KMEANS-ICA BASED AUTOMATIC METHOD FOR EOG DENOISING IN MULTI-CHANNEL EEG RECORDINGS

Elie Bou Assi, Sandy Rihana *IEEE Member*
Biomedical Engineering Department, Faculty Of Engineering, Holy Spirit University, USEK, Lebanon
ebouassi@net.usek.edu.lb, sandyrihana@usek.edu.lb

ABSTRACT
Electroencephalogram (EEG) recordings are contaminated by different internal and external noises and interferences. Therefore, they should be manipulated in order to restore them from these artifacts that could be eye blinks, electrocardiogram (ECG) and many others. Recent research is mainly oriented toward implementing methods in order to remove ocular artifacts whose frequency band overlap with the EEG frequency of interest. Independent Component Analysis (ICA) has already shown to be an effective way for removing the activity of these artifacts. However, when implementing an ICA-based method, the key relies on how to identify the ocular artifact components. Based on the components characteristics, different features such as correlation coefficients, distribution ratio, and maximum value have been identified in order to recognize in an automatic way the artifactual components and their subtraction from the original space to get ocular artifacts free EEG signals. Artifactual components were identified using an adaptive thresholding by means of K-means clustering. Qualitative and quantitative techniques of evaluation are presented and give promising results. The classification accuracy based on the correlation feature reached 99.54%.

KEY WORDS
Neural Sensory Systems and Rehabilitation, Brain Computer Interface, Electroencephalography, Independent Component Analysis, Kmeans clustering.

1. INTRODUCTION

Electroencephalogram (EEG) is a measurement of electrical currents that flow during synaptic excitations of the dendrites of many pyramidal neurons in the cerebral cortex [1]. Therefore, EEG is used for recording electrical activity of the brain via electrodes placed across the scalp. However, this electrical signal is always contaminated bythe internal and external noises and should be processed for further analysis and feature extraction. Ocular artifacts, especially eye blinks, are one of the main interferences in EEG and consist of low frequency high amplitude signals with a frequency range maximal below 4Hz and show propagation over the most anterior head regions [2].This type of artifacts cannot be filtered by traditional FIR filters such as high pass filters since they generally overlap with the EEG of interest or rely in the same EEG frequency band. Actually, some other artifacts can be minimized by training the subject to relax and to avoid facial expression during the experimentation; however it seems impossible for almost all people to control involuntary eye blinking. There are two main approaches used in order to deal with this type of artifacts; the first consists of a manual rejection of the epochs on the EEG channels that represent an ocular artifact. This is usually done by a field expert but it is not accurate and usually accompanied by a loss of data. The second consists of conventional Electro-oculogram (EOG) correction methods such as linear filtering, regression methods, Principal Component Analysis (PCA) and Independent Component Analysis (ICA) [3]. Many researchers have shown that ICA is very effective in eliminating the activity of a wide variety of artifactual sources from EEG recordings with results comparing favorable to those obtained with other methods [2]. However, when implementing an ICA-based method, the key relies on how to identify the artifactual components. Many researchers have reported their achievements in this field. Some of them added the acquisition of an EOG additional channel and adopted the relations between the EOG and the independent components (ICs) [4]. Others identified ocular artifacts ICs in a manual way by interpreting scalp topographies as well as frequency distribution characteristics in terms of visual inspection [5].This paper discusses the implementation of an automatic ocular artifacts rejection method based on ICA without the use of an additional channel such as an EOG channel. Therefore, ocular artifact components will be classified according to different features based on the proper characteristics of these ICs. Assuming that ocular artifact components are highly correlated with the prefrontal electrodes relies on the fact that these latter electrodes are the most contributed with EOG noise [1]. Furthermore, normal EEG activity is tightly distributed about its mean and ocular artifacts are ten times larger in amplitude than normal neurological signals. Based on these assumptions, three features were adopted for classification; the first one consists of calculating the average of the correlation of each IC with prefrontal electrodes, the second consists of calculating the maximum value of each IC and the third is the ratio between the peak amplitude and the variance of each IC. Based on these features, the ICs will be classified using adaptive thresholding by Kmeans clustering. Therefore, this study aims to develop an automatic Kmeans-ICA based method in order to remove ocular artifacts from multi-channel EEG recordings.

DOI: 10.2316/J.2014.216.818-0061

2. MATERIAL AND METHODS

2.1 Signal Acquisition

The signals database consists of an international database "PhysioNet.org" which includes a large amount of datasets regarding different physiological effects. "EEG Motor Movement/Imagery" signals were adopted. This dataset was recorded using the BCI2000 instrumentation system over 109 subjects. Each subject performed 14 different trials while 64 EEG channels were recorded and sampled at 160Hz [6].

2.2 Temporal Filtering

All EEG channels were band pass filtered between 1 and 30 Hz since this broad frequency band contains all frequency components necessary for a motor imagery classification. This was done by first implementing a high pass finiteimpulse response filter (FIR) with cutoff frequency equal to 1Hz; the main goal of this filter is to remove the disturbing very low frequency components such as those related to movement or breathing. Then a low pass FIR filter with cutoff frequency of 30Hz was implemented in order to remove high frequency artifacts such as power lines interferences.

2.3 Independent Component Analysis

Independent Component Analysis aims to separate a multivariate signal into additive components or activations in a way that these components are statistically independent and they are non-Gaussian. Almost all ICA algorithms start by a preprocessing phase that consists of whitening the data to remove any existing correlation. Actually, EEG signals consist of a linear mixture of real brain activity with different noise components, therefore when we whiten these linear mixtures, the variance on both axes is now equal and the correlation of the projection of the data on all axes is equal to zero. After whitening the data, the ICA algorithm consists of rotating the resulting axis of the matrix in order to minimize the Gaussianity of the projection on all axes. Therefore the full transformation from the original space is known as the weight matrix [7].

$$S = W*X \quad (1)$$

where the matrix X is the data in the original space, W is the weight matrix and S represents the sources activity. The main objective of the ICA algorithm is to find the weight matrix W that decomposes the multi-channel EEG signal into independent components assuming temporal and spatial independency. In this case, X is a matrix containing the EEG signals recorded at different electrode sites on the brain where each row of this matrix is one time waveform EEG signal. In this work, an Extended-Infomax ICA was implemented because of its ability to separate sub and sup Gaussian signals simultaneously [7]. Some of these ICs represent ocular artifacts and should be omitted from the signal reconstructed part. However the key relies on how to identify these artifactual components.

2.4 Automatic Artifact Identification

Based on different features, artifactual components are identified, and then classified into two groups using an adaptive thresholding (as described in section E):

- Average correlation: ocular artifact components are highly correlated with prefrontal electrodes since those are the electrodesmostly affected by EOG noise.Therefore, this first feature consists of calculating the average of correlation between each independent component and two prefrontal electrodes (FP1 and FP2). The correlation values expected should be greater for the ICs containing eye blinking artifacts. Cross correlation, $R_{xy,}$is widely used in order to estimate the degree of similarity of two time series [1]. The range of the correlation data is -1 to 1; therefore, more the signals are similar, the closer the correlation value is to 1. $R_{xy}$is the cross correlation of tow time series x (n) and y (n) as cited in equation 2.

$$R_{xy} = \sum_{n=-\infty}^{+\infty} x[n] * y[n+m] \quad (2)$$

Where, m denotes the number of samples by which y[n]is delayed. Then the average correlation feature vector was created by computing the correlation of each independent component with the prefrontal electrodes (FP1 and FP2) as shown in equation 3.

$$Avrg(i) = (R_{fp1,IC(i)} + R_{fp2,IC(i)})/2 \quad (3)$$

- Distribution ratio: this feature is based on the fact that normal EEG activity is tightly distributed about its mean [1]. Therefore a second classifying feature vector was adopted and is equal to the ratio between the peak amplitude and the variance of each IC as depicted in equation 4.

$$Dist_{ratio} = \frac{\max(IC(i))}{\delta^2(\mathrm{IC(i)})} \quad (4)$$

For i= 1, 2… N. Where max(IC (i)) is the maximum value or amplitude of the i^{th} independent component and $\delta^2(\mathrm{IC(i)})$ is its variance. Normal EEG components should give low values regarding this feature relatively to artifactual components.

- Maximum value: as its name implies, this feature vector consists on simply calculating the maximum value of each independent component based on the fact that ocular artifacts are greater in amplitude than normal neurological signals.

2.5 Adaptive Thresholding: Kmeans Clustering

In order to classify the ICs into ocular or non-ocular artifacts, we have to determine a threshold for all feature vectors. This was done by applying Kmeans algorithm that starts by partitioning the set of ICs into two groups using a random threshold. Then, it calculates the centroids of each group and assigns each component to the group that has the closest centroids based on the squared Euclidian distance. Once all the components are assigned, it recalculates each centroid and iterates until both centroids remain the same. Actually, it is an iterative partitioning in a way that the considered threshold will minimize the intra class variance and thus maximize the inter class variance [8]. The mathematical principle is explained below:

Considering a set of n independent components

IC's {IC (1), IC (2), IC (3)…….IC (n)};

The algorithm starts by partitioning this set into two groups and then computes the local means of each group as shown in equation 5:

$$C = \frac{1}{n_1} \sum_{i=1}^{n1} IC_1(i) \qquad (5)$$

Therefore the new grouping will be defined as:

$| IC_1 (k) – C_1 | < | IC_1 (k) – C_2 |$ for k=1...........n_1;

$| IC_2 (k) – C_2 | < | IC_2 (k) – C_1 |$ for k=1...........n_2;

Where C_1 and C_2 are the local means of each class, n is the total number of independent components.

2.6 Evaluation of the Proposed Method

In order to evaluate and validate the effectiveness of the proposed method, three aspects have beenconsidered:

- Performance of the ICs identification classification system: ROC analysishas been done for all the features [9]. Actually, manual classification was performedby a trainedexpert investigatoron 10 signals (640 samples as discussed in section III). The manual classification is based on the IC time course, power spectrum and scalp topography. These results were compared to those obtained by the proposed method in order to calculate the accuracy of the classification system.

- Performance of the denoising method in terms of EEG neural data conservation:Two features have been computed on segments of the signals that were originally artifact free. Actually, these segments should remain intact after EOG denoising. Correlation coefficients were calculated between the processed portions and the artifact free corresponding signals. For optimal performance, the correlation coefficient should be close to 1 [10]. Furthermore, coefficient of waveform similarity was calculated in order to evaluate the performance of the reconstruction of the signal [11].

$$\varepsilon_{dB} = \frac{1}{M} * \sum_{i=1}^{M} (1 - mean(x_{i,seg}[n] - x'_{i,seg}[n])) \quad (6)$$

Where x_{seg} is a clean original EEG epoch, x'_{seg} is the corresponding corrected EEG epoch and M is the considered number of electrodes.

- Performance of the artifact removal: it is difficult to evaluate quantitatively this criterion since the considered dataset consists of recorded data and therefore the exact time series of the EOG noise interfering with the clean neurological EEG signal is unknown. However, the processed and the original signals should be highly correlated at all frequencies except the band containing the artifact [11]. This is verified by computing the coherence of the signals as depicted in equation 7 where x and y are the Fourier coefficients of the two signals and w1 and w2 are the bounds of frequency. The coherence values expected should be the lowest under 4Hz which is the EOG bandwidth.

$$C_{xy} = \frac{0.5.\sum_{W1}^{W2} \bar{x}^* \bar{y} + \bar{x}\bar{y}^*}{\sqrt{\sum_{W1}^{W2} \bar{x}\bar{x}^* . \sum_{W1}^{W2} \bar{y}\bar{y}^*}} \qquad (7)$$

Furthermore, an additional feature, the percentage of signal reconstruction is calculated for all electrodes [12]. This percentage should be low for the electrodes mostly contaminated by the eye blinking noise such as the frontal electrodes.

$$PR = (1 - \varepsilon) * 100 \qquad (8)$$

$$\varepsilon = \frac{X(n) - X_R(n)}{X_R(n)} \quad (9)$$

Where x (n) is the original EEG signal and x_R (n) is the reconstructed EEG signal.

3. RESULTS AND DISCUSSION

All mentioned methods were implemented on the EEG dataset from the international database on Matlab© as well as in EEGLab toolbox distributed under the free GNU GPL license for processing data in Electroencephalography [13].

3.1 Manual Ocular Artifact Identification

ICA decomposition of the 64 input channels was performed using the extended Infomax algorithm which gives stable decomposition up to 100 channels [14]. Once the decomposition is done, the 2D scalp map of each component was plotted, the power spectrum as well as the ICs time series. A trained investigator classified the ICs and by means of visual inspection, components 1, 3, and 6 for example, in the considered dataset were assigned as ocular artifacts since there scalp maps show a far frontal projection typical of an eye artifact; furthermore, they occupy a low frequency range but are generally strongest under 4 Hz what was verified when plotting their power spectrum.

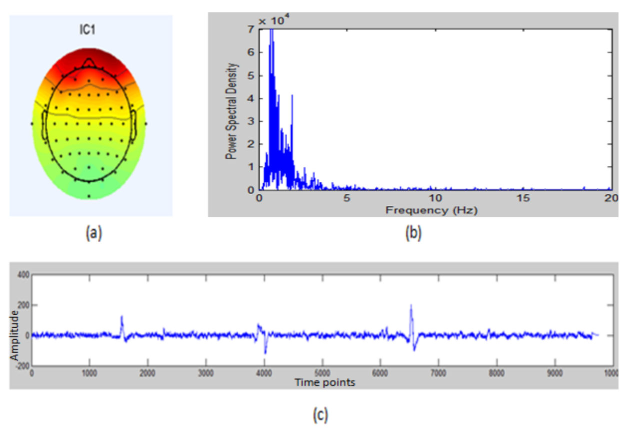

Figure 1 Example of an ocular artifact IC. (a) Scalp map; (b) Power spectrum; (c) signal time series.

3.2 Automatic Oocular Artifact Identification

- Average correlation: experimental results show that components 1, 3 and 6 in the considered dataset have the highest values regarding this feature. Fig. 2 shows the results of classification after applying Kmeans clustering on two classes; components 1, 3 and 6 were grouped in the first class which is the artifactual class.

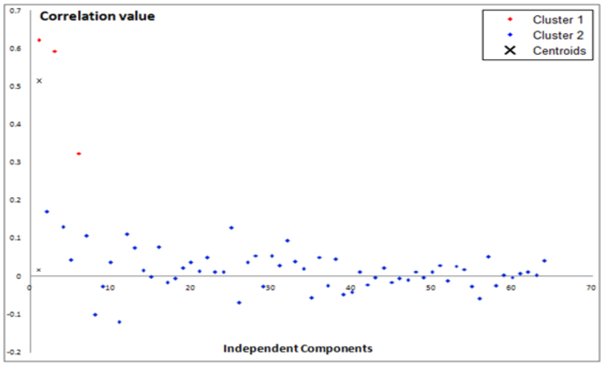

Figure 2 Kmeans clustering based on the correlationfeature.

- Distribution ratio: this feature vector shows that components 1, 3 and 6 have high values and are close to the artifact cluster centroid (10.41) however some non-ocular artifact ICs also show high values regarding this feature such as component 9 that shows a distribution ratio of 10.11.

- Maximum value: ocular artifacts are greater in amplitude than neurological signals. Experimental results show that components 1, 3 and 6 have the highest values regarding this feature. Kmeans clustering was performed on 3 classes. All components of class 1 are considered as artifactual.

3.3 Evaluation of the Proposed Method

- ROC Analysis: Accuracy, specificity and sensitivity were calculated for all feature vectors considering 640 samples. As each sample contains 10'000 data points, the analysis was conducted over 6'400'000 data points. Table 1 shows the results of the ROC analysis. Ocular artifacts were classified using the three features; the correlation feature gives the highest accuracy (percentage of correct predictions) of 99.54%. Distribution ratio gives the lowest accuracy (81.87%). However, some non-physiological artifacts such as extreme values and baseline drifts introduced during the acquisition of the EEG signals should be considered as it would affect the accuracy as well as the performance of the system. The results show that the classification is accurate, and therefore the considered, simple and easy to implement features could be used for the automation of an ICA-based denoising method.

Table 1 Results of the ROC analysis

Feature	ROC Analysis		
	Accuracy	*Specificity*	*Sensitivity*
Average correlation	**99.54%**	99.51%	100%
Distributio n ratio	81.87%	97.68%	14.7%
Maximum Value	98.75%	98.72%	100%

- Conservation of the EEG neurophysiological signals: The first feature which is the cross correlation between the free artifact segments before and after denoising have been computed over 64 electrodes.The average obtained value was 0.96. Fig. 3 shows a 2D mapping of the correlation all over the scalp illustrating high correlation at all electrodes except on the pre-frontal electrodes highly affected by the EOG noise. Furthermore, as a second objective criterion for measuring the conservation of the EEG neural signals, a coefficient of waveform similarity was calculated. The average computed over 64 sets was 0.037 showing that the clean signals, before and after denoising have 99.62% of similarity. The results demonstrate the conservation of the EEG intact signals due to the high values of similarity.

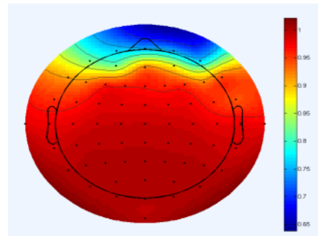

Figure 3 2D mapping of the correlation between the artifact free EEG signals before and after denoising.

- Ocular artifact removal: In order to evaluate the removal of the eye blinking artifact, coherence between the signals before and after denoising was plotted. Fig. 4 shows that the minimal coherence value is found in the bandwidth containing the artifact which is below 4 Hz.

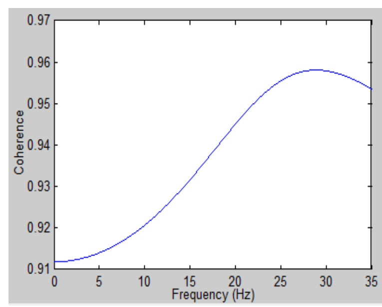

Figure 4 Coherence between the original and the processed signal showing a minimal value is the artifactual bandwidth.

Moreover, the percentage of signal reconstruction (PSR) was calculated over the 64 electrodes. Fig. 5 shows a 2D mapping of the PSR. Actually the percentage of the signal reconstruction is high on all electrodes between 80 to100 percent except on the prefrontal and the frontal electrodes, the most affected with eye blinking.

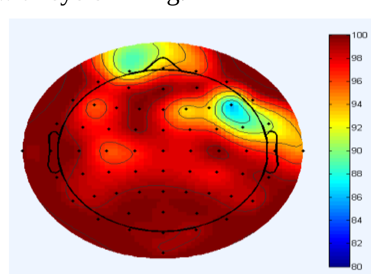

Figure 5 Mapping of the percentage of signal reconstruction evaluation criterion.

Furthermore, as a qualitative criterion for the evaluation of artifact removal, Fig. 6(a) shows an artifactual portion of the EEG signal containing an eye blinking artifact almost at time point 1600. Fig. 6(b) shows the corrected EEG signal with ocular artifact removal. While comparing Fig. 6(a) and Fig. 6(b), it is clear that the ocular artifact is perfectly removed and proves that Kmeans-ICA is effective and powerful in real contaminated EEG data providing a novel idea for preprocessing in EEG.

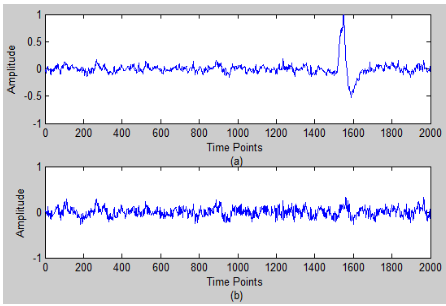

Figure 6 (a) Artifactual EEG segment before denoising, (b) Processed EEG segement. It is clear that the ocular artifact existing at time point 1600 has been removed.

On another hand, based on the EEG signal shown in fig. 1, the ocular artifact occurs on time point 6500; Fig. 7 shows the 2D scalp maps within the range of time points 6470 till 6550 with an increment of 10 time points of the signal before denoising. It shows the appearance of the eye blink (around time point 6500) mainly on the frontal lobe, how it is propagated through time from 6500-6540 as well as how it disappeared at the end of the eye blink as shown at time 6550.

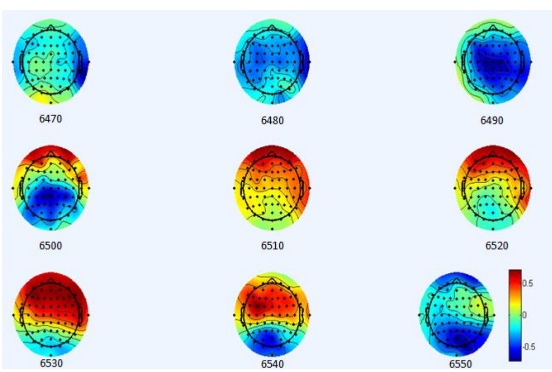

Figure 7 2D scalp maps of EEG before denoising showing the onset of the eye blinking and its propagation.

Fig. 8 shows the 2D scalp maps of the same data range with same increments after Kmeans-ICA denoising. While comparing Fig. 7 and Fig.8, it is clear that the far frontal activity, typical of an eye blinking artifact has been removed.

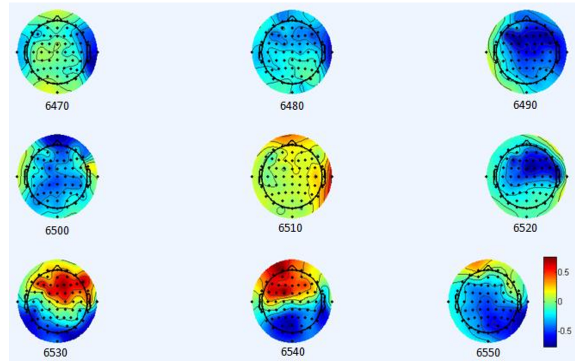

Figure 8 2D scalp maps after denoising.

4. CONCLUSION

The main contribution of this paper is the development of an adaptive, automatic ocular artifact removal method without the use of an additional reference such as an EOG channel. It is mainly based on Independent component analysis to which different traditional, easy to compute classifying features were added in order to automate the process. Classification is done using an adaptive thresholding by means of Kmeans clustering showing prominent and high accuracy. The results show that the proposed method is effective in terms of ocular artifacts classification, their removal as well as conservation of useful EEG neurological signals.

ACKNOWLEDGEMENT

Research supported by Higher Center Of Research-USEK

REFERENCES

[1]. J. Chambers, S. Sanei, *EEG Signal Processing. (*Chichester, England : John Wiley & Sons, Ltd., 2007).

[2]. Eleni Kroupi, Ashkan Yazdani, Touradj Ebrahimi, Jean-Marc Vesin**,** Ocular artifact removal from EEG: a comparison subspace projection and adaptive filtering method. *Eusipco, Proceedings of 19th European Signal Processing Conference*, Barcelona, Spain, 2011, EPFL-CONF-167489.

[3].Mehrdad Fatourechi, Ali Bashashati, Rabab K Ward, and Gary E Birch**,** EMG and EOG Artifacts in Brain Computer Interface Systems: A Survey, *Clinical Neurophysiology*, 118, 2007, 1388-2457.

[4]. Joyce Carrie, Irina Gorodnitsky, Marta Kutas, Automatic removal of eye movement and eye blink artifacts fromEEG data using blind component seperation,*Psychophysiology, 41*(2), 2004, 315-325.

[5]. S. Romero , MA. Mananas, J. Riba, A. Morte, S. Gimenez, S. Clos, MJ. Barbanoj,Evaluation of an automatic ocular filtering method for awake spontaneous EEG signals based on independent component analysis. *Proc. 26th Annual International Conference of the IEEE,*San Francisco, CA,2004, 925-928.

[6]. Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE**.***"Components of a New Research Resource for Complex Physiologic Signals.*" Circulation 101(23):e215-e220. *PhysioBank, PhysioToolkit, and PhysioNet.* [Online]

June 13, 2000 . http://circ.ahajournals.org/cgi/content/full/101/23/e215.
[7]. Clemens Brunner, Arnaud Delorme,Scott Makeig, Eeglab – an Open Source Matlab Toolbox for Electrophysiological Research, *Biomedical Engineering / Biomedizinische Technik,* 2013.
[8]. Rangaraj M. Rangayyan, Biomedical Signal Analysis.(Wiley-IEEE Press,Canada, 2002).
[9]. Kelly Zou, Aiyi Liu, Andriy Bandos Lucila Ohno-Machahdo, Howard Rockette, Statistical Evaluaion of Diagnostic Performance Topics in ROC Analysis(Boca Raton,Fl : CRC Press, 2012).
[10]. L. Vigon, M.R. Saatchi, J.E.W. Mayhew, R. Fernandes, Quantitative evaluation of techniques for ocular artefact filtering of EEG waveforms,*Science, Measurement and Technology, IEEE Proceedings 147*(5), 2000, 219 - 228.
[11]. J. Kelly, D. Siewiorek, A. Smailagic, J. Collinger, D. Weber and W. Wang, Fully Automated Reduction of Ocular Artifacts in High-Dimensional Neural Data,*IEEE Trans. on Biomed. Eng., 58,* 2011,*598 - 606.*
[12]. M. Gyanaprava, B.Kumar and M. Asit Kumar**,** Denoising of heart sound signal using wavelet transform,*International Journal of Research in Engineering and Technology*, *2*(4), 2013, 2319 - 1163.
[13]. A.Delorme, S. Makeig, EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independentcomponent analysis,*Journal of Neuroscience Methods*,*134*(1) 2004, 9-21.
[14]. Anthony Bell, Terrence Sejnowski**,** An Information-Maximization Approach to Blind Separation and Blind Deconvolution,*Neural Computation, 7*(6),1995,1129-1159.
[15]. J. Richman, J. Moorman,Physiological time-series analysis using approximate entropy and sample entropy, *American Journal of physiology, heart and circulatory physiology, 278*(6),2000, 2039-2049.