# A Method for Automatic Removal of Eye Blink Artifacts from EEG Based on EMD-ICA

Mumtaz Hussain Soomro, Nasreen Badruddin, *Member, IEEE*, Mohd Zuki Yusoff, Aamir Saeed Malik, *Senior Member, IEEE.*

Department of Electrical and Electronic Engineering, Universiti Teknologi Petronas, 31750 Tronoh, Perak, Malaysia.
mumtaz.muet@gmail.com, {Nasreen.b, mzuki_yusoff & aamir_saeed}@petronas.com.my

***Abstract*— The electroencephalography (EEG) recordings are mostly contaminated by eye blink artifacts. It is very difficult to analyze and interpret the EEG signal due to frequent occurrence of the eye blink artifact. In this paper, a new hybrid algorithm that automatically removes the eye blink artifact from the EEG, based on Empirical Mode Decomposition (EMD) and Independent Component Analysis (ICA) is proposed. The proposed algorithm is evaluated on simulated EEG to calculate correlation coefficient and signal-to-artifact ratio (SAR). A non-corrected EEG was simulated to have a SAR of -19.1673 dB. From the simulation results, the highest average correlation coefficient and SAR of corrected EEG from non-corrected EEG are obtained as 0.871094 and 2.71645 dB respectively by applying proposed algorithm. The results demonstrate that proposed method recovers the EEG data by removing the eye blink artifacts reliably. In addition, the proposed method is applied on real spontaneous EEG data with eye blink artifact.**

***Key words*—Electroencephalography (EEG), Eye Blink artifacts removal, Empirical mode decomposition (EMD), Independent Component Analysis (ICA), FastICA, Correlation Coefficient, Signal-to-artifact ratio (SAR).**

## I. INTRODUCTION

Electroencephalography (EEG) is an important tool to record electrical activity on the human brain. The EEG is recorded by placing the electrodes on scalp of a person using 10-20 electrode placement system with high temporal and low spatial resolution [1]. In medical domain, the EEG is used to investigate neurological disease such as epilepsy. It is very difficult to understand the EEG signal when it is corrupted with other non-cerebral signals, known as EEG artifacts or noise. The most prominent artifact is the eye blink artifact that mostly contaminates the EEG signal [2]. The EEG signal has many variations in terms of shape, amplitude and frequency. If EEG signal has amplitude below 20μV then it is consider as low, if it lies in the range of 20μV to 50μV then it is considered as medium and it is considered as high with amplitude above than 50μV [3-5].

The eye blink artifact has amplitude that is 10 to 100 times greater than cerebral potentials and can last up to 200 ms to 400 ms [5]. Therefore, the EEG signal has low signal-to-artifact ratio that makes it difficult to interpret the EEG signal. The artifact can be from the muscle activation, eye blink and an eye moment. The frequent occurrence of eye blink artifacts can seriously interfere with the interpretation of the EEG signals. Hence, several de-noising techniques are employed for the removal of the eye blink artifact in recent years. Some regression-based methods include time and frequency domain regression techniques were employed [6-7]. These regression-based approaches involve calibration tests to determine propagation coefficient $\gamma$ between Electrooculogram (EOG) channel and each channel of the EEG, where the each EEG channel is to be subtracted from separately recorded EOG to remove eye blink artifact. In fact, EOG signal to be subtracted from the recorded EEG signal is also contains some EEG information that introduces loss in desired information [7-8]

In [9] principle component analysis (PCA) is used to remove the eye blink artifacts. PCA transforms a set of multivariate data with $n$ correlated components into a set of uncorrelated components by finding the orthogonal directions of the largest variance in the EEG signals. The PCA outperformed the regression based methods. However, PCA separate the artifact but not completely, especially when EOG (eye blink) artifact and EEG have comparable amplitudes; due to lack of higher order statistical dependencies.

In [10-12] blind source separation approaches (i.e. Independent Component Analysis, ICA) are used to remove EEG artifacts. ICA is blind source separation method that separates the mixed signal into mutually statistical independent components. Compared to PCA, ICA recovers the underlying sources by maximizing the statistical independence by removing the constraint of orthogonality instead of forcing the components to be uncorrelated. However, ICA lacks the variance maximization property possessed by the PCA [4]. Furthermore, ICA needs visual inspection to select components manually for correction, thus creating challenges for implementing automated correction routines.

In [13], the authors proposed a novel technique to automatically remove eye-blink artifacts from the EEG data. They used EEG raw data and decomposed them into independent components by independent component analysis (ICA) and Peak Detection Algorithm of Independent Component (PDAIC) is used to identify eye-blink artifact components. This algorithm is automatic but it needs little visual inspection and also introduces computational complexity.

In [14-16], the authors have suggested that FastICA algorithm is more efficient and robust than other ICA

978-1-4673-5609-1/13/$31.00 ©2013 IEEE

algorithms, as FastICA is parallel, distributed, and computationally simple and also requires less memory.

In this paper, a new method for removal of eye blink artifact in EEG data is proposed that combines empirical mode decomposition (EMD) and fast independent component analysis (FastICA). First, the EMD is applied on the most affected channel, i.e. Fp1 to extract eye blink artifact as a template without any need of visual inspection. Then this extracted eye blink template will go to FastICA as an input along with other contaminated-EEG channels in order to separate the eye blink from other channels. The performance of proposed method is validated on simulated data in terms of correlation coefficient and signal-to-artifact ratio. Furthermore, the proposed method is tested on a real spontaneous contaminated EEG signals.

The rest of this paper is organized as follows: Section II "Theoretical Background" describes basic concepts of EMD and FastICA. Section III "Proposed Methodology" presents the procedure and implementation of the proposed algorithm in detail. Section IV "Results" presents simulation results and discussion.

## II. THEORETICAL BACKGROUND

### A. Empirical Mode Decomposition (EMD)

The EMD decomposes real signal whether it is linear, nonlinear or non-stationary, into a set of components, called "Intrinsic Mode Functions (IMFs)" [17-18]. The IMFs have time-variable amplitudes and frequencies [18]. The IMFs can be achieved by shifting out the dominant modes (rapidly oscillating components) from the real data, by subtracting iteratively from slowly oscillating components (less dominant modes). The slow oscillating components are known as local mean of the real data. When all dominant modes are extracted, an IMF has local mean of zero and shifting process will be stopped. More precisely, if $X(k)$ is input signal, then EMD decomposes the $X(k)$ into intrinsic mode functions denoted by $\{d_i(k)\}_{i=1}^{N}$ such that:

$$X(k) = \sum_{i=1}^{N} d_i(k) + r(k) \quad (1)$$

Where $r(k)$ denotes the residual monotonic function which reflects the average trend with in the original one signal. In order to get meaningful estimation of instantaneous frequency, the IMF should be designed as close symmetric around the local mean and their number of extrema and zero-crossing must be equal or differ at most by one [17]. The intrinsic mode functions can be obtained by shifting process, described as:

i. Find all extrema (minimum & maximum) of $X(k)$.

ii. Interpolate (using cubic spline interpolation) between maximum (minimums) to obtain signal upper envelope $e_u(k)$ (a lower envelop $e_l(k)$ )

iii. Calculate the local mean $m(k) = (e_u(k) + e_l(k))/2$

iv. Subtract $m(k)$ from $X(k)$ to construct oscillating signal $h(k) = X(k) - m(k)$

v. If $h(k)$ satisfies all stopping conditions, $d(k) = h(k)$ becomes an IMF; otherwise repeat step1 by setting $X(k) = h(k)$

### B. FastICA

Independent Component Analysis (ICA) is a blind source separation technique for separating multivariate observed random data into mutually statistically non-gaussian independent subcomponents. Let $X(t)$ be observed random vector and $S(t)$ be the mutually statistically independent source components for each sample value M, given as,

$$X = AS \quad (3)$$

Where, $A$ is $M \times M$ unknown mixing matrix and should be square matrix. The main goal of the ICA is to estimate the demixing matrix $W$ (i.e. inverse of $A$) which is to be transformed onto $X(t)$ to obtain the source components, such that:

$$S = WX \quad (4)$$

In order to implement ICA algorithm successfully, five assumptions must be satisfied. First, the sources must be mutually statistically independent and must have nongaussian distribution. Second, the mixing matrix should be square matrix that yields number of mixtures must be at least same as number of the sources. Third, there should not be external noise in the ICA model. Fourth, the observation vector $X$ should have zero mean, if it is not true, then in order to make model zero mean; the data have to be centered by subtracting the mean from each observable variable $x_i$. Fifth, the observable vector $X$ must be whitened. The main purpose of whitening the data to make the mixed observed variables uncorrelated; which is described in [5], given as:

$$F = JX \quad (5)$$

J is known as whitening matrix and F is new matrix which is white. Here, X is zero mean-vector, has covariance matrix $C_x$ is given by:

$$C_x = E\{XX^T\} = EDE^T$$
$$J = D^{-1/2}E^T \quad (6)$$

Where, E is orthogonal matrix of Eigen vectors of the covariance matrix $C_x$ and D is diagonal matrix and the diagonal values are the Eigen values of the $C_x$.

FastICA is an improved version of ICA. It is fast fixed-point iterative algorithm that maximizes the non-gaussianity which is measured by approximation of negentropy. The implementation of FastICA algorithm is summarized in as below:

i. Center the data to make its mean to be zero.

ii. Whiten the data to give H using formula (5) by choosing bigger $m$ elements from their corresponding Eigen vector.

iii. Choose an initial random demixing matrix $W$ for every $w_i, i = 1,2,3,...,m$, each of unit norm. Then the demixing matrix is to be orthogonalized.

iv. Let $w^+ = E\{Hg(w^T H)\} - E\{g'(w^T H)\}w$, hyperbolic tangent function g is chose to renew $W_i$.

v. Symmetric decorrelation (i.e. orthogonalization) of the matrix $W = w_i^T$, $i=1,2,..,m$ is performed by $W = W / \sqrt{WW^T}$.

vi. If not converged, go back to step 4.

The symmetric orthogonalization is achieved by first applying the iterative step of one unit algorithm on every vector $w_i$, and then all $w_i$ are to be orthogonalized by special symmetric decorrelation methods. In this way, demixing matrix $W$ is obtained by the iteration which separates the observed signals into independent components.

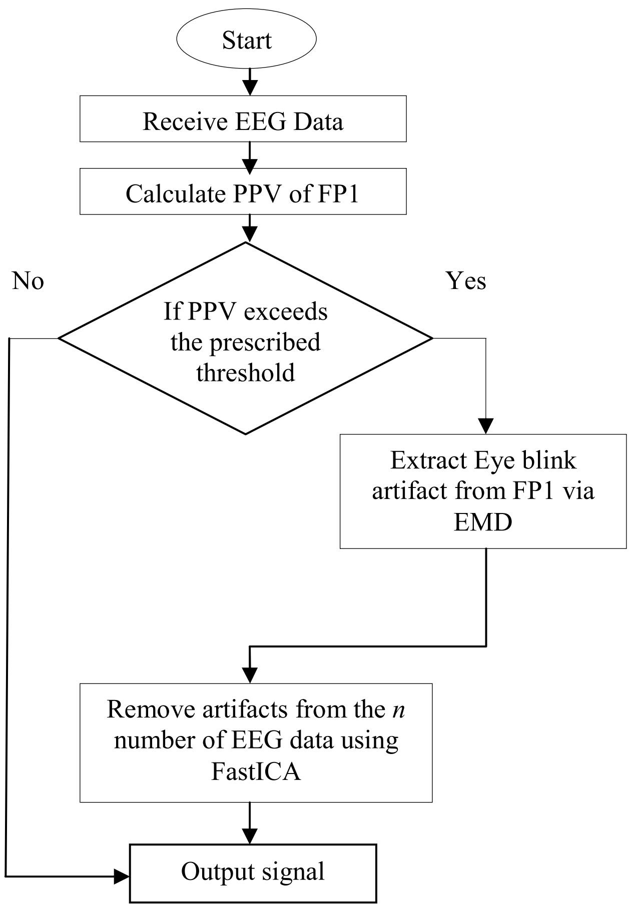



Fig. 1. Flow chart of EMD-FastICA method

## III. Proposed Methodology

In this paper, a new hybrid algorithm EMD-FastICA is presented. The procedure for extracting and removing the eye blink artifact is summarized as follows:

i. Once the contaminated-EEG segment is detected by calculating peak to peak value (PPV) of FP1 or FP2 EEG channel (as the eye blink artifact is the most dominant over Fp1 or Fp2) comparing with prescribed threshold.

ii. If PPV is larger than prescribed threshold value, apply EMD to the recorded EEG segment to extract eye blink artifact as a template.

iii. Then apply FastICA on the recorded EEG data along with extracted eye blink template to obtain mixing matrix and independent components.

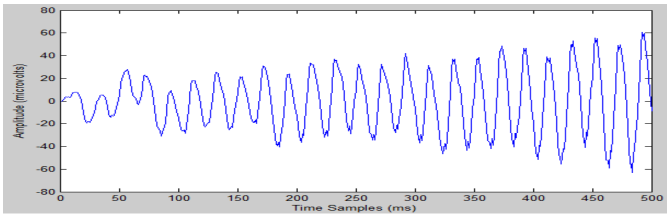
iv. Reconstruct non-artifactual EEG components.

The flow chart for EMD-FastICA method is illustrated in Fig. 1.

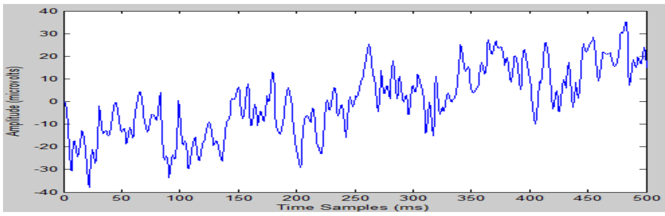
## IV. Results

### A. Simulated Data

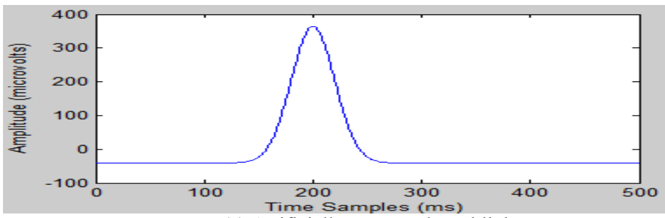
The main aim of simulated data is to evaluate the performance of the proposed method in removal of eye blink artifacts in terms of correlation coefficient and signal-to-artifact ratio (SAR). In the simulation study, the artificially EEG and eye blink is generated by autoregressive and Gaussian function which is mentioned in [19]. Here, two EEG signals are generated artificially along with artificially generated eye blink as shown in Fig. 2.




(a) Artificially generated EEG channel1



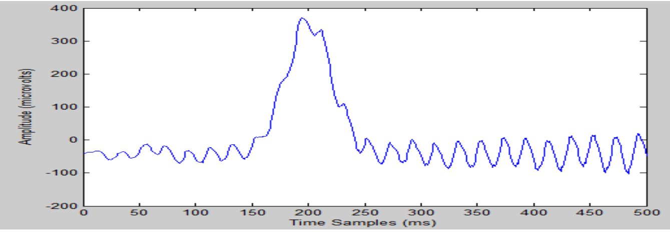

(b) Artificially generated EEG channel2



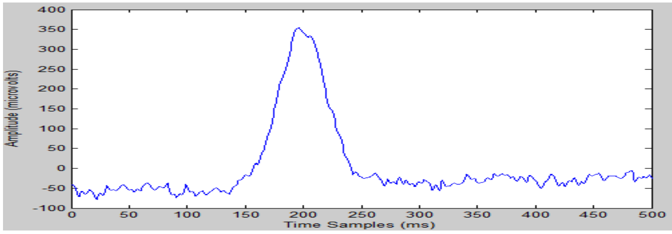

(c) Artificially generated eye blink

Fig. 2. Artificially generated EEG channel 1 & 2 and eye blink shown in (a), (b) and (c) respectively.

Generally, the artificially generated EEG and eye blink are mixed to each other to create contaminated EEG that needs to be processed and extracted using proposed technique as shown in Fig. 3.



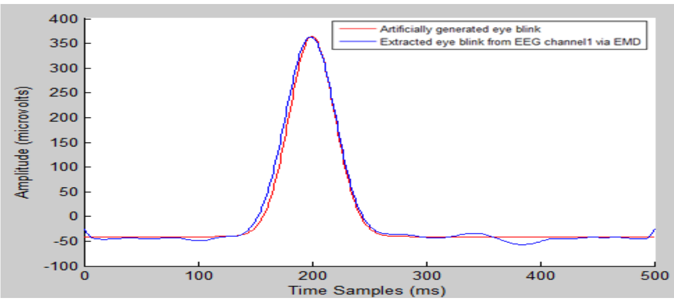

(a)



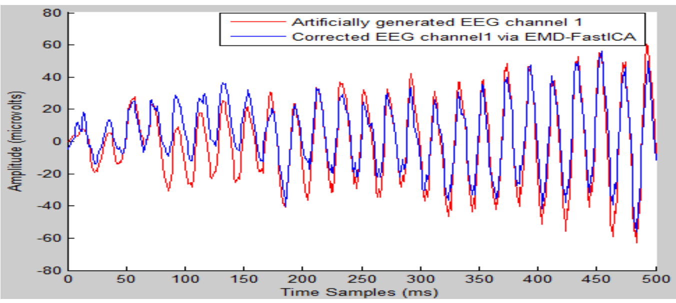

(b)

Fig. 3. Contaminated EEG channel 1 & 2 shown in (a) and (b) respectively

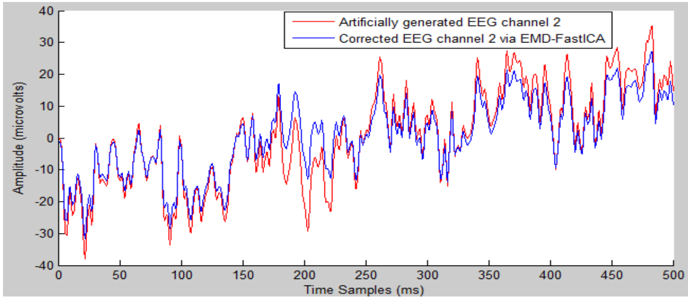
After that, the proposed algorithm is applied on the contaminated EEG channels shown in Fig. 3 that gives following results, shown in Fig. 4. According to methodology of proposed algorithm shown in the flow chart Fig. 1, first EMD is applied on channel 1 to extract eye blink that works as template in FastICA algorithm for removal of eye blink from the EEG channels.




(a) Extracted eye blink via EMD (blue) and artificially generated eye blink (red).




(b) Corrected EEG channel 1 via EMD-FastICA (blue) and artificially generated EEG channel 1(red)




(c) Corrected EEG channel 2 via EMD-FastICA (blue) and artificially generated EEG channel 2(red)

Fig. 4. Visual comparison of EMD-FastICA results corresponding to (a) extracted eye blink via EMD (b) artificially generated EEG channel 1 and (c) artificially generated EEG channel 2.

TABLE I. Average Correlation Coefficient and Elapsed Time for 100 Runs

| Methods | Correlation Coefficient EEG channel1 | Correlation Coefficient EEG channel 2 | Elapsed Time (Seconds) |
|---|---|---|---|
| FastICA | 0.830294 | 0.319057 | **0.008103** |
| EMD-FastICA | **0.871094** | **0.859781** | 0.657801 |

TABLE II. Average Signal-to-Artifact Ratio (SAR) for 100 Runs

| Channels | SAR Non-corrected | SAR FastICA | SAR EMD-FastICA |
|---|---|---|---|
| Channel 1 | -19.1673 | 0.84276 | **2.71645** |
| Channel 2 | -21.0274 | -2.48361 | **1.04761** |

## *B. Validation*

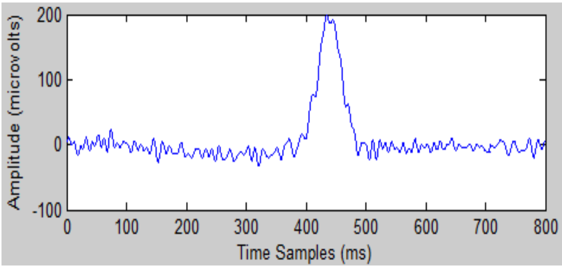
In order to validate the proposed algorithm, the correlation coefficient is used to calculate the similarity between EEG sources and the corrected EEG data wave forms.

Besides the correlation, in order to quantify the degree of the removal of eye blink artifact, the signal-to-artifact ratio is defined as:

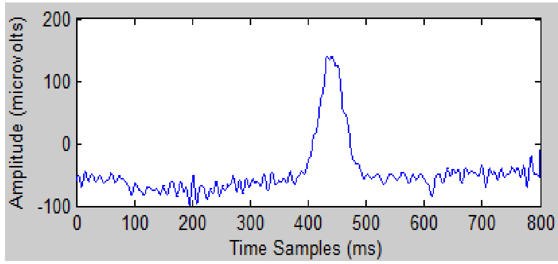
$$SAR = 10 \cdot \log\left(\frac{std\,(x)}{std\,(x - \hat{x})}\right) \tag{7}$$

Where $x$ is the original artificially generated EEG and $\hat{x}$ is the corrected EEG. Table I shows the 100 runs averaged values of correlation coefficient. The average correlation coefficients for two different corrected EEG channels 1 and 2 using the proposed algorithm are obtained as 0.87026 and 0.84986 respectively; shows that the artificially generated EEG and the corrected EEG data are relatively closed. Table II shows the 100 runs averaged SAR. The better improvement in SAR is obtained by applying EMD-FastICA shows that the proposed algorithm removes so for artifacts.

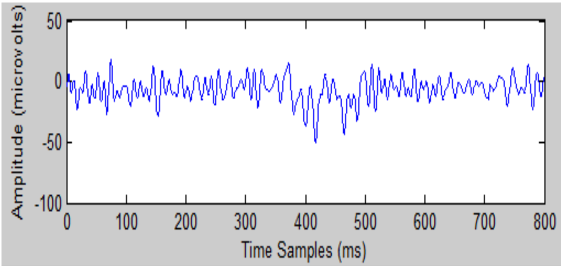
Considering the correlation coefficient and SAR results, we concluded that EMD-FastICA based hybrid algorithm removes the most eye blink artifacts from the contaminated EEG signals by preserving the most EEG signals.




Fp1

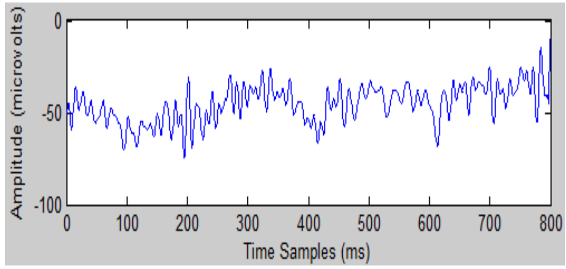



F3




EMD-FastICA

(a)




EMD-FastICA

(b)

Fig. 5. Visual inspection between before and after applying EMD-FastICA algorithm using 800 ms real-time original EEG data, (a) Fp1 and (b) F3 original EEG channels.

### *C. Real Data*

Fig.5 shows an example of 800 ms seconds spontaneous real time EEG signals sample at the rate of 250Hz before and after applying the EMD-FastICA method. Fp1 and F3 channels are used as these channels are the most contaminated with eye blink artifact. The visual inspection of the proposed method's results are shows in Fig.5, demonstrates that EMD-FastICA algorithm removes the eye blink artifact efficiently from the real time spontaneous EEG signals by preserving the actual neural signals.

## V. Conclusion

In this paper, a new method for removal of eye blink artifacts in EEG data based on empirical mode decomposition FastICA (EMD_FastICA) is presented. The simulation results show that the proposed method has achieved the highest averaged correlation coefficient of 0.871094 and improved SAR of 2.71645 dB for corrected EEG, while the non-corrected EEG was simulated to have a SAR of -19.1673 dB. Furthermore, the proposed algorithm is easily realized and corrects the eye blink artifact automatically without any visual inspection. However, the elapsed time is slightly greater than FastICA. Therefore, this technique is not suitable for real-time removal of eye blink artifacts. As future work, we will investigate other techniques to remove eye blink artifacts in less time to make it suitable for real-time applications.

## V. Acknowledgment

The authors would like to thank Dr. Tahamina Begum of Department of neurosciences, Hospital Universiti Sains Malaysia, Kota Bharu for EEG data.

## References

[1] S. Sanei and J.A. Chambers, EEG Signal Processing. NewYork:Wiley, 2007.

[2] Li YD, Ma ZW, Lu WK, "Automatic removal of the eye blink artifact from EEG using an ICA-based templated matching approach". PhysiolMeas, vol. 27, pp. 425–436, 2006.

[3] Babu, P. A. and K. V. S. V. R. Prasad, "Removal of Ocular Artifacts from EEG Signals Using Adaptive Threshold PCA and Wavelet Transforms". Communication Systems and Network Technologies (CSNT), IEEE International Conference, pp.572-575, 3-5, 2011.

[4] Kiamini, M., S. Alirezaee, B. Perseh., M. Ahmadi, "A wavelet based algorithm for Ocular Artifact detection In the EEG signals". Multitopic Conference, INMIC, IEEE International, 2008.

[5] K. Naraharisetti "Removal of Ocular artefacts from EEG Signal using Joint Approximate Diagonalization of Eigen Matrices (JADE) and Wavelet Transform", Canadian Journal on Biomedical Engineering & Technology Vol. 1, No. 4, July 2010.

[6] G.Gratton, M.G. Coles, E. Donchin, "A new method for off-line removal of ocular artifact". Electroencephalogr Clin. Neurophysiol. Vol. 55, pp. 486- 484, 1983.

[7] Wallstrom, G. L., R. E. Kass, "Automatic correction of ocular artifacts in the EEG: a comparison of regression-based and component-based methods." International Journal of Psychophysiology, vol. 53, pp. 105-119, 2004.

[8] Fatourechi, M., A. Bashashati, "EMG and EOG artifacts in brain computer interface systems: A survey." Clinical Neurophysiology, vol. 118, pp. 480-494, 2007.

[9] T. D. Lagerlund, F. W. Sharbrough and N. E. Busacker, "Spatial filtering of multichannel electroencephalographic recordings through principal component analysis by singular value decomposition," J. Clin. Neurophysiol. vol. 14, pp. 73-82, 1997.

[10] Delsanto, S., F. Lamberti, "Automatic ocular artifact rejection based on independent component analysis and eye blink detection". Neural Engineering Conference Proceedings. First International IEEE EMBS Conference, 2003.

[11] Joyce CA, Gorodnitsky IF, Kutas M, "Automatic removal of eye movement and blink artifacts from EEG data using blind component separation". Psychophysiology vol. 41, pp. 313–325, 2004.

[12] Li YD, Ma ZW, Lu WK, Li YD, "Automatic removal of the eye blink artifact from EEG using an ICA-based templated matching approach". PhysiolMeas vol 27, pp. 425–436, 2006.

[13] Junfeng, G., L. Pan, "Automatic removal of eye-blink artifacts based on ICA and peak detection algorithm". Informatics in Control, Automation and Robotics (CAR), 2nd IEEE International Asia Conference, pp.22-27, 6-7 March 2010.

[14] A. Hyvärinen, E. Oja, "Independent component analysis: algorithms and applications", Neural Networks, Vol. 13, Issues 4–5, pp. 411-430, June 2000.

[15] D Langlois, S Chartier, D Gosselin, "An Introduction to Independent Component Analysis: Infomax and FastICA Algorithms", Tutorials in Quantitative Methods for Psychology, Vol. 6(1), pp. 31-38, 2010.

[16] G Geetha and S N Geethalakshmi, "Artifact Removal from EEG using Spatially Constrained FastICA and Fuzzy Shrink Thresholding Technique", Article in International Journal of Applied Information Systems, vol. 4, No. 11, pp. 25-29, December 2012.

[17] Rehman, N. and D. P. Mandic, "Empirical Mode Decomposition for Trivariate Signals." IEEE Transactions on Signal Processing, Vol. 58, issue No. 3, pp.1059-1068, 2010.

[18] P. Trnka, and Hofreiter, "The Empirical Mode Decomposition in Real-Time", In Proceedings of the 18th International Conference on Process Control, Tatranská Lomnica, Slovakia, ISBN 978-80-227-3517-9, pp. 284–289, 2011.

[19] Mohd Zuki, "Generalized Subspace Approach for Measurement of Latencies in Visual Evoked Potentials", *Thesis*, IRC, UTP, Malaysia 2010.