

Comparison of Independent Component Analysis Algorithms for Removal of Ocular Artifacts from Electroencephalogram

V. Krishnaveni¹, S. Jayaraman¹, P. M. Manoj Kumar¹, K. Shivakumar¹, K. Ramadoss²

¹Department of Electronics & Communication Engineering, PSG College of Technology, Coimbatore – 641 004 India, venimurthy@hotmail.com

²PSG Institute of Medical Sciences and Research, Coimbatore – 641 004 India

Abstract: *The electroencephalogram (EEG) is useful for clinical diagnosis and in biomedical research. EEG recordings are distorted by electrooculogram (EOG) artifacts causing serious problem for EEG interpretation and analysis. An often preferable method is to apply Independent Component Analysis (ICA) to multichannel EEG recordings and remove a wide variety of artifacts from EEG recordings by eliminating the contributions of artifactual sources onto the scalp sensors. The estimated sources should be as independent as possible, for better removal of artifacts from EEG. In this paper, the actual independence of the components obtained from various ICA algorithms like OGWE, MS-ICA, SHIBBS, Kernel-ICA, JADE and RADICAL are assessed and compared by a recently introduced Mutual Information (MI) Estimator based on k-neighbor statistics without using the probability density functions. The results show that RADICAL algorithm performs best at separating the source signals from the observed (mixed) EEG signals and is recommended for.*

Keywords: *Electroencephalogram, electrooculogram,, ocular artifacts, blind source separation, independent component analysis, mutual information*

1. Introduction

The removal of artifacts or noise in biomedical signals like electrocardiogram (ECG) and electroencephalogram (EEG) is a challenging and a crucial task, and their wrong interpretation could prove lethal. EEG is a recording of the electrical activity of the brain from the scalp. It gives a non-invasive insight into the intricacy of the human brain and is a valuable tool for clinicians for numerous applications, from the diagnosis of neurological disorders, to the clinical monitoring of depth of anesthesia. The EEG is susceptible to various artifacts, causing problems for analysis and interpretation. In current data acquisition, eye movement and blink related artifacts are often dominant over other electrophysiological contaminating signals (e.g., heart and muscle activity, head and body movements), as well as external

interference due to power sources. Eye movements and blinks produce a large electrical signal, known as electrooculogram (EOG), which spreads across the scalp and contaminates the EEG. These contaminating potentials are commonly referred to as ocular artifacts (OA).

Numerous methods have been proposed by researchers to remove ocular artifacts in EEG and are reviewed in [1, 2]. A brief write-up about the existing techniques for correction of OAs in EEG is given below. Eye fixation method in which the subject is asked not to blink or move his eyes, or to keep his eyes close, is often unrealistic or inadequate, and is a fact that the subject is concentrating on fulfilling these requirements might itself influence his/her EEG. Another common strategy is to reject all EEG epochs containing artifacts larger than some arbitrarily selected EEG voltage level. When limited data are available, or when blinks and

eye movements occur too frequently, as in children, the rejection of epochs contaminated with OAs usually results in a considerable loss of information and may be impractical for clinical data. Since EEG and EOG occupy the same frequency band, use of analog and digital filters is ineffective. Use of potentiometers to balance out the effect of eye movements, is subjective, since the required adjustments were made manually by observing the EEG [3].

Traditional ocular artifact correction procedures use a regression based approach. Widely used methods for removing OAs are based on regression in time domain [4, 5] or frequency domain [6,7] techniques. Regression analyses are used to compute propagation factors or transmission coefficients in order to define the amplitude relation between one or more electrooculogram (EOG) channels and each EEG channel. Correction involves subtracting the estimated proportion of the EOG from the EEG. One concern often raised about the regression approach is bidirectional contamination. If ocular potentials can contaminate EEG recordings, then brain electrical activity can also contaminate the EOG recordings. Therefore, subtracting a linear combination of the recorded EOG from the EEG may not only remove ocular artifacts but also interesting cerebral activity. In order to reduce the cerebral activity in the EOG, Lins et al. [8] suggested low-pass filtering the EOG signal used to compute regression coefficients. However, they recognized that low-pass filtering removes all high frequency activity from the EOG signal, both of cerebral and ocular origin. A new filtering approach for regression based correction using Bayesian adaptive regression splines [9, 10] uses a locally defined nonlinear filter to remove high frequency activity when the amplitude fluctuations are small and retain high frequency activity when the amplitude fluctuations are large. Such adaptively filtered EOG essentially isolates activity typically associated with ocular artifacts and removes cerebral activity. The use of such

adaptive filtering prior to applying regression correction may substantially reduce problems from bidirectional contamination. Use of adaptive digital filters for OA removal [11], also requires a suitable EOG reference model for training the filter.

Another class of methods is based on decomposing the EEG and EOG signals into spatial components, identifying artifactual components and reconstructing the EEG without the artifactual components. For example, Lins et al. [8] and Lagerlund et al. [12] used Principal Component Analysis (PCA) to identify the artifactual components. In addition, the dipole modeling technique of Berg and Scherg [13, 14] used PCA to compute topographies of eye activity. Statistically, PCA decomposes the signals into uncorrelated, but not necessarily independent components that are spatially orthogonal and thus it cannot deal with higher-order statistical dependencies. However PCA cannot completely separate eye artifacts from brain signals especially when they both have comparable amplitudes [12].

A newer approach uses independent component analysis (ICA), which was developed in the context of blind source separation problems to form components that are as independent as possible [15,16]. Scott Makeig et.al [17] reported the first application of ICA for EEG data analysis by using the algorithm of Bell and Sejnowski [18] for ICA. It is based on a new unsupervised neural network learning algorithm. They showed that ICA can separate neural activity from muscle and blink artifacts in spontaneous EEG data.

Jung et.al, [19] proposed a new and generally applicable method for removing a wide variety of artifacts from EEG records. This method is based on an extended version of infomax algorithm [18] and can be used for performing blind source separation on linear mixtures of independent source signals with either sub-Gaussian or super-Gaussian distributions. They showed that ICA can effectively detect, separate and remove activity in EEG records from a wide variety

of artifactual sources. Vigario [20] used FastICA algorithm [21] for identification of artifacts in EEG and MEG. They showed that the FastICA algorithm can be used for extracting different types of artifacts from EEG and MEG data, even when these artifacts are smaller than the background brain activity. Compared to PCA, ICA removes the constraint of orthogonality and forces components to be approximately independent rather than simply uncorrelated. However, the ICA components lack the important variance maximization property possessed by PCA components. In addition, the ICA algorithms discussed above [17,19, and 20] requires the user to manually select the artifacts from the estimated components for correction, thus creating challenges for implementing automated correction routines.

An ICA based method for removing artifacts semi automatically was presented by Delorme et.al [22]. Although it is automated to flag (mark) trials as potentially contaminated, these trials are still examined and rejected manually via a graphical interface. Belouchrani et.al [23], proposed an alternative approach for signal source separation, the SOBI (Second Order Blind Identification) algorithm which uses decorrelation across several time points as its computational step. Carrie Joyce et.al [24] used SOBI algorithm along with correlation metrics and Nicolaou et.al [25] used TDSEP [26] along with Support Vector Machine (SVM) for automatic removal of artifacts. The results of these studies does not imply that SOBI/TDSEP is the overall best approach for decomposing EEG sensor data into meaningful components, and has not been completely validated by the authors.

The limitation of ICA algorithms is that there is no guarantee that any particular algorithm can capture the individual source signals in its components [24]. The estimated source signals (obtained from ICA algorithm) should be as independent as possible (or least dependent on each other) for better removal of artifacts from EEG,

since, either by visual inspection, or by automated procedure, only the estimated sources are classified as EEG or artifacts, but, the actual independence of the components (estimated sources) obtained from ICA algorithms used in [17, 19, 20, 22, 24, 26] are not tested and quantified. As discussed above, a number of ICA/BSS based EEG/EOG analyses have been published till date, but not with sufficient background to enable the EEG practitioner to choose the best algorithm [24].

This paper compares different ICA algorithms MS-ICA [27], OGWE [28], JADE [29], SHIBBS [30], Kernel-ICA [31] and RADICAL [32] for removal of ocular artifacts from EEG. For assessing the actual independence of the components obtained from ICA/BSS, Mutual Information (MI) is used. MI leads to basic performance tests for any ICA problem and hence different ICA/BSS algorithms can be ranked by how well they perform i.e. whether they find the most independent components. However, MI was not extensively used for measuring interdependence, mainly because of the difficulty in estimating it reliably [33]. In this paper, an efficient methodology to estimate Mutual Information (MI) between two or more signals [33] is used and an efficient ICA algorithm for removal of ocular artifacts from EEG is found.

The paper is outlined as follows. In Section II, a theoretical review of various ICA/BSS algorithms like MS-ICA, OGWE, JADE, SHIBBS, Kernel-ICA and RADICAL is described. In Section III, the algorithm for the estimation of mutual information is given. Results are discussed in Section IV and the paper is concluded in Section V.

2. Independent Component Analysis

Independent Component Analysis (ICA) [16] involves the task of computing the matrix projection of a set of components onto another set of so called independent component. Here, the objective is to maximize the statistical independence of the outputs. If the inputs to the ICA are known to

be linear instantaneous mixture of a set of sources, the ICA process provides an estimate of the original sources. Here, and in the context of this paper, neither the original sources nor the mixture matrix are known. This is the Blind Separation of Sources (BSS) [16] where the aim is to obtain a non-observable set of signals, the so-called sources, from another set of observable signals regarded as mixtures. The BSS problem can be easily tackled by exploiting the higher order signal statistics and optimization techniques.

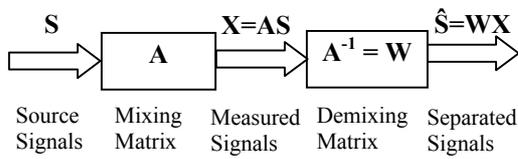


Fig. 1 : Schematic illustration of the mathematical model used to perform ICA decomposition

The original source vector \mathbf{S} is of size $M \times N$ and the mixing matrix \mathbf{A} is of size $M \times M$, where M is the number of statistical independent sources and N is the number of samples in each source.

The result of the separation process is the demixing matrix \mathbf{W} which can be used to obtain the estimated statistical independent sources, $\hat{\mathbf{S}}$ from the mixtures. This process is described by Equation 1 and a schematic illustration of the mathematical model is shown in Figure 1.

$$\mathbf{X} = \mathbf{A}\mathbf{S} \rightarrow \hat{\mathbf{S}} = \mathbf{W}\mathbf{X} \quad (1)$$

Preprocessing for ICA:

Some preprocessing is useful before attempting to estimate \mathbf{W} [34].

(i) The observed signals should be centered by subtracting their mean value $E\{\mathbf{x}\}$

$$\tilde{\mathbf{X}} = \mathbf{X} - E[\mathbf{X}] \quad (2)$$

(ii) Then they are whitened, which means they are linearly transformed so that the components are uncorrelated and has unit variance.

(iii) Whitening can be performed via eigenvalue decomposition of the covariance matrix, $\mathbf{V}\mathbf{A}\mathbf{V}^T$, \mathbf{V} is the matrix of orthogonal eigenvectors and $\mathbf{\Lambda}$ is a diagonal matrix with the corresponding eigenvalues. The whitening is done by multiplication with the transformation matrix \mathbf{P}

$$\mathbf{P} = \mathbf{V}\mathbf{\Lambda}^{-1/2}\mathbf{V}^T \quad (3)$$

$$\mathbf{Z} = \mathbf{P}\tilde{\mathbf{X}} \quad (4)$$

This is closely related to principal component analysis [35]. The covariance of the whitened data $E[\mathbf{Z}\mathbf{Z}^T]$ equals the identity matrix. For simplicity, let \mathbf{X} be the centered mixed vector $\tilde{\mathbf{X}}$, i.e. $\mathbf{X} = \tilde{\mathbf{X}}$

Theoretical review of the ICA Algorithms

The algorithms studied in this paper are OGWE, MS-ICA, SHIBBS, Kernel-ICA JADE and RADICAL. A brief description of various algorithms is given below.

OGWE Algorithm:

In OGWE (Optimized Generalized Weighted Estimator) [28], the marginal entropy contrast function (Φ^{ME}) is written in terms of second-order and fourth-order cumulants, and then it is minimized for all possible distributions for the sources \mathbf{S} [30], it follows that

$$\phi^{ME}(\mathbf{Y}) \approx \frac{1}{48} \phi_{24}^{ME}(\mathbf{Y}) = -\frac{1}{48} \sum_i (C_{iiii}^Y)^2 \quad (5)$$

where, for zero-mean signals, $C_{iii}^Y = E[Y_i^4] - 3E[Y_i^2]^2$ are the marginal cumulants or autocumulants [9].

In the two dimensional case, the pair of normalized sources $s_i = [\bar{s}_p(t) \bar{s}_q(t)]^T$ in polar coordinates may be written as $(r(t), \alpha(t))$ so that the outputs yield

$$\begin{pmatrix} Y_p(t) \\ Y_q(t) \end{pmatrix} = \mathbf{R}(\theta) \begin{pmatrix} r(t) \cos(\beta(t)) \\ r(t) \sin(\beta(t)) \end{pmatrix} = \mathbf{R}(\theta) Z_t \quad (6)$$

where $Z_t = [Z_p(t) Z_q(t)]^T$ are the whitened mixtures, and matrix \mathbf{V} performs a rotation of θ so that $\rho(t) = \theta + \beta(t)$ is the angle of vector

y. Note that ideally, at separation $\theta + \beta(t) = \alpha(t)$.

(i) The whitening matrix \mathbf{P} is computed to whiten the vector \mathbf{X} and the vector $\mathbf{Y} = \mathbf{P}\mathbf{X}$ is formed.

(ii) One Sweep. For all $g = m(m-1)/2$ pairs, i.e., for $1 \leq p < q \leq m$, the following steps have to be done:

(a) The Given angle $\theta_{pq} = \theta_{GWE}$ is computed,

(with $[z_p \ z_q]^T = [y_p \ y_q]^T$) as follows:

$$\hat{\theta}_{GWE}(\omega_r, \omega_\xi) = \frac{1}{4} \angle(\omega_\xi \omega_r \xi_r + (1 - \omega_\xi) \xi_\eta) \quad (7)$$

$$0 < \omega_\xi < 1, \omega_r \pm 1, \gamma$$

$$\hat{\theta}_{SICA} = \hat{\theta}_{GWE}(\gamma, \frac{3}{7}) \quad (8)$$

where $\angle(\cdot)$ supplies the principal value of the argument.

$$\begin{aligned} \xi_r &= E[r^4(t) e^{j4\beta(t)}] \\ \xi_\eta &= E^2[r^4(t) e^{j2\beta(t)}] \\ \gamma &= E[r^4(t)] - 8 \end{aligned} \quad (9)$$

(b) If $\theta_{pq} > \theta_{\min}$, the pair (Z_p, Z_q) is rotated by θ_{pq}

according to Eq.(5) and also the rotation matrix \mathbf{R} is updated. The value of θ_{\min} is selected in such a way that rotations by a smaller angle are not statistically significant. Typically $\theta_{\min} = 10^{-2}/\sqrt{N}$ where N is the number of samples.

(iii) End if the number of iterations n_{it} satisfies $n_{it} \geq 1 + \sqrt{M}$ or no angle θ_{pq} has been updated, stop. Otherwise go to step(ii) for another sweep.

(iv) Then the demixing matrix $\mathbf{W} = \mathbf{R}\mathbf{P}$ and the independent sources are estimated as $\hat{\mathbf{S}} = \mathbf{W}\mathbf{X}$

ICA-MS Algorithm

Molgedey and Schuster [27] proposed an approach based on dynamic decorrelation which can be used if the independent source signals have different autocorrelation functions. The main advantage of this

approach is that the solution is simple and constructive, and can be implemented in a fashion that requires the minimal user intervention (parameter tuning).

Let \mathbf{X}_τ be the time shifted version of the mixed vector \mathbf{X} . The delayed correlation approach is based on solving the simultaneous eigenvalue problem for the correlation matrices $\mathbf{X}_\tau \mathbf{X}^T$ and $\mathbf{X}\mathbf{X}^T$ [28]. This is implemented by solving the eigenvalue problem for the quotient matrix $\mathbf{Q} \equiv \mathbf{X}_\tau \mathbf{X}^T (\mathbf{X}\mathbf{X}^T)^{-1}$. From (1), $\mathbf{X}\mathbf{X}^T = \mathbf{A}\mathbf{S}\mathbf{S}^T \mathbf{A}^T$ and $\mathbf{X}_\tau \mathbf{X}^T = \mathbf{A}\mathbf{S}_\tau \mathbf{S}^T \mathbf{A}^T$ are obtained.

If the sources furthermore are independent, the diagonal source cross-correlation matrix is obtained at lag zero in the limit $\lim_{N \rightarrow \infty} N^{-1} \mathbf{S}\mathbf{S}^T = \mathbf{C}(\mathbf{0})$. Similarly, $\lim_{N \rightarrow \infty} N^{-1} \mathbf{S}_\tau \mathbf{S}^T = \mathbf{C}(\tau)$ produces the diagonal crosscorrelation matrix at lag τ . Hence, to zeroth order in $1/N$,

$$\mathbf{X}_\tau \mathbf{X}^T (\mathbf{X}\mathbf{X}^T)^{-1} \approx \mathbf{A}\mathbf{C}(\tau)\mathbf{A}^T (\mathbf{A}^T)^{-1} \mathbf{C}(\mathbf{0})^{-1} \mathbf{A}^{-1} \quad (10)$$

with $\mathbf{C}(\tau)\mathbf{C}(\mathbf{0})^{-1}$ being a diagonal matrix. If the eigenvalue problem is solved for the quotient matrix [36].

$\mathbf{Q} \equiv \mathbf{X}_\tau \mathbf{X}^T (\mathbf{X}\mathbf{X}^T)^{-1} \approx \mathbf{A}\mathbf{C}(\tau)\mathbf{C}(\mathbf{0})^{-1} \mathbf{A}^{-1}$ (11) then the direct scheme is obtained to estimate \mathbf{A} , \mathbf{S} . Let

$$\mathbf{Q}\Phi = \Phi\Lambda \quad (12)$$

and $\Phi = \mathbf{A}$ and $\Lambda = \mathbf{C}(\tau)\mathbf{C}(\mathbf{0})^{-1}$ up to scaling factors are identified.

Then the demixing matrix \mathbf{W} is the inverse of the mixing matrix \mathbf{A} . The sources can be estimated as $\hat{\mathbf{S}} = \mathbf{W}\mathbf{X}$.

JADE Algorithm

Yet another signal source separation technique is the Joint Approximation Diagonalisation of Eigen matrices (JADE) algorithm [29]. This exploits the fourth order moments in order to separate the source signals from mixed signals. The operation of JADE is as given below:

(i) The Whitening matrix \mathbf{P} and the set $\mathbf{Z} = \mathbf{P}\mathbf{X}$ are estimated.

- (ii) The fourth cumulants of the whitened mixtures \hat{Q}_i^Z are computed. Their m most significant eigen values λ_i and their corresponding eigen matrices V_i are determined. An estimate of the unitary matrix R is obtained by maximizing the criteria $\lambda_i V_i$ by means of joint diagonalisation. If $\lambda_i V_i$ cannot be exactly jointly diagonalised, the maximization of the criteria defines a joint approximate diagonalisation.
- (iii) An orthogonal contrast is optimized by finding the rotation matrix R such that the cumulant matrices are as diagonal as possible, that is, the equation

$$R = \arg \min_R \sum_i \text{Off}(R^T \hat{Q}_i^Z R) \quad (13)$$

- (iv) The matrix A is estimated as $\hat{A} = RP^{-1}$ and the components are estimated as $\hat{S} = \hat{A}^{-1}X$.

SHIBBS Algorithm:

Another signal separation technique is Shifted Block Blind Separation (SHIBBS) [29] to estimate the demixing matrix W .

- (i) A fixed set $X = \{X_1, \dots, X_m\}$ of $m \times n$ matrices is selected. A Whitening matrix P and set $Z = PX$ are estimated.
- (ii) The set $\{\hat{Q}^Z(X_m) | 1 \leq p \leq M\}$ of M cumulant matrices is estimated and a joint diagonalizer R of it is found.
- (iii) If R is close enough to the identity transform, stop. Otherwise, the data is rotated using the equation $Z = R^T Z$ and step (ii) is repeated.
- (iv) Then the demixing matrix $W = RP$ is used to estimate the independent components $\hat{S} = WX$

The SHIBBS algorithm is implemented in the same way as JADE is done. But the joint diagonalization of the significant eigen matrices is done without going through the estimation of the whole cumulant set and through the computation of its eigenmatrices.

Kernel-ICA Algorithm:

The Kernel-ICA algorithm [31] uses the contrast functions based on Canonical Correlation Analysis (CCA) [37] in a Reproducing Hilbert Kernel Space (RKHS) [31,37]. The outline of Kernel Canonical Correlation Analysis (KCCA) is given as follows:

- (i) Let x_1, x_2, \dots, x_m be the data vectors and $K(x_i, x_j)$ be the kernel.
- (ii) Data is whitened with the help of whitening matrix P .
- (iii) The contrast function $C(W)$ is minimized with respect to W .
- (iv) The contrast function is minimized in the following way:

a) The centered Gram matrices [31] K_1, K_2, \dots, K_m of the estimated sources $\{y_1, y_2, \dots, y_m\}$, where $y^i = Wx^i$ are computed.

b) The minimal eigenvalue of the generalized eigenvector equation $\hat{\lambda}_F^k(K_1, \dots, K_m)$ is defined as

$$K_k \alpha = \lambda D_k \alpha \quad (14)$$

c) Then

$$C(W) = \hat{I}_{\lambda_F}(K_1, \dots, K_m) = -\frac{1}{2} \log \hat{\lambda}_F^k(K_1, \dots, K_m)$$

- (v) The demixing matrix W , is then formed $W = WP$. Then the independent components are estimated as $\hat{S} = WX$

RADICAL Algorithm:

The RADICAL (Robust, Accurate, Direct Independent Component Analysis Algorithm) [32] estimates the independent sources using differential entropy estimator based on 'm'-spacing estimator. The contrast function in (15) is to be minimized by RADICAL is almost equivalent to Vasicek estimator [38],

$$\hat{H}_{RADICAL}(Z^1, \dots, Z^N) \equiv \frac{1}{N-m} \sum_{i=1}^{N-m} \log \left(\frac{N+1}{m} (Z^{(i+m)} - Z^{(i)}) \right) \quad (15)$$

The data vectors X_1, X_2, \dots, X_M are assumed to be whitened using the whitening matrix P .

Let m be the size of spacing. The value of 'm' is taken as \sqrt{N} where N is the number of samples in each source.

Let \mathfrak{R} be the number of replicated points per original data point to eliminate the local minima problem [32].

Let σ be the standard deviation of replicated points. For $N < 1000$, $\sigma = 0.35$ and for $N \geq 1000$, $\sigma = 0.175$, where N is the number of samples in each source before replication.

Let K be the number of angles for which cost function has to be evaluated. The optimum value of K here is 350.

- (i) For each of $M-1$ sweeps (or until convergence), where M is the number of sources.
- (ii) For each of $M(M-1)/2$ jacobi rotations for dimensions (p,q) .
 - (a) A pair of whitened mixture is taken (Z_p, Z_q) .
 - (b) Create Z' by replicating \mathfrak{R} points with Gaussian noise for each original point.
 - (c) For each θ in K number of angles, the augmented data are rotated to this angle

$$\mathbf{Y} = \mathbf{R}(\theta) \times \mathbf{Z}' \quad (16)$$
 and the contrast function is evaluated.
 - (d) The Jacobean matrix for the optimal θ is formed and it is incorporated into the rotation matrix \mathbf{R} . The optimal θ is one which yields the minimum Vasicek estimator value [38] for the rotated pair.
- (iii) The final rotational matrix \mathbf{R} is the accumulation of all the jacobi rotations of optimal θ .
- (iv) The demixing matrix $\mathbf{W} = \mathbf{R}\mathbf{P}$ and the estimated sources $\hat{\mathbf{S}} = \mathbf{W}\mathbf{X}$ are obtained.

3. Estimating Mutual Information

If X and Y are two random variables with joint distribution $\mu(x, y)$ and marginal distributions $\mu_x(x)$ and $\mu_y(y)$, then Mutual Information $I(X, Y)$ between X and Y is defined as

$$I(X, Y) = \iint \mu(x, y) \times \log(\mu(x, y) / \mu_x(x)\mu_y(y)) dx dy \quad (17)$$

The algorithm proposed by Kraskov et.al [33] estimates $I(X, Y)$ from the set $\{Z_i\}$ alone without explicit estimation of the unknown densities and is outlined below.

For any set of N bivariate measurements $z_i = (x_i, y_i)$, the k^{th} closest neighbor of each Z_i is found according to the metric

$$\|z - z'\| = \max\{\|x - x'\|, \|y - y'\|\} \quad (18)$$

The k^{th} nearest neighbor is then projected onto the X and Y axes giving the distances $\varepsilon_x(i)/2$ and $\varepsilon_y(i)/2$. The estimate for MI is given by

$$\hat{I}(x, y) = \psi(k) - (1/k) - \langle \psi(n_x) + \psi(n_y) \rangle + \psi(N) \quad (19)$$

Where $n_x(i)$ and $n_y(i)$ be the number of points with $\|x_i - x_j\| \leq \varepsilon_x(i)/2$ and $\|y_i - y_j\| \leq \varepsilon_y(i)/2$ and $\psi(\cdot)$ is the digamma function

$$\psi(x) = \Gamma(x^{-1}) \times (d\Gamma(x) / dx)$$

and $\langle \dots \rangle = N^{-1} \sum_{i=1}^N E[\dots(i)]$.

4. Results and Discussion

ICA is a statistical method for transforming an observed multi-component data set into independent components that are statistically as independent as possible. The components estimated by an ICA algorithm should be least dependent on each other, for better removal of artifacts from EEG and so the actual dependencies between the obtained components is to be estimated and it is most often ignored. Hence it becomes necessary to estimate the actual dependencies between the components and to find the best ICA algorithm that transforms the observed data set into components that are least dependent on each other. There are various measures to evaluate the independence among the estimated sources. Some of the measures are kurtosis, negentropy, Mutual Information, etc [39]. Kurtosis is the fourth-order cumulant. In terms of robustness and asymptotic variance, the cumulant based estimator tend to be far from optimal. Intuitively there are

two main reasons for this. Firstly, higher order cumulants measure mainly the tails of the distributions, and are largely unaffected by structure in the middle of the distribution. Secondly, estimators of the higher order cumulants are highly sensitive to outliers [39]. Their value may depend on only a few observations in the tails of distribution which may be outliers. Negentropy involves estimation of probability density function which is very difficult. Cumulant-based approximations of negentropy are inaccurate and in many cases too sensitive to outliers. Among these measures, Mutual Information (MI) is the best choice to measure the independence of the estimated sources. However, MI was not extensively used for measuring interdependence because estimating MI from statistical samples is not easy. In the ICA literature very crude approximations to MI based on cumulant expansions are popular because of their ease of use. In this paper, an efficient methodology to estimate MI [33] based on k-nearest neighbor distances without estimating the probability densities is used to assess the actual independence of the components obtained from various ICA algorithms like OGWE, ICA-MS, JADE, SHIBBS, Kernel-ICA and RADICAL

In order to find the best ICA algorithm, for removal of ocular artifacts from EEG, 17 sets of recorded EEG data with ocular artifacts are taken for analysis (32 electrodes, 128 Hz, 5 sec). The independent components for various ICA algorithms are obtained and MI for the mixed signals $\hat{i}(X_1, X_2, \dots, X_N)$ and for the components $\hat{i}(\hat{S}_1, \hat{S}_2, \dots, \hat{S}_N)$ are estimated by using Eq. (19) choosing $k = 6$. Practical considerations for selecting k are discussed in [33]. MI is zero, if two random variables are strictly independent. Hence it is expected that $\hat{i}(\hat{S}_1, \hat{S}_2, \dots, \hat{S}_N)$ will decrease when compared to $\hat{i}(X_1, X_2, \dots, X_N)$. The mutual information of the estimated components are computed for all the algorithms and is tabulated in Table 1 for 17 data sets. Lesser the MI, better the algorithm is for separating the original sources from mixtures, given the

assumptions of that algorithm. The results show that the RADICAL algorithm performs best at separating the source signals from the observed EEG signals. Fig 2 shows the mixed EEG signals and Fig 3 shows the independent components obtained using RADICAL algorithm. Ocular artifacts are identified visually from the RADICAL estimated source components shown in Fig 3, and channel 1 corresponds to EOG and are removed from the mixed (observed) signals to get clean EEG signal as shown in Fig 4. After denoising, the background information in each lobe has to be retained. Otherwise, there will be a loss of original EEG data. EOG interference will be dominant in the EEG recorded from the electrodes (F3, Fz, F4, C3, C4, CP2) placed on the patient's forehead. Hence a frontal channel EEG recording (F3) is shown along with the corrected EEG data for various algorithms in Fig. 5. In order to check whether the background EEG information is retained, the contaminated and corrected versions of EEG signals are visually inspected. Since the exact amplitude cannot be determined by ICA, the data have been centered by subtracting their means from them.

From Fig. 2 it is clear that OAs occupy the lower frequency range from 0 Hz to 6-7 Hz for the eye movement artifacts and typically up to the alpha band (8-13 Hz), excluding very low frequencies, for the eye blinks. The power of EOG in the low frequency band is reduced from 44.9204 dB to 25.0952 dB ensuring the exact preservation of the high frequency content of the original EEG signal, while removing the low frequency ocular artifacts. Power spectral density was found by periodogram smoothing by applying the Blackman window to the autocorrelation estimate and then taking Fourier transform. The periodogram averaging was done by segmenting the data to obtain several records followed by windowing spectral leakage and finally by averaging the periodogram to reduce variances.

5. Conclusion and Future Scope

Ocular artifact correction is a challenging task. A variety of techniques have been proposed in the literature for the same. However there is no general consensus amongst researchers upon the selection of the best, appropriate and feasible technique which enables the satisfactory removal of ocular artifacts and preservice of EEG information intact. In this paper, quantitative analysis has been carried out to evaluate six ICA algorithms for removal of ocular artifacts from EEG by using a reliable Mutual Information Estimator, and the results show that RADICAL perform best at separating the original sources from the observed signals. Once the components are as independent as possible, then the components can be classified either as artifacts or EOG, and can be removed from the mixed signals to obtain the artifact free EEG data. In [40], it is shown that JADE outperforms the well-known ICA/BSS algorithms such as Infomax, Extended Infomax, FastICA, SOBI, TDSEP. But in this paper, it is shown that RADICAL has

emerged superior when compared with JADE and other ICA algorithms such as, OGWE, SHIBBS, MS-ICA and Kernel-ICA on the basis of Mutual Information Estimation. In this paper, the inspection of artifact channel is done visually, once the independent components are separated. However as an improvement over the current process, this inspection of artifact channel can be automated by using Wavelets, Kalman Predictor, Neural Networks etc., The advantage of automated correction procedure is that it eliminates the subjectivity associated with non-automated correction procedures and can be used during on-line EEG monitoring for clinical purposes. Power Spectral Density as a performance metric gives only a rough estimate in providing an inference relating to the relative superiority of various ICA algorithms in removing ocular artifacts from EEG. Further it is our considered opinion that the usefulness of the best separation algorithm for removing ocular artifacts from EEG can also be justified quantitatively by proposing a suitable performance metric for validating the de-noised EEG signals.

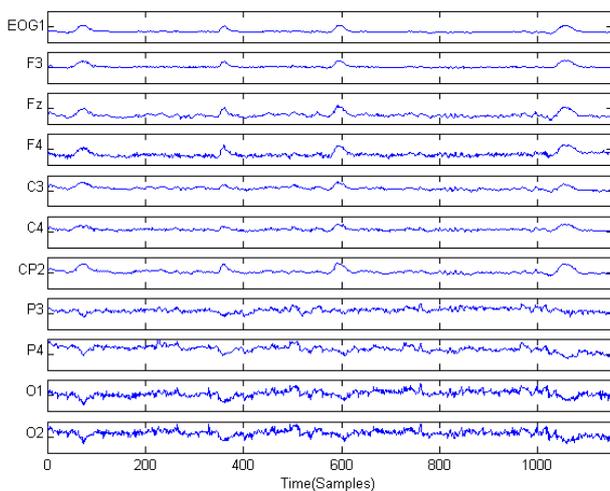


Fig. 2: Recorded (Mixed) EEG signals

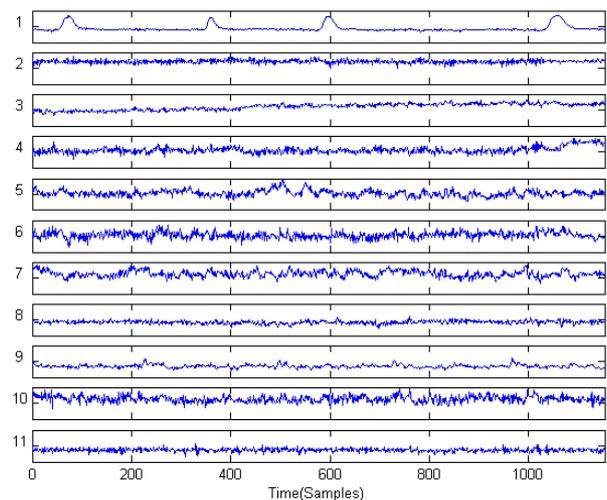


Fig. 3: Independent components obtained from RADICAL

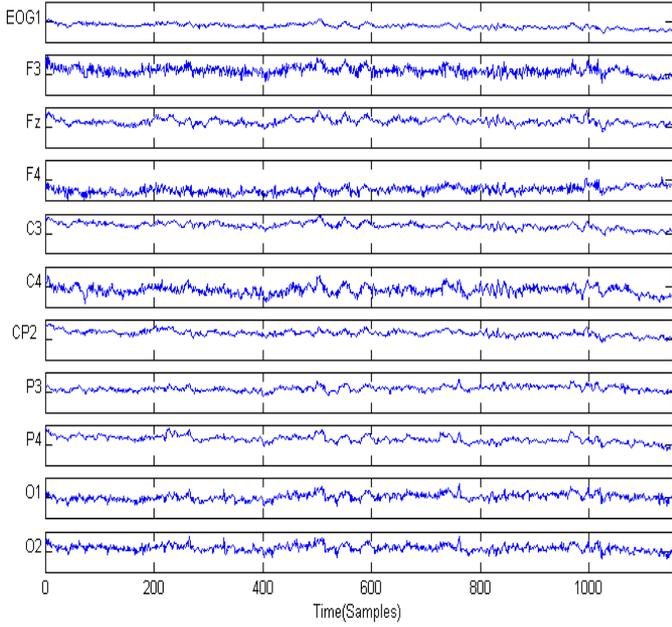


Fig. 4: Denoised EEG components obtained from RADICAL

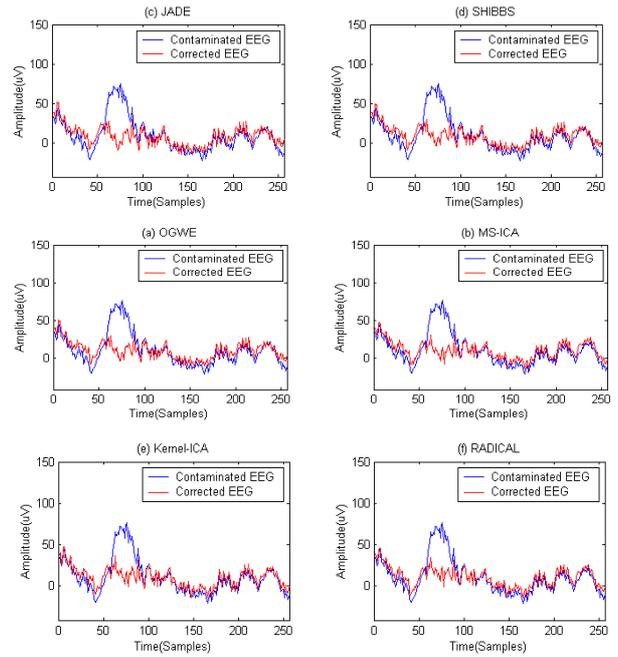


Fig. 5: Contaminated and Corrected EEG for different ICA algorithms

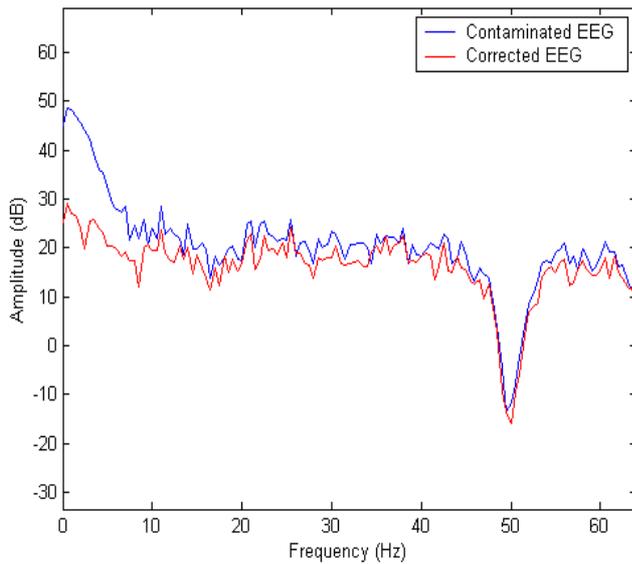


Fig. 2: Power Spectral Density spectrum signal taken from Frontal lobe channel (F3)

Tab. I: Mutual information of the estimated independent sources for various ICA Algorithms

EEG Dataset (.set files)	Original MI	MI of Independent Components					
		OGWE	ICA-MS	JADE	SHIBBS	Kernel-ICA	RADICAL
Cba1ff01a	8.9436	1.2368	2.0677	1.2073	1.2927	1.0548	0.9365
Cba1ff01b	7.7816	1.0190	1.8554	0.9729	1.0004	0.9051	0.8339

Cba1rej	9.3163	1.2149	2.1165	1.2332	1.1866	1.1249	0.9163
Cba2ff01a	8.4019	0.4093	1.2497	0.4547	0.4300	0.4675	0.3757
Cba2ff01b	7.5887	0.6253	1.2812	0.6511	0.6582	0.6901	0.5182
Clm2ff01a	8.8179	1.4946	2.9867	1.4754	1.5071	1.3209	1.1597
Ega1ff01a	8.4623	1.1700	2.2377	1.1207	1.0594	1.4795	0.9243
Ega1ff01b	7.0800	0.5814	1.2092	0.5880	0.5757	0.4799	0.4702
Ega2ff01a	8.3326	1.0428	1.8619	0.9964	1.0016	1.0707	0.8385
Ega2ff01b	8.4093	0.9175	1.3222	0.8118	0.8346	0.9003	0.7493
Fsa1ff01a	8.4249	0.5703	1.6082	0.5332	0.5500	0.5375	0.4913
Fsa1ff01b	9.0367	0.5130	1.6609	0.4638	0.4565	0.4636	0.4064
Fsa2ff01a	9.7269	0.3527	1.1514	0.4202	0.3970	0.3476	0.2847
Fsa2ff01b	8.1775	1.0664	1.7065	1.0572	1.0492	0.8094	0.8015
Gro1ff01a	10.6050	1.8100	3.6184	1.8408	1.8472	1.6607	1.2398
Gro1ff01b	11.1994	1.7463	2.8006	2.0251	1.9968	1.8256	1.6164
Gro2ff01a	8.9454	1.4400	2.6700	1.5252	1.4938	1.3810	1.1983

Acknowledgments

The authors thank the Management and the Principal of PSG College of Technology for providing the necessary facilities for this study. The authors are thankful to Dr Arnaud Delorme and Dr Scott Makeig from SCCN, for providing EEG dataset and EEGLAB toolbox.

References

- [1] Croft RJ, Barry RJ, "Removal of ocular artifact from the EEG: a review, *Clinical Neurophysiology*, 30(1), 2000, pp 5-19.
- [2] Dr A Kandaswamy, Mrs.V Krishnaveni, Dr.S.Jayaraman, Mr. N.Malmurugan and Dr.K.Ramadoss. "Removal of Ocular Artifacts from EEG - A Survey" To be published in *IETE Journal of Research*, Vol 52, No.2, March-April 2005.
- [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] Verleger R, Gasser T, Mocks J, "Correction of EOG artifacts in event-related potentials of the EEG: aspects of reliability and validity," *Psychophysiology*, 19, 1982, pp 472-480.
- [6] Whitton JL, Lue F, Moldofsky H, "A spectral method for removing eye

- movement artifacts from the EEG,” *Electroencephalography and Clinical Neurophysiology* 44, 1978, , pp 735-741.
- [7] Woestengurg JC, Verbaten MN, Slangen JL, ‘The removal of the eye movement artifact from the EEG by regression analysis in the frequency domain,” *Biological Physiology*, 16, 1983, pp 127-147.
- [8] 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,
- [9] DiMatteo, I., Genovese, C.R., Kass, R.E., 2001, “Bayesian curve fitting with free-knot splines,” *Biometrika* 88, 1055– 1073.
- [10] Wallstrom, G., Kass, R., Miller, A., Cohn, J., Fox, N., 2002, “Correction of ocular artifacts in the EEG using Bayesian adaptive regression splines,” *Bayesian Statistics*, vol. 6. Springer-Verlag, New York, NY.
- [11] Rao KD, Reddy DC, “On-line method for enhancement of electroencephalogram signals in presence of electro-oculogram artifacts using non-linear recursive least square technique,” *Med. Biol. Engg. Comput*, 35, 1995, pp 488-491.
- [12] Lagerlund, T.D., Sharbrough, F.W., Busacker, N.E., 1997, “Spatial filtering of multichannel electroencephalographic recordings through principal component analysis by singular value decomposition,” *Journal of Clinical Neurophysiology* 14 (1), 73–82.
- [13] Berg, P., Scherg, M., 1991, “Dipole modelling of eye activity and its application to the removal of eye artefacts from the EEG and MEG,” *Clinical Physics and Physiological Measurement* 12 (Suppl. A), 49– 54.
- [14] Berg, P., Scherg, M., 1994, “A multiple source approach to the correction of eye artifacts,” *Electroencephalography and Clinical Neurophysiology* 90 (3), 229– 241.
- [15] Jutten, C., Herault, J., 1991, “Blind separation of sources: Part I: Anadaptive algorithm based on neuromimetic architecture,” *Signal Processing* 24, 1–10.
- [16] Comon P., “Independent Component Analysis: A new concept?,” *Signal Processing* 36(3), 1994, pp 287-314.
- [17] Scott Makeig, Tzyy-Ping Jung, Anthony J Bell, Terrence J Sejnowski, “Independent Component Analysis of Electroencephalographic data,” *Advances in Neural Information Processing Systems* 8 MIT Press, Cambridge MA, Vol (8), pp 145-151, 1996.
- [18] Bell, A., Sejnowski, T., 1995, “An information-maximization approach to blind separation and blind deconvolution,” *Neural Computation* 7, 1129– 1159.
- [19] Jung, T.-P., Makeig, S., Humphries, C., Lee, T.-W., McKeown, M.J., Iragui, V., Sejnowski, T.J., 2000, “Removing electroencephalographic artifacts by blind source separation,” *Psychophysiology* 37, 163–178.
- [20] Vigário, R.N., 1997, “Extraction of ocular artefacts from EEG using independent component analysis,” *Electroencephalography and Clinical Neurophysiology* 103, 394– 404.
- [21] A. Hyvärinen and E. Oja, “A fast fixed-point algorithm for independent component analysis,” *Neural Computation*, vol. 9, 1997: 1483–1492.
- [22] Delorme.A, Makeig.. S & Sejnowski T, “Automatic artifact rejection for EEG data using high-order statistics and independent component analysis”, *Proceedings of the Third International ICA Conference*, December 9-12, 2001, San Deigo.
- [23] Beloucharani, K Meriam, J F Cardoso and E Moulines, “A blind source separation technique using second order statistics”, *IEEE Transactions on*

- Signal Processing, 45, Feb 1997: 434-444.
- [24]Carrie A.Joyce, Irina F Gorodnitsky and Marta Kutas, “Automatic removal of eye movement and blink artifacts from EEG data using blind component separation”,*Psychophysiology*, Volume 41: Issue 2 March 2004, :313-325
- [25]N.Nicolaou and S.J.Nasuto, “Temporal Independent Component Analysis for automatic artefact removal from EEG”, 2nd International Conference on Medical Signal and Information Processing, Malta, September 2004: 5-8.
- [26]A Ziehe and K R Muller, “TDSEP – an efficient algorithm for blind separation using time structure” in *Proceedings of ICANN ’98*, December 1998: 675-680.
- [27]L. Molgedey and H. Schuster, “Separation of independent signals using time-delayed correlations,” *Physical Review Letters*, vol. 72, no. 23, pp. 3634–3637, 1994.
- [28]Juan J. Murillo-Fuentes and Rafael Boloix-Tortosa, Francisco J. Gonz’alez-Serrano, “Initialized Jacobi Optimization in Independent Component Analysis”.
- [29]Jean-François Cardoso, “High-order contrasts for independent component analysis”, *Neural Computation*, vol. 11, no 1, Jan. 1999, pp. 157—192.
- [30]Cardoso, J.-F., Bose, S., & Friedlander, B. (1996), “On optimal source separation based on second and fourth order cumulants,” *Proc.IEEEWorkshoponSSAP*. Corfu, Greece.
- [31]F. R. Bach and M. I. Jordan, “Kernel independent component analysis.” *J. of Machine Learning Research*, 3:1–48, 2002.
- [32]Erik G. Miller and John W. Fisher III, “Independent components analysis by direct entropy minimization,” *Tech. Rep. UCB/CSD-03-1221*, University of California at Berkeley, January 2003.
- [33]Alexander Kraskov, Harald Stogbauer and Peter Grassberger, “Estimating Mutual Information”,. ArXiv:cond-mat/0305641 v1 28th May 2003
- [34]Vinther, Michael, “Independent Component Analysis of Evoked Potentials in EEG”.
- [35]Joliffe I T, “Principal Component Analysis,” Springer Verlag, New York, 1986.
- [36]J. Larsen L.K. Hansen and T. Kolenda, “On independent component analysis for multimedia signals,” *Multimedia Image and Video Processing*, CRC Press, vol. Chapter 7, pp. 175–200, 2000.
- [37]B. Schölkopf and A. J. Smola, “Learning with Kernels.” MIT Press, 2001.
- [38]Oldrich Vasicek, “A test for normality based on sample entropy,” *Journal of the Royal Statistical Society, Series B*, vol. 38, no. 1, pp. 54–59, 1976.
- [39] A. Hyvärinen and E. Oja, “A survey on independent component analysis,” Helsinki University of Technology.
- [40]Mrs V Krishnaveni, Dr S Jayaraman, Mr N Malmurugan, Ms Chaitanya Mathi, Dr K Ramadoss, “Quantitative Evaluation of Signal Separation Algorithms for the removal of ocular artifacts from EEG” To be published in *Technology, Journal of PSG College of Technology & Polytechnic, Coimbtore, INDIA*.