stimuli," *Journal of neural engineering*, vol. 9, no. 2, pp. 026011, 2012.
12. M. Lopez-Gordo, E. Fernandez, S. Romero, F. Pelayo, and A. Prieto, "An auditory brain–computer interface evoked by natural speech," *Journal of neural engineering*, vol. 9, no. 3, pp. 036013, 2012.
13. M. Schreuder, T. Rost, and M. Tangermann, "Listen, you are writing! speeding up online spelling with a dynamic auditory bci," *Frontiers in neuroscience*, vol. 5, pp. 112, 2011.
14. A. Kübler, A. Furdea, S. Halder, E. M. Hammer, F. Nijboer, and B. Kotchoubey, "A brain–computer interface controlled auditory event-related potential (p300) spelling system for locked-in patients," *Annals of the New York Academy of Sciences*, vol. 1157, no. 1, pp. 90–100, 2009.
15. F. Faugeras and L. Naccache, "Dissociating temporal attention from spatial attention and motor response preparation: A high-density eeg study," *NeuroImage*, vol. 124, pp. 947–957, 2016.
16. X.-W. Wang, D. Nie, and B.-L. Lu, "Emotional state classification from eeg data using machine learning approach," *Neurocomputing*, vol. 129, pp. 94–106, 2014.
17. I. Mehmood, M. Sajjad, S. W. Baik, and S. Rho, "Audio-visual and eeg-based attention modeling for extraction of affective video content," in *Platform Technology and Service (PlatCon), 2015 International Conference on*. IEEE, 2015, pp. 17–18.
18. C. Pokorny, D. S. Klobassa, G. Pichler, H. Erlbeck, R. G. Real, A. Kübler, D. Lesenfans, D. Habbal, Q. Noirhomme, M. Riseti *et al.*, "The auditory p300-based single-switch brain–computer interface: paradigm transition from healthy subjects to minimally conscious patients," *Artificial intelligence in medicine*, vol. 59, no. 2, pp. 81–90, 2013.
19. A. Delorme and S. Makeig, "Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis," *Journal of neuroscience methods*, vol. 134, no. 1, pp. 9–21, 2004.
20. L. Venables and S. H. Fairclough, "The influence of performance feedback on goal-setting and mental effort regulation," *Motivation and Emotion*, vol. 33, no. 1, pp. 63–74, 2009.
21. A. R. Clarke, R. J. Barry, R. McCarthy, and M. Selikowitz, "Eeg analysis in attention-deficit/hyperactivity disorder: a comparative study of two subtypes," *Psychiatry research*, vol. 81, no. 1, pp. 19–29, 1998.
22. K.-H. Lee, L. M. Williams, M. Breakspear, and E. Gordon, "Synchronous gamma activity: a review and contribution to an integrative neuroscience model of schizophrenia," *Brain Research Reviews*, vol. 41, no. 1, pp. 57–78, 2003.
23. A. Kübler, B. Kotchoubey, J. Kaiser, J. R. Wolpaw, and N. Birbaumer, "Brain–computer communication: Unlocking the locked in." *Psychological bulletin*, vol. 127, no. 3, pp. 358, 2001.
24. L. Lu, H.-J. Zhang, and H. Jiang, "Content analysis for audio classification and segmentation," *Speech and Audio Processing, IEEE Transactions on*, vol. 10, no. 7, pp. 504–516, 2002.
25. S. Corchs, G. Ciocca, M. Fiori, and F. Gasparini, "Video salient event classification using audio features," Proc. SPIE 9027, Imaging and Multimedia Analytics in a Web and Mobile World 2014, 90270P (March 3, 2014).
26. A. F. Rossi, L. Pessoa, R. Desimone, and L. G. Ungerleider, "The prefrontal cortex and the executive control of attention," *Experimental Brain Research*, vol. 192, no. 3, pp. 489–497, 2009.
27. M. Murugappan, N. Ramachandran, and Y. Sazali, "Classification of human emotion from eeg using discrete wavelet transform," *Journal of Biomed.Sci.Eng*, vol. 3, pp. 390–396, 2010.
28. R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 10, pp. 1175–1191, 2001.
29. S. Pazhanirajan and P. Dhanalakshmi, "Eeg signal classification using linear predictive cepstral coefficient features," *International Journal of Computer Applications*, vol. 73, no. 1, 2013.