Artifact correction of EEG signals using Kernel FastICA based on Mutual Information

J. Sheshagiri Babu , *Member, IEEE* and KVSVR Prasad, *Fellow, IETE*

Abstract— The signals collected from scalp are often called as electroencephalography (EEG) signals .These are usually contaminated by other signals like electrooculography (EOG) signals which are often called as artifacts. The recorded brain signals are mixtures of both EEG and EOG signals. Since the required EEG signal is having an extra unwanted EOG information, this leads to wrong diagnosis of brain activity. The main objective here is removal of this unwanted EOG signal. There are several signal processing methods to solve this problem. But in this paper the authors considered methods of independent component Analysis (ICA) . This problem is considered as a cocktail party problem and hence blind source separation method like ICA is best suitable for the EEG separation. In this paper, ICA algorithms like FastICA, Joint Approximate Diagonalization of Eigen matrices (JADE) Analysis and KernelFastICA are used based on mutual information. These two algorithms are simulated with 2 EEG datasets using MatLab. The results conclude that KernelFastICA gives better results compared to FastICA and JADE Analysis.

***Keywords*—**Blind source separation, EEG, EOG, Entropy, SNR, PDF, Mutual information, Independent component analysis

I. INTRODUCTION

The Electroencephalography (EEG) signal is an electrophysiological monitoring method to record electrical activity of the brain. It is typically noninvasive, with the electrodes placed along the scalp, although invasive electrodes are sometimes used in specific applications. EEG measures [1] voltage fluctuations resulting from ionic current within the neurons of the brain. In clinical contexts, EEG refers to the recording of the brain's spontaneous electrical activity over a period of time, as recorded from multiple electrodes placed on the scalp. Diagnostic applications generally focus on the spectral content of EEG, that is, the type of neural oscillations (popularly called "brain waves") that can be observed in EEG [2] signals. Most of the cerebral signal observed in the scalp EEG falls in the range of 1–20 Hz. Waveforms are subdivided into bandwidths known as alpha, beta, theta, and delta to signify the majority of the EEG used in clinical practice. The alpha waves have the frequency spectrum of 8-13 Hz and can be measured from the occipital region in an awake person when the eyes are closed. The frequency band of the beta waves is 13-30 Hz; these are detectable over the parietal and frontal lobes. The delta waves have the frequency range of 0.5-4 Hz and are detectable in infants and sleeping adults. The theta waves have the frequency range of 4-8 Hz and are obtained from children and sleeping adults. EEG signals have small amplitudes and strong randomness so they can be very easily contaminated with various artifacts. artifacts. One of the most common artifacts influencing the quality of EEG signals are electrooculography (EOG) activities whose magnitude is usually much higher than that of EEG signals.EOG has a burst of high energy in low frequency, which seriously affects the EEG waves, like theta waves and delta waves.

In order to reduce the interference of EOAs, subjects are asked not to blink for a long time or to blink as infrequently as possible, which causes eyes discomfort able. Especially for some specific patients, such as children with attention deficit hyperactivity disorder (ADHD), it is difficult to

ISSN 2229-5518

obey it. Hence, many EOAs often appear in EEG signals[3]-[4]. The common clinical practice is to directly reject EEG segments with eye artifacts. However, it may lead to some loss of important EEG information. Therefore, it is very essential to effectively remove EOAs from EEG signals and preserve underlying brain activity signals with little distortion in the preprocessing of EEG signals.

This paper is organized as section II with Independent component analysis(ICA) algorithm and its methods, Section III describes the reconstruction parameters, Section IV illustrates the results with existing ICA algorithms with Kernel FastICA with 2 different EEG datasets provided in Physionet.org, Section V gives the conclusion and future scope, and Section VI is the references.

II. Independent component analysis (ICA) methods

There exist many algorithms for ICA[5]-[6].They are based on maximizing contrast functions or minimizing batch computations with higher order cumulants. Few of them are FastICA, Joint Approximate Diagonalization of Eigen matrices (JADE) Analysis and KernelFastICA.

A. FastICA:

This algorithm is a fixed point algorithm which is based on the iteration for maximum Non-Gaussianity [7]. It optimizes negative entropy (Negentropy) or kurtosis to measure non gaussianity. Negentropy of Y is given by

$$J(y) = H(y_{gauss}) - H(y) \tag{1}$$

The estimation of Negentropy is quite complex and hence some approximations have to be used. The descent method of approximating negentropy is based on higher order moments is given by

$$J(y) \approx 1/12E\{y^3\} + 1/48kurt(y)^2 \tag{2}$$

Here the random variable y is assumed to be of zero mean and unit variance. However, the validity of such approximations may be rather limited. In particular these approximations suffer from the nonrobustness encountered with kurtosis.

To avoid this new approximations were developed by Aapo Hyvarinen [4] based on maximum entropy principles. The approximation in general is given as

$$J(y) \approx \sum_{i=1}^{p} k_i [E\{G_i(y)\} - E\{G_i(v)\}]^2 \tag{3}$$

Where k_i are some positive constants, y and v are Gaussian variables of zero mean and unit variance with G as a non quadratic function.

In cases where we use only one quadratic function G, the approximation becomes

$$J(y) \propto [E\{G(y)\} - E\{G(v)\}]^2 \tag{4}$$

Taking $G(y)=y^4$,one then obtains exactly (2). i.e kurtosis approximation. Choosing a proper G gives a better result than in (2). The following G have proved very useful:

$$G_1(u) = \frac{1}{a_1}\log\cosh a_1 u \,,\, G_2(u) = -\exp(-u^2/2) \tag{5}$$

Where $1 \le a_1 \le 2$ is a suitable constant.

Fast ICA algorithm:

For a unit vector w with projection $w^T x$ maximizes non gaussianity [7]-[8]. Nongaussianity is here measured by the approximation of negentropy $J(w^T x)$. As discussed the variance of $w^T x$ is unity. For whitened data this is equivalent to constraining the norm of w to be unity.Denote by g the derivative of the non quadratic function G used in (4); for example the derivatives of the functions in (5) are:

$$g_1(u) = \tanh(a_1 u) \,,\, g_2(u) = u\exp(-u^2/2) \tag{6}$$

Where $1 \le a1 \le 2$ is a suitable constant, often taken as a1=1. The basic Fast IC analysis algorithm is as below:

Step1. Choose an initial weight(random) vector w.

Step2. Let $w^+ = E\{xg(w^T x)\} - E\{g'(w^T x)\}w$

Step3. *Let* $w = w^+ / \|w^+\|$

Step4. If not converged go back to step2.

Joint approximate diagonolization of eigen matrices(JADE) algorithm:

It is built on cumulant based contrast function[9]-[10]. The fourth order cross cumulants for random variables x_i, x_j, x_k, x_l is defined as

$$cum(x_i,x_j,x_k,x_l) = E(x_i,x_j,x_k,x_l) - E(x_i,x_j)E(x_k,x_l) - E(x_i,x_k)E(x_j,x_l) - E(x_i,x_l)E(x_j,x_k) \tag{7}$$

The auto cumulants for the random variable is defined as

ISSN 2229-5518

$$cum(x_i,x_j,x_k,x_l)=E(x_i,x_i,x_i,x_i)-E(x_i,x_i)E(x_i,x_i)-E(x_i,x_i)E(x_i,x_i)-E(x_i,x_i)E(x_i,x_i)$$
$$=E(x_i^4)-3((E(x_i^2))^2$$
$$kurt(X)=E(X^4)-3((E(X^2))^2 \qquad (8)$$

If X is Gaussian then kurtosis becomes zero.The steps involving JADE kurtosis are as follows:

Step1.Initially estimate a whitening matrix $\hat{W}$ and set $Z=W^{\wedge}X$

Step2. Estimate a maximal set $\{\hat{Q}_i Z\}$ of cumulant matrices

Step3. Find the rotation matrix $\hat{V}$ such that the cumulant matrices are as diagonal as by solving (optimizing) orthogonal contrast

$$\hat{V}=\arg\min\sum_i off\,(V^T\hat{Q}_i ZV)$$

Kernel FastICA algorithm:

It uses the canonical correlation analysis (CCA) in a reproducing Hilbert kernel space(RHKS) [11]-[12]. The outline of Kernel canonical correlation analysis(KCCA) is given as follows

i) Let X1, X2, X3,……, Xm be the data vectors and K(Xi, Xj) be the Kernel.
ii) Data is whitened with the help of whitened matrix P.
iii) The contrast function C(w) is minimized w.r.t. ‘w’
iv) The contrast function C(w) is minimized in the following way

a) The centered gram matrices K_1, K_2, K_3,……., K_m of the estimated sources {y_1, y_2, y_3,……, y_m}, where $y^i=Wx^i$ are computed.

b) The minimal eigen value of the generalized eigen vector equation

$$C(w)=\hat{I}_{\lambda F}(k_1,k_2,.....k_m)=-\frac{1}{2}\log\hat{\lambda}_F^k(k_1,k_2,.....k_m) \qquad (9)$$

is defined as $K_k\alpha=\lambda D_k\alpha$ (10)

(c)Then

$$C(w)=\hat{I}_{\lambda F}(k_1,k_2,.....k_m)=-\frac{1}{2}\log\hat{\lambda}_F^k(k_1,k_2,.....k_m) \qquad (11)$$

The de-mixing matrix W, is then formed W=WP. Then the independent components are estimated as

$$\hat{S}=WX\,. \qquad (12)$$

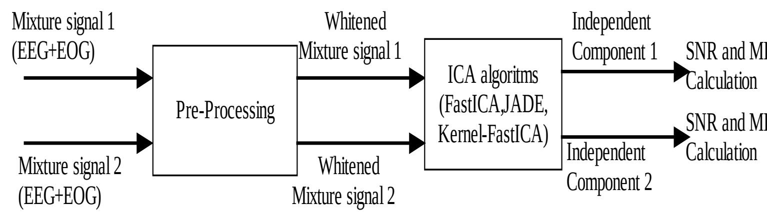

Fig.1.Block diagram for the model

III.Parameters Calculation:

Signal-to-Noise Ratio(SNR): The efficiency of the proposed method can be calculated using the statistical parameter called Signal to Noise Ratio (SNR). High SNR value gives information that the signal content is more compared to noise component. Mathematically, it can be expressed as

$$SNR=10\log\frac{\sum_{k=1}^{m}(X_{denoised}^k)^2}{\sum_{k=1}^{m}(X_{original}^k-X_{denoised}^k)^2} \qquad (13)$$

Where $X_{original}$ and $X_{denoised}$ are taken at K^{th} channel.

Table 1. Comparison of Kernel FastICA with other algorithms with SNR values

EEG data	SNR Values		
	FastICA	**JADE**	**Kernel FastICA**
EEG dataset1	3.13	3.17	3.33
EEG dataset2	3.08	3.13	3.17

Infomax algorithm:

The principle of this algorithm says that the sources are independent and hence they do not have mutual information[13]. The minimization of the mutual information leads to separation of sources. Mutual information is zero if they are statistically independent.

Information measure given by Entropy is defined as

$$I(X) = \int P(x)\log\frac{1}{P(x)}dx \qquad (14)$$

The Joint Entropy is given by

$$I(X,Y) = \int P(x,y)\log\frac{1}{P(x,y)}dxdy \qquad (15)$$

The Mutual Information is given by

$$M(X,Y) = I(X) + I(Y) - I(X,Y) \qquad (16)$$

Minimizing mutual information is nothing but maximizing joint entropy or likelihood or network entropy.

Table 2. Comparison of Kernel FastICA with other algorithms with Mutual Information of Independent components

EEG data	Mutual Information of Original EEG Signal	Mutual Information of Independent components		
		FastICA	**JADE**	**Kernel FastICA**
EEG dataset1	6.81	1.13	0.83	0.53
EEG dataset2	5.76	0.99	0.69	0.39

IV.**RESULTS AND DISCUSSION:**

The EEG data usually contaminated with EOG artifacts which is a major risk in diagnosing human brain activity. In this paper EEG with EOG artifacts is taken and simulated for noise free EEG signal using various ICA algorithms in MATLAB. Here, two data sets of EEG available in website www. physionet.org are analysed and corrected for Pure EEG signal.The Table 1 and Table 2 show that SNR is high and Mutual information is less for Kernel FastICA respectively compared to other two algorithms.

***Case(i)*: For EEG dataset1**

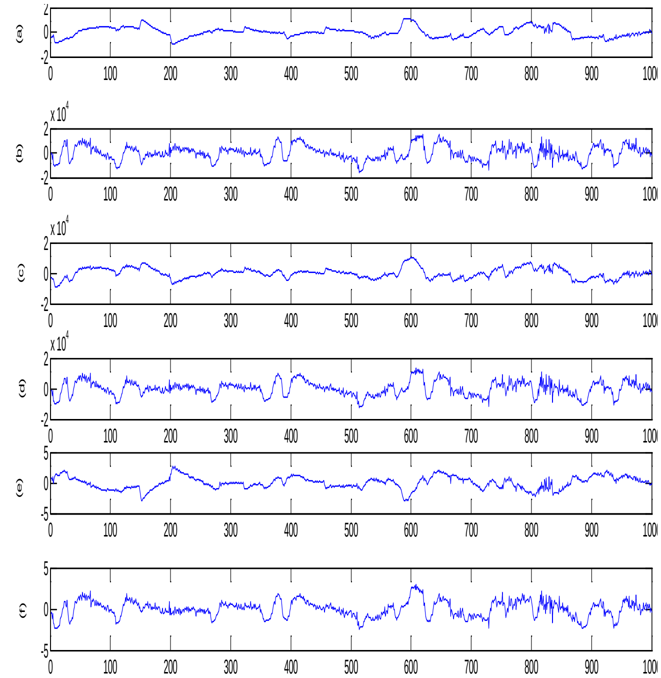

*Fig.1.*FastICA-algorithm
(a) Source signal (EEG) 1 (b) Source signal (EOG) 2 (c) Mixed signal 1 (d) Mixed signal 2 (e) Seperated Independent component 1 (f) Seperated Independent component 2

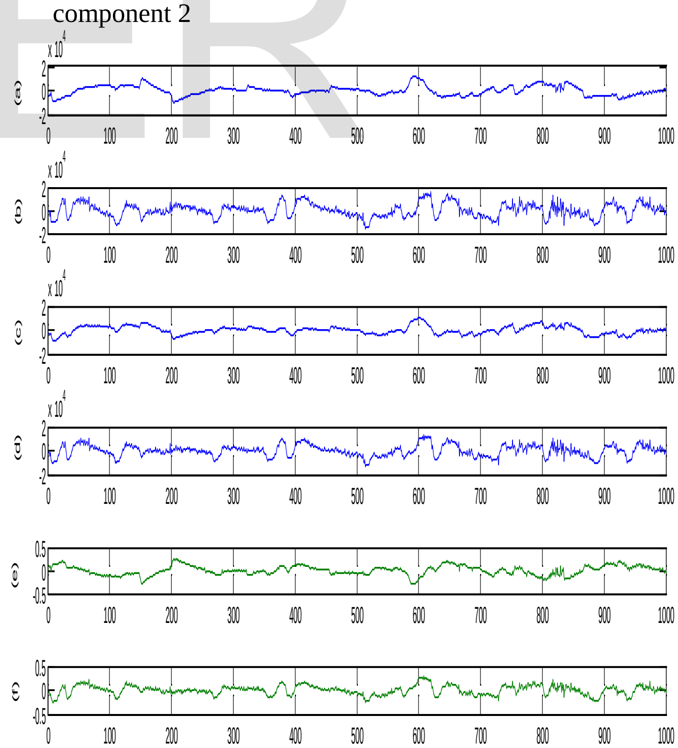

*Fig.2.*JADE-algorithm
(a) Source signal (EEG) 1 (b) Source signal (EOG) 2 (c) Mixed signal 1 (d) Mixed signal 2 (e) Seperated Independent component 1 (f) Seperated Independent component 2 ((e),(f) vary with an voltage of 0.01µV to FastICA)

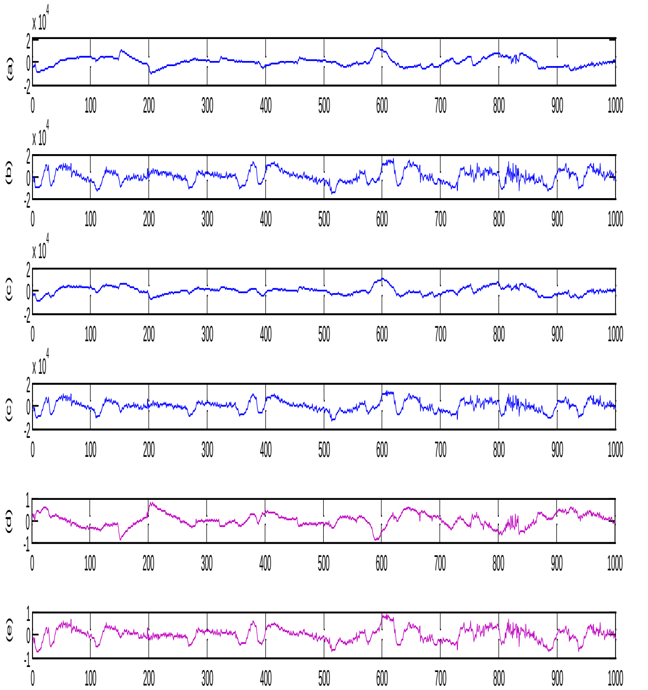

Fig.3.Kernel-FastICA-algorithm

(a) Source signal (EEG) 1 (b) Source signal (EOG) 2 (c) Mixed signal 1 (d) Mixed signal 2 (e) Seperated Independent component 1 (f) Seperated Independent component 2 ((e),(f) vary with an voltage of 0.3μV to JADE)

Case (ii): For EEG data2

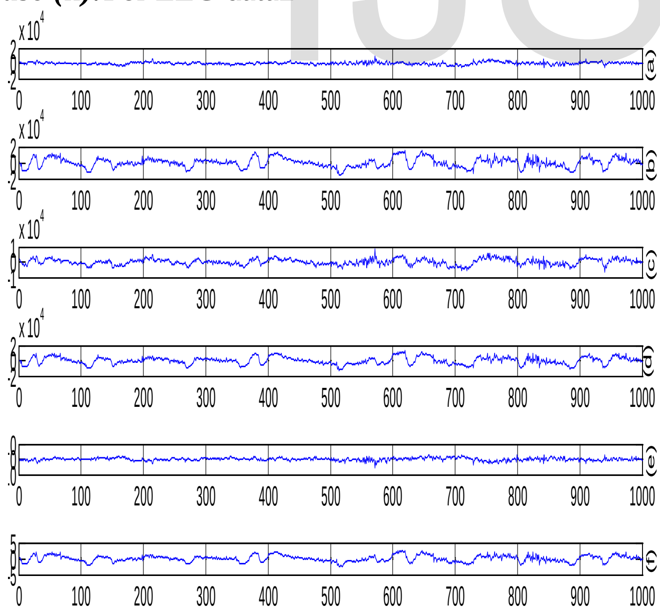

Fig.1.Fast-ICA-algorithm

(a) Source signal (EEG) 1 (b) Source signal (EOG) 2 (c) Mixed signal 1 (d) Mixed signal 2 (e) Seperated Independent component 1 (f) Seperated Independent component 2

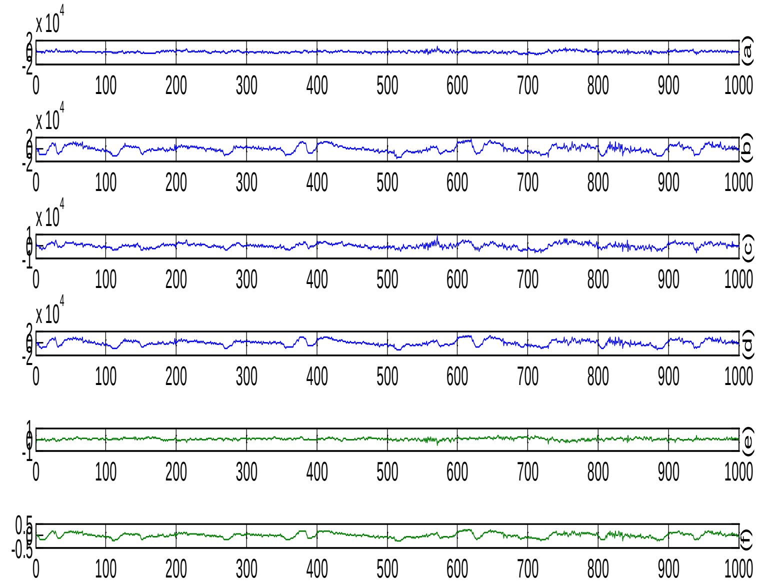

*Fig.2.*JADE-algorithm

(a) Source signal (EEG) 1 (b) Source signal (EOG) 2 (c) Mixed signal 1 (d) Mixed signal 2 (e) Seperated Independent component 1 (f) Seperated Independent component 2 ((e),(f) vary with an voltage of 0.01μV to FastICA)

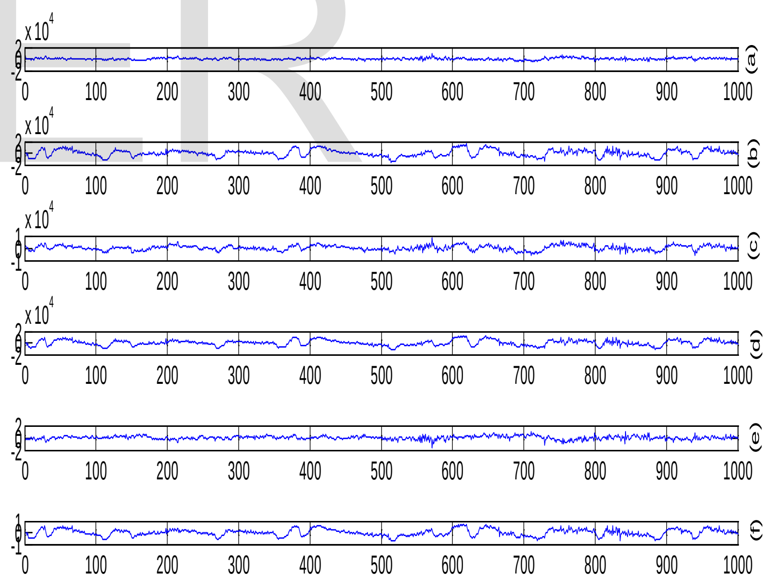

Fig.3.Kernel-FastICA-algorithm

(a) Source signal (EEG) 1 (b) Source signal (EOG) 2 (c) Mixed signal 1 (d) Mixed signal 2 (e) Seperated Independent component 1 (f) Seperated Independent component 2 ((e),(f) vary with an voltage of 0.3μV to JADE)

V. CONCLUSION AND FUTURE SCOPE

ICA is a linear estimation of independent sources from mixtures. It is statistical method of separating the sources from a randomly observed mixed data by transforming it into independent components with interesting distributions.It is also called as BSS or Latent variable model. Minimization of mutual Information of sources leads to ICA model. A computationally efficient methods like FastICA, JADE and Kernel FastICA are proposed and results show that Kernal FastICA gives better results (High SNR values and low mutual information) compared to other two methods.

The proposed method gives better results in correction or removal of EOG artifacts very efficiently in a faster manner but it is quite complex to implement. It can further be improved or modified to get better results with less computational complexity.

VI. REFERENCES:

[1] Lins, O.G., Picton, T.W., Berg, P., Scherg, M., 1993, “Ocular artifacts in recording EEGs and event-related potentials: II. Source dipoles and source components,” Brain Topography 6 (1), 65– 78,

[2] Croft RJ, Barry RJ, “Removal of ocular artifact from the EEG: a review, Clinical Neurophysiology, 30(1), 2000, pp 5-19.

[3] Girton D G, Kamiya J, “A simple on- line technique for removing eye movement artifacts from the EEG,” Electroencephalography and Clinical Neurophysiology, 34, 1973, pp 212-216.

[4] Gratton. G, Coles MG, Donchin E, “A new method for off-line removal of ocular artifact”, Electroencephalography and Clinical Neurophysiology, 55(4), 1983, pp 468-484.

[5] Comon P., “Independent Component Analysis: A new concept?,” Signal Processing 36(3), 1994, pp 287-314.

[6] A. Hyvärinen and E. Oja, “A survey on independent component analysis,” Helsinki University of Technology.

[7] A. Hyvärinen and E. Oja, “A fast fixed-point algorithm for independent component analysis,” Neural Computation, vol. 9, 1997: 1483–1492.

[8] A Hyvarinen., “fast and Robust fixed point algorithm for independent component analysis,” IEEE transactions on Neural Networks, Vol.10, pp.626-634, 1999.

[9] Jean-François Cardoso, “High-order contrasts for independent component analysis”, Neural Computation, vol. 11, no 1, Jan. 1999, pp. 157—192.

[10] Amari , and Cichocki., “A new learning algorithm for blind source separation,” Advances in Neural Information Processing, MIT press, pp.757-763, vol.8, 1996.

[11] F. R. Bach and M. I. Jordan, “Kernel independent component analysis.” J. of Machine Learning Research, 3:1–48, 2002.

[12] B. Schölkopf and A. J. Smola, “Learning with Kernels.” MIT Press, 2001.

[13] A.J.Bell and T.J.Sejnowski, “An information-maximization approach to blind separation and blind deconvolution,” Neural Computation, vol.7, pp.1129-1159, 1995.

J.Sheshagiri babu was born in Ramagundam, India, on 03 July, 1985. He received the B.Tech. degree (2007) in Electronics and Communication Engineering, JNTU, Hyderabad and the M.Tech. degree (2009) in Digital Communications, KU, Warangal. Since 2012, he has been working towards the Ph.D. degree at JNTU, Hyderabad in biomedical signal processing. He is a member of the faculty at the Kakatiya Institute of technology and Science, Warangal. His research interests include biomedical (EEG) signal processing, Independent Component Analysis, Auto Regressive (AR) and Auto Regressive Moving Average (ARMA) Modelling.

He is a *member* of IEEE. He has published and presented over 10 research papers in various IEEE international conferences.

K.V.S.V.R.Prasad was born in Srikakulam, India, on 27 March, 1945. He received the B.Sc. degree (1963) in M.P.C., Andhra University, the B.E. degree (1967) in Telecommunication Engineering, Osmania University, the M.E. degree (1977) in Microwave Engineering, Osmania University and the Ph.D. degree (1985) in Microwave Engineering, Indian Institute of Technology (IIT), Kharagpur. He is currently an emeritus Professor of DMSSVH college of engineering, Machilipatnam. His research interests include Microwave Engineering, Signal processing, and Comunication Engineering.

Dr.KVSVR prasad is a *Fellow* of the Institution of Electronics and Telecommunication Engineers (IETE). He has published over 30 research papers in various IEEE international conferences and Journals.