

# EYEBLINK ARTEFACT REMOVAL FROM EEG USING INDEPENDENT COMPONENT ANALYSIS

R. Benazir Begam<sup>1</sup>, B.Thilakavathi<sup>2</sup>

<sup>1</sup>PG Scholar, M.E Communication Systems, Rajalakshmi Engineering College, TamilNadu, India

<sup>2</sup>Associate Professor, Department of ECE, Rajalakshmi Engineering College, TamilNadu, India

## Abstract

The electrical activity of the human brain is recorded using Electroencephalogram (EEG). Most Often, an EEG signal is contaminated with some unwanted signals called artefact. Artefacts are due to eyeblink (EB), eyeball movement, sweat, power line and muscle. It is very essential to remove such artifact from an EEG signal without losing integrity. In this project, detection and removal of eyeblink artefact has been carried out. Linear feature has been taken to detect an EB artefact in an EEG signal and it's been validated with Spectrogram. Using independent component analysis (ICA), EB artifacts are removed. This project uses some of the algorithms like Algorithm for multiple unknown source extraction (AMUSE), Second order blind identification (SOBI), Joint approximate diagonalization of Eigen matrices (JADE) and Second order non stationary source separation (SONS) which performs ICA for removal of EB artifact. Correlation coefficient is calculated between an original signals and reconstructed signals which are obtained using the above four algorithms. It is found that SOBI and JADE algorithm performs better when compared to other two algorithms.

**Keywords:** artefact, channel, eyeblink(EB), independent component analysis(ICA) , spectrogram.

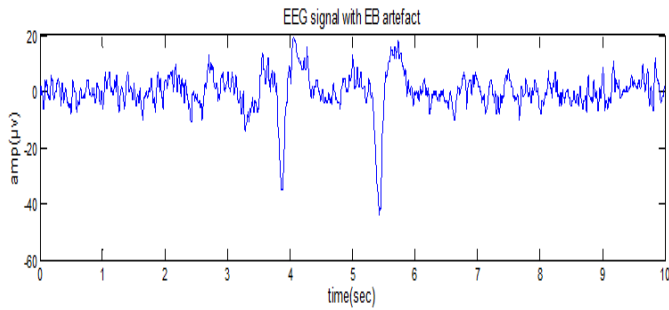
-----\*\*\*-----

## 1. INTRODUCTION

The Electroencephalograph (EEG) is an essential way to determine the activities of the human brain. [1]. Signals that are detected by EEG but not belong to a cerebral origin are called artefacts. The amplitude of artefacts is larger than the size of amplitude of the cortical signals. The artefacts can be divided into two, one artefact is related to the subject and the other artefact is related to the system. The subject-related or internal artefacts are body movements, muscle movement, heart beating, eyeball movement, eyeblink and sweating. The system artefacts are 50Hz interference, impedance fluctuation, cable defects, and electrical noise from the electronic components. Often in the preprocessing stage these artefacts are highly mitigated and the informative information is restored [2]. In current data acquisition, EB artefacts are more dominant over other electrophysiological contaminating signals such as heart and muscle activity, head and body movement, as well as external interferences due to power sources. Hence, devising a method for successful removal of EB artefact from EEG recordings is an essential. Thus, our aim focuses to eliminate EB in this paper. This paper is organized as follows. Section 2 contains information about the signals which are used for simulation. In section 3, we discuss feature extraction for the identification of EB artefact and for validating the results, we use spectrogram. Section 4, explains the different blind source separation (BSS) algorithms for the removal of EB artefacts. The results obtained are shown in section 5 followed by conclusion in section 6.

## 2. DATA COLLECTION

The digital EEG is collected from Aarthi Scans, Chennai, for 10 different subjects and all the subjects are age matched. Sampling frequency of the EEG is 256 Hz. The length of the eyeblink data is 3 minutes. The sensitivity of the equipment of the equipment is set to 7.5µm/mm and the filter frequency range kept between 0.1Hz-70Hz. The Standard 10-20 electrode system is used in monopolar montage with linked ear as reference which is shown in Fig 1. Therefore, totally 16 (FP2 F4 C4 P4 O2 F8 T4 T6 FP1 F3 C3 P3 O1 F7 T3 T5) electrodes with A2 and A1 are used for this study. We have considered 10 seconds of EEG data for offline analysis. The Eyeblink artefacts from an EEG signal has been identified and removed without losing the useful information. Identification process has been done using MATLAB and removal process has been carried out in ICALAB. The presence of EB in the EEG signal is identified by the peaks in the signal. In the Fig.1, EB is present in 3-4 & 5-6 intervals.



**Fig-1:** Eyeblink EEG of FP2 channel for one subject

### 3. FEATURE EXTRACTION

The original EEG signal is a time domain signal and the signal energy distribution is scattered. The signal features are buried away in the noise. In order to extract the features, the EEG signal is analyzed to give a description of the signal energy as a function of time or/and frequency [3]. Features are of two types. One is, linear features and the other is, non linear features. In this paper, we use only linear feature.

#### 3.1 Linear Features

Linear feature denotes the amplitude-dependent properties of an EEG signal [4]. The EEG signal is stochastic, and each set of samples is called realizations or sample functions  $\{x(t)\}$  [5]. Some of the linear features are mean, variance, skewness and kurtosis.

The expectance ( $\mu$ ) is the mean of the realizations and is called first-order central momentum [5]. Mean of the signal gives the average information content. The mean of the given data sequence can be calculated using formula in equation 1. Here,  $n$  denotes the length of the sample.

$$\text{mean} = \frac{x(1) + x(2) + \dots + x(n)}{n} \quad (1)$$

Variance measures how far a set of samples are spread out. The second-order central momentum is the variance of the realizations. The square root of the variance is the standard deviation ( $\sigma$ ), which measures the spread or dispersion around the mean of the realizations [5]. The variance of the given data sequence can be calculated using formula in equation 2

$$\text{variance} = E(x - \mu)^2 \quad (2)$$

Where  $\mu$  denotes mean and  $E$  denotes statistical expectation. Skewness is a measure of the lack of symmetry of the distribution. The skewness of the given data sequence can be calculated using formula in equation 3.

$$\text{skewness} = E \left\{ \left[ \frac{x - \mu}{\sigma} \right]^3 \right\} \quad (3)$$

Where  $\mu$  and  $\sigma$  are the mean and standard deviation respectively, and  $E$  denotes statistical expectation.

Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution. The kurtosis of the given data sequence can be calculated using formula in equation 4.

$$\text{kurtosis} = E \left\{ \left[ \frac{x - \mu}{\sigma} \right]^4 \right\} \quad (4)$$

Where  $\mu$  and  $\sigma$  are the mean and standard deviation respectively, and  $E$  denotes statistical expectation.

#### 3.2 Short Time Fourier Transform

Short time Fourier transform (STFT) is the time-dependent Fourier transform for a sequence, and it is computed using a sliding window. The Short Time Fourier Transform evaluates the frequency (and possibly the phase) change of a signal over time. To achieve this, the signal is cut into blocks of finite length called window, and then the Fourier transform of each block is computed. To improve the result, blocks are overlapped and each block is multiplied by a window that is tapered at its end points [6].

The mathematical representation of Short Time Fourier Transform (STFT) is written as:

$$x(t, f) = \sum_{r=-L}^L x(s-t)w(s)e^{-j2\pi ft} \quad (5)$$

Where  $f \in \{0, 1/L f_s, \dots, L-1/L f_s\}$  is a frequency,  $w(s)$  is a window that tapers smoothly to zero at each end and  $t$  represents discrete time and  $L$  indicates the length of the window. The graphical display of the magnitude of the Short Time Frequency Transform is called the spectrogram of the signal [6]. Using this, we can identify the presence of eyeblinks in EEG signal.

### 4. INDEPENDENT COMPONENT ANALYSIS

Independent component analysis (ICA) is found to be useful in extracting independent components of the signal. This is very helpful in eliminating EB artefact from EEG signal. Independent Component Analysis, as the name implies, can be defined as the method of decomposing a set of multivariate data into its underlying statistically independent components. Hyvarinen and Oja [7] rigorously define ICA using the statistical "latent variables" model. Under this model, we

observe  $n$  random variables  $x_1, x_2, \dots, x_n$  etc which are linear combinations of  $n$  random latent variables  $s_1, s_2, \dots, s_n$  as:

$$X_i = a_{i1}s_1 + a_{i2}s_2 + \dots + a_{in}s_n \text{ for all } i = 1, \dots, n$$

Where  $a_{ij}, j = 1, \dots, n$  are some real coefficients. By definition, the sources  $s_i$  are *statistically independent*. The “latent variables” are the sources,  $s_i$ , which are also called the independent components. They are called “latent” because they cannot directly be observed. The independent components,  $s_i$ , and the mixing coefficients,  $a_{ij}$ , are not known and must be determined (or estimated) using only the observed data  $x_i$ .

The ICA latent variables model is better represented in matrix form. If  $\mathbf{S} = [s_1, s_2, s_3, \dots, s_n]^T$  represents the original multivariate data that is transformed through some transformation matrix  $\mathbf{H}$  producing  $\mathbf{X}$  such that:

$$\mathbf{X} = \mathbf{H}\mathbf{S} \quad (6)$$

Then ICA tries to identify an unmixing matrix  $\mathbf{W}$  such that:

$$\mathbf{W} = \mathbf{H}^{-1} \quad (7)$$

so that the resulting matrix  $\mathbf{Y}$  is:

$$\mathbf{Y} = \mathbf{W}\mathbf{X} = \mathbf{W}(\mathbf{H}\mathbf{S}) = \mathbf{S} \text{ (since } \mathbf{W} = \mathbf{H}^{-1}) \quad (8)$$

As stated earlier, the only thing ICA demands is that the original signals  $s_1, s_2, \dots, s_n$ , be at

any time instant  $t$  *statistically independent* and the mixing of the sources be linear.

There are several algorithms which extracts independent component. In this paper, we are going to use four algorithms namely Algorithm for Multiple Unknown Signals Extraction algorithm (AMUSE), Second Order Blind Identification (SOBI), Second Order Nonstationary Source separation (SONS), Joint approximate of Diagonalization of Eigen matrices (JADE).

#### 4.1 Amuse

This is a batch-type (non-iterative) blind separation learning algorithm based on second order statistics. This algorithm is well suited for temporally correlated sources [8]. The learning procedure in the AMUSE is consists of following steps:

1. Whitening procedure is applied to the input signal  $x(t)$ , which is described by  $Z(t) = Qz(t)$  where,  $Q = S^{-1/2}G^T$  and  $G$  is obtained from the Eigen value decomposition:  $C_{xx} = E[x(t)x^T(t)] = GSG^T$ .
2. Singular Value Decomposition is applied to the (p-lag) time delayed correlation matrix

$C_{zz}(p) = E(z(t)z^T(t-1)) = U\Sigma V^T$  where,  $\Sigma$  is the diagonal matrix that contains singular values and  $U$  &  $V$  are two new orthogonal matrices.

3. The demixing matrix is calculated as  $W = U^TQ$ .

#### 4.2 Sobi

The objective of this procedure is to find the orthogonal matrix  $U$  which diagonalizes a set of matrices. The SOBI algorithm can be performed as follows [9]

1. Perform robust orthogonalization  $X(k) = Qx(k)$ .
2. Estimate the set of covariance matrices:

$$\hat{R}_{\bar{x}}(p_1) = \left(\frac{1}{N}\right) \sum_{k=1}^N \bar{x}(k)\bar{x}^T(k-p_1) = Q\bar{R}_x(p_1)Q^T \text{ for}$$

a preselected set of time lags  $(p_1, p_2, \dots, p_L)$ .

3. Perform JAD:  $R_{\bar{x}}(p_1) = U D_1 U^T, \forall t$  i.e., estimate the orthogonal matrix  $U$  using one of the available numerical algorithms.
4. Estimate the source signals as  $\bar{x}(k) = U^T Q x(k)$  and the mixing matrix as  $\hat{A} = Q^+ U$ .

#### 4.3 Sons

A very flexible and efficient algorithm developed by Choi and Cichocki referred to as Second Order Nonstationary Source separation (SONS). The method jointly exploits the nonstationarity and the temporal structure of the sources under assumption that additive noise is white or the undesirable interference and noise are stationary signals [10].

The main idea of the SONS algorithm is to exploit the nonstationarity of signals by partitioning the prewhitened sensor data into non-overlapping blocks for which we estimate time-delayed covariance matrices. The SONS algorithm can be obtained as follows

1. The robust whitening method is applied to obtain the whitened vector  $\bar{x}(k) = Qx(k)$  as in AMUSE algorithm.
2. Divide the whitened data  $\bar{x}(k)$  into  $K$  non-overlapping blocks and calculate  $\{M_{\bar{x}}(k_r, \tau_j)\}$  for  $r = 1 \dots K$  and  $j = 1 \dots J$ . In other words, at each time-windowed data frame, we compute  $J$  different time-delayed correlation matrices of  $\bar{x}(k)$ .
3. Find a unitary joint diagonalizer  $V$  of  $\{M_{\bar{x}}(k_r, \tau_j)\}$  using the joint approximate diagonalization method which satisfies  $V^T M_{\bar{x}}(k_r, \tau_j) V = \Lambda_{r,j}$  where  $\{\Lambda_{r,j}\}$  is a set of diagonal matrices.
4. The demixing matrix is computed as  $W = V^T Q$ .

#### 4.4 Jade

The Jade algorithm (Cardoso and Saouloumiac) is an ICA method for identification of the demixing matrix. Jade method is based on a two stage process: (1) diagonalization of the covariance matrix, this is the same as PCA and (2) diagonalization of the kurtosis matrices of the observation vector sequence [11]. The Jade method is composed of the following stages :

1. Initial PCA stage. At this stage first the covariance matrix of the signal X is formed. Eigen analysis of the covariance matrix yields a whitening matrix  $W = \Lambda^{-0.5}U^T$ , where the matrices U and  $\Lambda$  are composed of the Eigen vectors and Eigen values of the covariance of X. The signal is transformed through the matrix W as  $Y = WX$ . The covariance matrix of Y is  $E[YY^T] = E[WXX^TW^T] = I$ .
2. Calculation of kurtosis matrices. At this stage the fourth order (kurtosis) cumulant matrices  $Q_i$  of the signal are formed.
3. Diagonalization of kurtosis matrices. At this stage a single transformation matrix V is obtained such that all the cumulant matrices are as diagonal as possible. This is achieved by finding a matrix V that minimizes the off-diagonal elements.
4. Apply ICA for signal separation. At this stage a separating matrix is formed as  $WV^T$  and applied to the original signal.

### 5. RESULTS & DISCUSSIONS

Linear parameters are calculated for all frontal channels of all subjects. The Fig. 1 shows FP2 channel of 1 subject. Linear parameters are calculated for FP2 channel and listed in Table1. We can see that variance and skewness exhibit higher value in an eyeblink interval during 3-4 & 5-6 time slots. Similarly we calculated linear parameters for remaining channels. High variance is present only in frontal channels when compared to all other channels. Hence, we can take frontal channels for further analysis. In order to validate this result, we go for spectrogram.

Fig 2 shows the Spectrogram of the FP2 channel. Spectrogram possesses higher intensity values in 4<sup>th</sup> and 6<sup>th</sup> interval which corresponds to 3-4 & 5-6 time slots. Similarly spectrogram is performed for all other frontal channels and identifies the EB. Hence, Spectrogram and linear parameters identifies the eyeblink artifacts perfectly

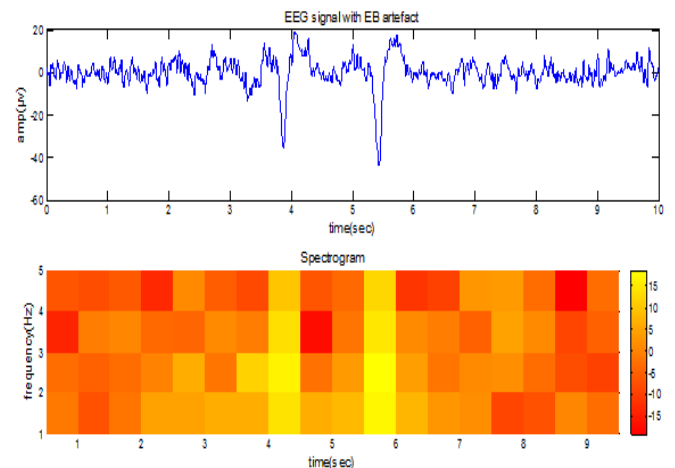
Correlation coefficient is a measure that determines the degree to which two variable's movements are associated. Correlation coefficient should be calculated between an original signal and a reconstructed signal to validate our algorithm. Using the above four algorithms, reconstructed signals has been obtained for three subjects and correlation coefficient is calculated. Correlation coefficient between original and reconstructed

signals of four algorithms is calculated and tabulated for three subjects.

Fig. 11 shows the frontal channels of eyeblink EEG of subject 1. The correlation coefficients of the four different algorithms are listed in Table 2. From the table, we see that JADE algorithm exhibit high correlation coefficient among all. Fig. 12 shows the frontal channels of subject 2. The correlation coefficients of the four different algorithms are listed in Table 3. From the table, we see that SONS algorithm exhibit high correlation coefficient among all. Fig. 13 shows the frontal channels of subject 3. The correlation coefficients of the four algorithms are listed in Table 3. From the table we see that SONS, JADE and AMUSE algorithm exhibit high correlation coefficient randomly in all channels since this signal has more number of eyeblinks.

**Table 1:** Linear parameters of channel FP2-A2

| Time(sec) | 1     | 2      | 3      | 4       | 5      | 6       | 7      | 8      | 9     | 10     |
|-----------|-------|--------|--------|---------|--------|---------|--------|--------|-------|--------|
| Mean      | -0.58 | -0.296 | 1.621  | -2.843  | 3.589  | 0.019   | -0.433 | -0.601 | -1.23 | 1.480  |
| Variance  | 8.267 | 13.825 | 23.499 | 108.555 | 40.540 | 203.689 | 13.375 | 13.683 | 9.513 | 15.521 |
| Kurtosis  | -0.13 | -0.048 | -0.249 | -1.299  | 0.732  | -1.570  | -0.190 | -0.059 | 0.512 | 0.176  |
| Skewness  | 2.499 | 2.572  | 2.695  | 4.786   | 2.620  | 5.009   | 2.942  | 2.705  | 3.115 | 2.868  |



**Fig-2:** Spectrogram of Channel FP2-A2

Using AMUSE, SOBI, JADE, SONS algorithms, independent components are calculated and shown in Fig. 3-Fig. 6. From the independent components, EB components are identified manually, and it is removed by masking it. Then the signal is reconstructed with the help of above discussed algorithms and it is shown in Fig. 7-Fig. 10.

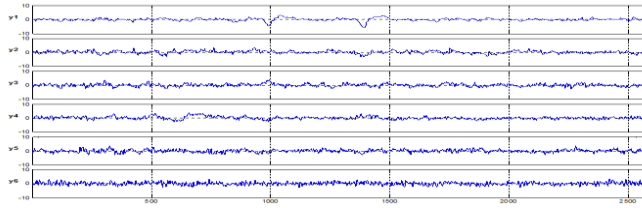


Fig-3: Independent components using AMUSE

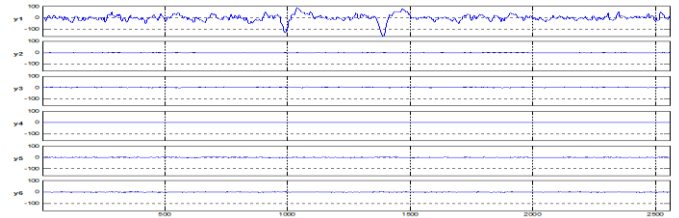


Fig-8: Reconstructed signals using SOBI

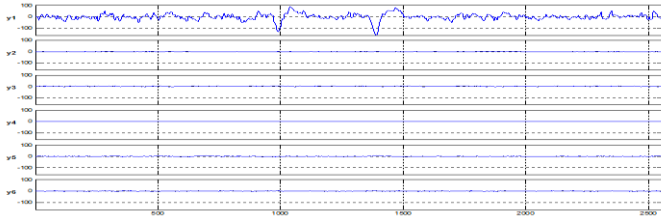


Fig-4: Independent components using SOBI

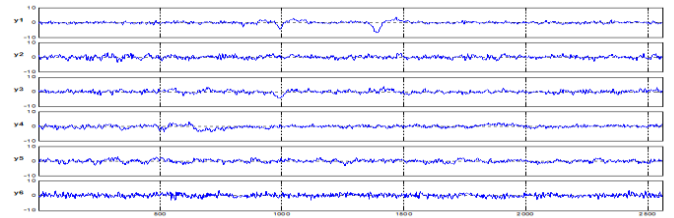


Fig-9: Reconstructed signals using SONS

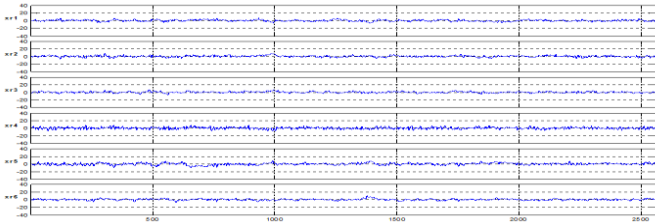


Fig-5: Independent components using SONS

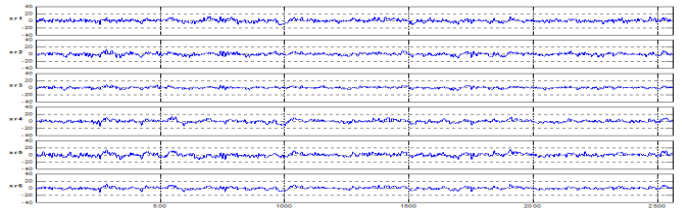


Fig-10: Reconstructed signals using JADE

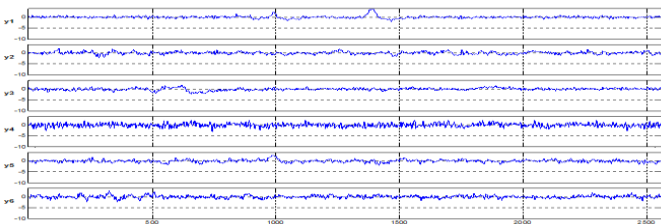


Fig-6: Independent components using JADE

Subject 1:

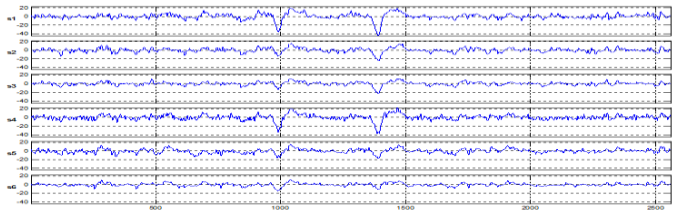


Fig-11: Frontal channels of Subject 1

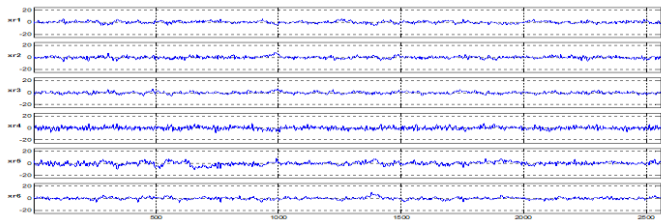
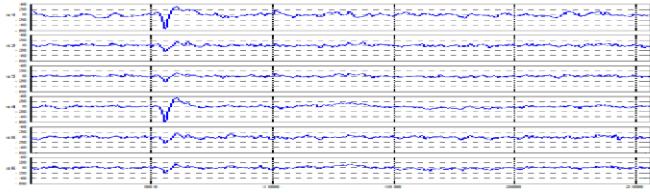


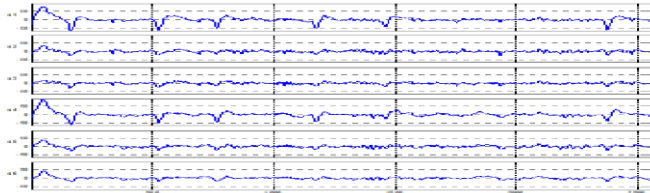
Fig-7: Reconstructed signals using AMUSE

Table-2: Correlation coefficient of Subject 1

|       | FP2 - A2  | F4 - A2  | F8 - A2  | FP1 - A1  | F3 - A1  | F7 - A1  |
|-------|-----------|----------|----------|-----------|----------|----------|
| AMUSE | 0.270765  | 0.409475 | 0.438802 | 0.376182  | 0.601497 | 0.556342 |
| SOBI  | 0.1840056 | 0.408887 | 0.43346  | 0.3747713 | 0.315364 | 0.512532 |
| JADE  | 0.4527972 | 0.68589  | 0.605962 | 0.566375  | 0.86541  | 0.908589 |
| SONS  | 0.4163418 | 0.648592 | 0.561503 | 0.5499127 | 0.847232 | 0.894228 |

**Subject 2:****Fig.12:** Frontal channels of Subject 2**Table-3:** Correlation coefficient of subject 2

|       | FP2 - A2  | F4 - A2  | F8 - A2  | FP1 - A1  | F3 - A1  | F7 - A1  |
|-------|-----------|----------|----------|-----------|----------|----------|
| AMUSE | 0.6504477 | 0.837845 | 0.684436 | 0.241863  | 0.854798 | 0.54401  |
| SOBI  | 0.2562086 | 0.876288 | 0.744538 | 0.2395411 | 0.894201 | 0.630297 |
| JADE  | 0.5982288 | 0.809112 | 0.752787 | 0.4083513 | 0.84866  | 0.862293 |
| SONS  | 0.6694539 | 0.896104 | 0.792883 | 0.4847259 | 0.859199 | 0.903347 |

**Subject 3:****Fig.13:** Frontal channels of Subject 3**Table-4:** Correlation coefficient of subject 3

|       | FP2 - A2  | F4 - A2  | F8 - A2  | FP1 - A1  | F3 - A1  | F7 - A1  |
|-------|-----------|----------|----------|-----------|----------|----------|
| AMUSE | 0.1566871 | 0.707029 | 0.657817 | 0.4031842 | 0.727544 | 0.685222 |
| SOBI  | 0.2657366 | 0.75834  | 0.749294 | 0.3086295 | 0.69493  | 0.622188 |
| JADE  | 0.4566634 | 0.60931  | 0.799215 | 0.4595917 | 0.690847 | 0.461916 |
| SONS  | 0.5330759 | 0.779825 | 0.897478 | 0.4116516 | 0.546588 | 0.275174 |

**6. CONCLUSIONS**

EB artefact has been identified using linear parameters. Out of four parameters, Variance identifies the artefact most prominently. Spectrogram has been obtained in validating our result for the artefact detection. EB artefact removal is done by four algorithms namely AMUSE, SOBI, JADE and SONS. It is

inferred that, if the number of EB artefacts present in an EEG signal is less, both the JADE and SONS algorithm provides higher correlation coefficient, whereas it is not true for more number of EBs present in EEG. In these cases it is difficult to extract original signal after EB removal. Hence, there is a tradeoff between both correlation coefficient and EB removal. This work can be enhanced with hybrid approach of several algorithms to remove EB artefact effectively.

**REFERENCES**

- [1] Saeid Sanei and J.A. Chambers "EEG Signal Processing" Wiley,2007.
- [2] Danny Oude Bos "Automated EEG Artefact Detection and Removal" Internship Radboud University, Nijmegen, 2007.
- [3] Abdul-Bary Raouf Suleiman, Toka Abdul-Hameed Fatehi "Features Extraction Techniques of EEG Signal for BCI Applications" University of Mosul, Iraq.
- [4] Baikun Wan, Dong Ming, Hongzhi Qi, Zhaojun Xue, Yong Yin, Zhongxing Zhou, nd Longlong Cheng " Linear and Nonlinear Quantitative EEG Analysis" IEEE Engineering In Medicine And Biology Magazine, pp 58-63, September 2008.
- [5] Manousos A. Klados, Christos L. Papadelis, Panagiotis D. Bamidis, "A New Hybrid Method for EOG Artifact Rejection" Proceedings of the 9th International Conference on Information Technology and Applications in Biomedicine, 2009.
- [6] Neelam mehala, ratna dahiya "A Comparative Study of FFT, STFT and Wavelet Techniques for Induction Machine Fault Diagnostic Analysis" Proc. of the 7th WSEAS Int. conf. on computational intelligence, man-machine systems and cybernetics (CIMMACS '08), pp 203-208,2008.
- [7] Aapo Hyvärinen, J. Karhunen, *Independent Component Analysis*, 2001 John Wiley & Sons.
- [8] Zhe Chen, Simon Haykin, Jos J. Eggermont, Suzanna Becker "Correlative Learning: A Basis for Brain and Adaptive Systems"
- [9] Seungjin Choi, Andrzej Cichocki, Hyung-Min Park and Soo-Young Lee, "Blind Source Separation and Independent Component Analysis: A Review" Neural Information Processing - Letters and Reviews Vol.6, No.1, pp 8-9, January 2005.
- [10] Lawrence J. Hirsch and Richard P. Brenner "EEG basics" Atlas of EEG in Critical Care, pp1-12, 2010.
- [11] Professor Saeed V. Vaseghi "Advanced digital signal processing and noise reduction" Wiley, 2008, pp 487-88.