Comparative Evaluation of Various Independent Components (ICA) Tech for the removal of artifacts of EEG Signals

B.Paulchamy[1] Ilavennila [2] J.Jaya[3] R.Saravanakumar[4]
1. Research Scholar, Anna University, Coimbatore,
2. Asst professor, PSG College of Technology, Coimbatore,
3. Research Scholar, Anna University, Chennai.
4. Lecturer ,Hindusthan Institute of Technology,Coimbatore.

Summary
In this paper, Independent Component Analysis (ICA) is applied to EEG signals collected from different mental tasks in order to remove the artifacts from the EEG signals. ICA is a statistical and computational technique for revealing hidden factors that underlie sets of random signals. In the ICA model the data samples are assumed to be linear mixture of some unknown latent variables, and the mixing system is also unknown. The latent variables are assumed to have a nongaussian distribution. These variables are the independent components of the observed data which can be found, up to some degree of accuracy, using different algorithms based on ICA techniques. There are several algorithms based on different approaches for ICA widely in use for all sort of applications. These algorithms include, but not limited to, the popular Fast-ICA, Joint Approximate Diagonalization of Eigen values (JADE), Infomax, and Extended Infomax etc. Fast-ICA is based on the optimization of negentropy of the datasets. Infomax and Extended Infomax are based on the minimization of mutual information between the data variables. JADE is based on the fourth-order cumulate matrices of the input data. A framework for accommodating four ICA algorithms is developed to estimate the convergence speed of the algorithms and hence selects the best algorithm for the specific type of data.
Key words: Artifacts, EEG Signal, Fast ICA, JADE, BCI, Infomax, Entropy

1. Introduction

Electroencephalography (EEG) is a medical imaging technique that reads the scalp electrical activity generated by the brain structures [1]. Electrical impulses generated by nerve firings in the brain diffuse through the head and can be measured by electrodes placed on the scalp. The EEG gives a coarse view of neural activity and has been used to non-invasively study cognitive processes and the physiology of the brain. The greatest advantage of EEG when compared with other medical imaging techniques is its speed.
When the EEG signals measured by electrodes placed on the scalp and are always under the influences of artifacts. Those artifacts include: line noise from the power line, eye blinks, eye movements, heartbeat, breathing, and other muscle activities. Because of these artifacts contained in EEG, the pattern detection in EEG produced from normal mental states is a very difficult problem.

2. Literature Review

Artifacts in EEG are commonly handled by discarding the affected segments of EEG. The simplest approach is to discard a fixed length segment, perhaps one second, from the time an artifact is detected. The recognition of the eye blink and eye movement artifacts are generally effected by detecting a voltage increase in the elctrooculogram (EOG) channel above a threshold, generally 100 µV. Other artifacts are generally ignored or manually marked by a practitioner and discarded. Discarding segments of EEG data with artifacts can greatly decrease the amount of data available for analysis.

3. Proposed Methodology

3.1. Independent component analysis

ICA is an extension of PCA but it is more powerful that PCA in the field of signal analysis. In mid 90's several new ICA algorithms [17] were introduced with impressive demonstrations on problems like separating different speech signals from a mixed signal. The applications of ICA include but not limited to the fields of biomedical, telecommunications, audio and video signal processing feature extraction, data mining, and functional time series analysis. Generally, ICA technique can be regarded as a technique to separate signals from a mixture. There are several ICA algorithms in use. Some of these algorithms are FastICA, JADE and First Order Blind Identification (FOBI), Maximum Likelihood and Infomax, algorithms based on Kernel methods, and algorithms using time structure like Second Order Blind Identification (SOBI) .

3.2 ICA Model: A simple mathematical representation of ICA model [18] is as follows:Consider a simple linear model which consists of **N** sources of **T** samples i.e. $S_i = [s(1)_i \ldots s(t)_i \ldots s_i(T)]$. The symbol **t** here

Manuscript received March 5, 2010
Manuscript revised March 20, 2010

represents time but it may represent some other parameter like space. **M** weighted mixtures of the sources are observed as **X**, where $X_i = [x_i(1)....x_i(t).....x_i(T)]$. This can be represented as in Eqn.(1)

$$x = \mathbf{A}s + \mathbf{n} \quad (1)$$

Where $X = \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_M \end{bmatrix}$, $S = \begin{bmatrix} s_1 \\ s_2 \\ . \\ . \\ . \\ s_N \end{bmatrix}$ and $n = \begin{bmatrix} n_1 \\ n_2 \\ . \\ . \\ . \\ n_K \end{bmatrix}$ (2)

and **n** represents the Additive White Gaussian Noise (AWGN). It is assumed that there are at least as many observations as sources i.e., $M \geq N$. The $M \times N$ matrix **A** is represented as in Eqn.(3)

$$A = \begin{bmatrix} a_{11}a_{12}...a_{1N} \\ a_{21}a_{22}...a_{2N} \\ . \\ . \\ a_{M1}a_{M2}a_{MN} \end{bmatrix} \quad .(3)$$

A is called the mixing matrix. The estimation of the matrix **S** with knowledge of X is the linear source separation problem. This is schematically shown in Fig. 1 A is the mixing matrix and **B** is the unmixing matrix.

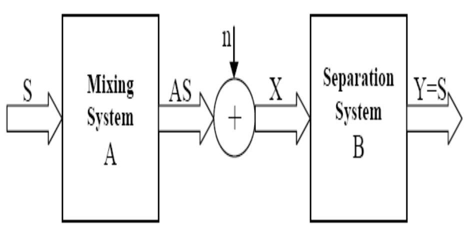


Fig.1. Illustration of Mixing and Separation System for ICA

The following assumptions [18] ensure that the ICA model estimates the independent components meaningfully. Actually the first assumption is the only true requirement which ICA demands. The other assumptions ensure that the estimated independent components are unique.

The latent variables (or independent components) are statistically independent and the mixing is linear.There is no more than one gaussian signal among Thelatent variables and the latent variables have cumulative density function not much different from a logistic sigmoid

The number of observed signals, *m*, is greater than or equal to the number of latent variables, *n* (i.e. $m \geq n$). If *n* > *m*, we come to a special category of Independent Component Analysis called ICA with over-complete bases .In such a case the mixed signals do not have enough information to separate the independent components. There have been attempts to solve this particular problem but no rigorous proofs exist as of yet. If *m* > *n* then there is redundancy in the mixed signals. The ICA model works ideally when *n* = *m*.

The mixing matrix is of full column rank, which means that the rows of the mixing matrix are linearly independent. If the mixing matrix is not of full rank then the mixed signals will be linear multiples of one another.

The propagation delay of the mixing medium is negligible.

3.3. Pre-processing of Data for ICA

Before applying an ICA algorithm on the data, it is usually very useful to do some pre-processing [18]. In this section, some pre-processing techniques that make the problem of ICA estimation simpler and better conditioned are given.

4. ICA Algorithms

4.1. Algebraic ICA Algorithm

An algebraic solution to ICA is proposed by Taro Yamaguchi. This is a non-iterative algorithm but becomes extremely complex to compute when the number of sources goes more than two. For two sources separation it works very fast .Two observed signals x_1 and x_2 are given by linear mixture of two independent original signals s_1 and s_2 as in Eqn.(4):

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} 1\alpha \\ \beta 1 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} (4)$$

where α and β are unknown mixing rates. The algebraic solution to α and β are given by Eqn.(5) and Eqn.(6) . $\beta = \frac{\alpha C_2 - C_3}{\alpha C_3 - C_1}$

(5)

$$(C_2C_{10} - C_{11}C_3)\alpha^4 + (3C_9C_3 - 3C_8C_2 - C_3C_{10} + C_1C_{11})\alpha^3 + (3C_6C_2 + 3C_8C_3 - 3C_9C_1 - 3C_7C_3)\alpha^2 + (C_5C_3 + 3C_7C_1 - 3C_6C_3 - 3C_2C_4)\alpha + (C_3C_4 - C_1C_5) = 0$$

(6)

Where $C_1, C_2, C_3, C_4, C_5, C_7, C_8, C_9, C_{10}$, and C_{11} are as shown in Eqn.(7)

$$C_1 = E[X_1^2] - \{E[X_1]\}^2,$$
$$C_2 = E[X_2^2] - \{E[X_2]\}^2,$$
$$C_3 = E[X_1 X_2] - E[X_1]E[X_2],$$
$$C_4 = E[X_1^4] - E[X_1^3]E[X_1],$$
$$C_5 = E[X_1^3 X_2] - E[X_1^3]E[X_2],$$
$$C_6 = E[X_1^3 X_2] - E[X_1^2 X_2]E[X_1],$$
$$C_7 = E[X_1^2 X_2^2] - E[X_1^2 X_2]E[X_2],$$
$$C_8 = E[X_1^2 X_2^2] - E[X_1 X_2^2]E[X_1],$$
$$C_9 = E[X_1 X_2^3] - E[X_1 X_2^2]E[X_2],$$
$$C_{10} = E[X_1 X_2^3] - E[X_1]E[X_2^3],$$
$$C_{11} = E[X_2^4] - E[X_2^3]E[X_2] \quad (7)$$

Where E [.] denotes the expectation operation. α and β are obtained by solving the Eqn.(5), Eqn.(6) and Eqn.(7) with the Ferrari method. Excluding the solutions having non-zero imaginary parts and negative sizes the proper solution is selected. Original independent signals are computed from Eqn (6) by solving value of α and β .

4.2 Fast ICA

The advantage of using the gradient method to maximize negentropy is that the inputs **z**(t) can be used in the algorithm at once, thus enabling fast adaptation in non-stationary environment. However convergence is slow and depends on a good choice of learning rate γ .

To make this method efficient, a fast-fixed point algorithm is devised, also called Fast ICA for Negentropy. To understand this algorithm it should be noted that at a stable point of the gradient algorithm, the gradient must be pointing towards **w** or in other words it must be a scalar multiple of **w**. That means that adding the gradient of negentropy in **w** would not change its direction at the stable points and hence convergence can be obtained. The t Eqn.(8) can be written as ,

$$W \leftarrow E\{ZG'(W^T Z)\}$$
$$W \leftarrow \frac{W}{\|W\|} \quad (8)$$

The coefficient γ is omitted because it would be eliminated by the normalization anyways. Iteration in Eqn.(8) does not however have as good of a convergence as the one using kurtosis. The reason behind this is the non-polynomial moments (G's) do not have same nice algebraic properties as cumulants like kurtosis. Hence to get a better convergent algorithm the iteration in Eqn.(9) has to be modified. This modification can simply be done by adding some multiple of **w** to the both sides of the iterant term in Eqn.(9) and then changing the value of multiple to find a better convergence speed as in Eqn.(9),

$$W = E\{zG'(W^T z)\} \Leftrightarrow (1+\alpha)W = E\{zG'(W^T z)\} + \alpha W \quad (9)$$

Adding a multiple of **w** to the both sides of Eqn. (10) would not change the direction of the vector and after normalization in the next step **w** will be constrained to the unit sphere again. A suitable value of α and thus the Fast ICA algorithm can be found using Newton's method for solving equations. Newton's method can briefly explained as follows:

To find a maxima or minima of any function with respect to some variable, first the function is expanded using Taylor's series and the terms above the quadratic terms are dropped to keep it manageable (since higher order terms don't contribute a lot in the total value of the function). Let E (not expectation) be a cost or error function which has to be minimized around vector **w**(n) having m elements and n being the number of iteration. The change in the cost function can be written as in Eqn.(12)

$$\Delta E(Wn) = E(W(n+1)) - E(W(n))$$
$$\approx E(W(n) + g^T(n)\Delta W(n) + \frac{1}{2}\Delta W(n)^T H(n)\Delta W(n)) - E(W(n)) \quad (10)$$

where **g**(n)= *mx1* gradient vector of cost function evaluated at **w**(n), and, **H**(n) is an *mxm* 2nd order derivative matrix of the cost function *E(***w***(n))* evaluated at **w**(n) , called *Hessian Matrix*. Hessian Matrix **H** is given by Eqn.(11),

$$H = \Delta^2 E(W(n))$$
$$= \begin{bmatrix} \frac{\partial^2 E}{\partial w_1^2} \frac{\partial^2 E}{\partial w_1 \partial w_2} \cdots\cdots \frac{\partial^2 E}{\partial w_1 \partial w_m} \\ \frac{\partial^2 E}{\partial w_1^2} \frac{\partial^2 E}{\partial w_1 \partial w_2} \cdots\cdots \frac{\partial^2 E}{\partial w_1 \partial w_m} \\ . \\ . \\ . \\ \frac{\partial^2 E}{\partial w_1^2} \frac{\partial^2 E}{\partial w_1 \partial w_2} \cdots\cdots \frac{\partial^2 E}{\partial w_m^2} \end{bmatrix} \quad (11)$$

Differentiating ΔE (W (n)) w.r.t Δ w (n) to find out the minimal value of ΔE gives the condition as in Eqn.(12),

$$g(n) + H(n)\Delta W(n) = 0$$

$$\Rightarrow \quad \Delta W(n) = -H^{-1}(n)g(n)$$

$$W(n+1) = W(n) - H^{-1}(n)g(n)$$

$$W(n+1) = W(n) - \left[\frac{\partial^2 E(Wn))}{\partial W^2}\right]^{-1}\left[\frac{\partial E(W(n))}{\partial W}\right] \quad (12).$$

Expression in Eqn.(12) is the Newton's method for updating the vector **w** to move towards the minimization of the cost function. The advantage of Newton's method is fast convergence but as it can be seen that it is computationally more intensive since one has to calculate inverse of Hessian matrix at each step.

In order to avoid the cost and time consuming calculation of the inverse of Hessian matrix in the Newton's method, an approximation of this method is developed that avoids the use of matrix inversion without sacrificing its essence to employ ICA algorithm. The approximation of Newton's method calls for the use of Lagrangian rule for constrained optimization. Lagrangian rule for constrained optimization can be briefly described as follows.Assume a cost function E(**w**) (E is not expectation) which is suppose to be minimized or maximized under some constraint H_i(**w**) = 0, where i = 1,2,3, …., k. One can write the Lagrangian function based on the given information as in Eqn.(13).

$$L(W, \lambda_1, \lambda_2, ..., \lambda_k) = E(W) + \sum_{i=1}^{k} \lambda_i H_i(W) \quad (13)$$

where $\lambda_1, \lambda_2, ..., \lambda_k$ are called Lagrangian multipliers. The minimum (maximum) point of Eqn.(14) where its gradient is zero with respect to both **w** and all of the λ_i gives the solution to the original constrained problem, i.e., minimization of E(**w**) under some constraint H_i(**w**) = 0. The gradient of $L(W, \lambda_1, \lambda_2, ..., \lambda_k)$ with respect to λ_i gives the ith constraint function H_i (**w**), so putting all these to zero will give the original constraint condition. When gradient of $L(W, \lambda_1, \lambda_2, ..., \lambda_k)$ is taken with respect to **w** and equate it to zero, one will get the

$$\frac{\partial L(W, \lambda_1, \lambda_2, ..., \lambda_k}{\partial W} = \frac{\partial E(W)}{\partial W} + \sum_{i=1}^{k} \lambda_i \frac{\partial H_i(W)}{\partial W} = 0 \quad (14)$$

Hence the minimization problem has been changed into two sets of equations that are much easier to solve. A possible way to solve these two sets, one given by the constraints, the other by Eqn.(15), is some appropriate iteration method like Newton iteration. From Eqn.(16) the optima of $E\{G(\mathbf{w}^T\mathbf{z})\}$ for constraint $\|\mathbf{w}\|^2 = 1$ can be evaluated as in Eqn.(15),

$$\frac{\partial E\{G(W^T z)\}}{\partial W} + \lambda \frac{\partial(\|W\|^2 - 1)}{\partial W} = 0$$

$$\Rightarrow E\{zG^{'}(W^T z)\} + 2\lambda w = 0$$

$$E\{zG^{'}(W^T z)\} + \beta w = 0 \quad (15)$$

where $H(\mathbf{w}) = \|\mathbf{w}\|^2 - 1 = 0$ is the only constraint to find out the extrema of $E\{G(\mathbf{w}^T\mathbf{z})\}$. To solve Eqn.(16) one can use Newton's method to find the optima with respect to **w**. Let $F = E\{zG^{'}(W^T z)\} + \beta w$, the derivative of F, i.e., the second derivative of Lagrangian function can be evaluated as in Eqn.(16),

$$\frac{\partial F}{\partial W} - E\{zz^T G^{"}(W^T z)\} + \beta I \quad (16)$$

Thus the Newton iteration from Eqn. (13) can be written as in Eqn. (17),

$$W \leftarrow W - \frac{\left[\frac{\partial L}{\partial W}\right]}{\left[\frac{\partial F}{\partial W}\right]} = W - \frac{E\{zG^{'}(W^T z)\} + \beta W}{E\{zz^T G^{"}(W^T z)\} + \beta I} \quad (17)$$

To simplify the calculations, since the vector **z** is sphered then $\frac{\partial F}{\partial W}$ can be approximated as in Eqn.(18)

$$\frac{\partial F}{\partial W} = E\{zz^T G^{"}(W^T z)\} + \beta I$$

$$\approx E\{zz^T\}E\{G^{"}(W^T z)\} + \beta I$$

$$= E\{G^{''}(W^T z)\} + \beta I = [E\{G^{''}(W^T z)\} + \beta]I \quad (18)$$

Hence the gradient becomes a diagonal matrix and can easily be inverted. Thus the algorithm becomes as in Eqn.(19),

$$W \leftarrow W - \frac{E\{zG^{'}(W^T z)\} + \beta W}{[E\{G^{''}(W^T z)\} + \beta]I} \quad (19)$$

Multiplying both sides of Eqn.(21) by $-[E\{G^{''}(W^T z)\} + \beta]I$ and simplifying the resulting expression can be written as Eqn.(20),

$$-W[E\{G^{''}(W^T z)\} + \beta]I \leftarrow E\{zG^{'}(W^T z)\} - EG^{''}(W^T z)\}W \quad (20)$$

Left hand side of Eqn. (20) is nothing but a new variable to which right hand side value will be assigned, hence the Fast ICA algorithm based on negentropy will become as in Eqn.(21),

$$W_{new} \leftarrow E\{zG^{'}(W^T z)\} - E\{G^{''}(W^T z)\}W_{old} \quad (21)$$

Following is a brief summary of Fast ICA algorithm based on negentropy for finding one maximally non-Gaussian direction, i.e., estimating one independent component.

4.3 JADE

JADE is an algorithm that uses significant eigenpairs of the cumulant tensor F(M) to find out the estimated values of independent components. In this algorithm the tensor eigenvalue decomposition is considered as more of a preprocessing step. Eigenvalue decomposition can be viewed as diagonalization. The idea is to diagonalize F(M) for any M using the matrix W. In other words, $WF(M)W^T$ is diagonal. This is because the matrix **F** consists of a linear combination of terms of the form $\mathbf{w}_i\mathbf{w}_i,^T$ assuming that the ICA model holds. Hence the goal is to take a set of significant eigenmatrices, M_i, and try to make the matrices $WF(M)W^T$ as diagonal as possible. They might not be made exactly diagonal since the model doesn't hold exactly because of some sampling errors.

Let $Q = WF(M)W^T$, then maximization of the sum of the square of diagonal elements of the Eqn.(22),

$$J_{JADE}(W) = \sum_i \left\| diag(WF(M_i)W^T \right\|^T \quad (22)$$

Maximization of J_{JADE} is then one method of joint approximate diagonalization of the $F(M_i)$.

Two of the main steps in the JADE algorithm are to find the significant eigenpairs of the cumulant tensor $\{\lambda\ r, M\ r\ |1 \le r \le n\}$, and to jointly diagonalize the JADE criterion J_{JADE}(W). These two steps that lead to the JADE algorithm are discussed next.

4.4 Infomax Method

The Infomax method has been proposed for performing linear ICA based on a principle of maximum information preservation (hence its name). However, it can also be seen as a maximum likelihood method, or as a method based on the minimization of mutual information. Infomax uses a network whose structure is depicted in Fig.6.1 (the figure shows the case of two components; extension to a larger number of components is straightforward). W is a linear block, yielding the Eqn.(23)

$$Y=WX \quad (23)$$

This block thus performs just a product by a square matrix (we shall designate both the block and the matrix by the same letter since this will cause no confusion). After optimization, the components of Y are expected to be as independent from one another as possible. Blocks ϕ_i are auxiliary, being used only during the optimization phase. Each of them implements a nonlinear function (that we shall also designate by ϕ_i). These functions must be increasing, with values in [0; 1]. The optimization of W is made by maximizing the output entropy, H (Z).

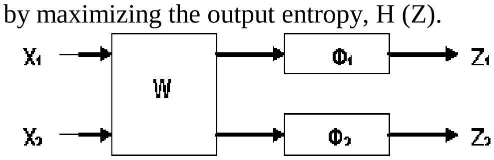


Fig.2 Structure of the Infomax ICA System

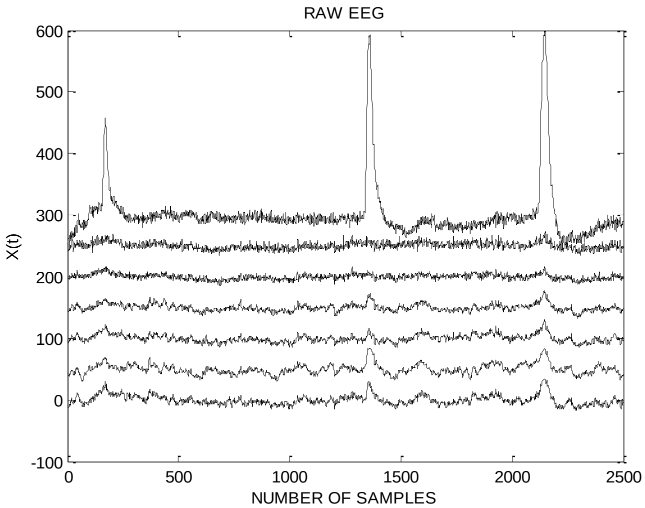
4.5 Extended Infomax Method

The algorithm of Bell and Sejnowski which uses a sigmoidal activation function is specifically suited to separate signals with super-Gaussian distribution (i.e. positive kurtosis). Lee and Sejnowski [23] proposed an extension of Infomax ICA that is able to separate with sub and as well as super Gaussian distribution. This preserves the ICA architecture of Infomax algorithm, but it uses a learning rule derived by Girolami and Fyfe. It determines the sign changes (positive to negative and vice versa) required by the algorithm to handle both sub and super Gaussian distributions.

5. Results

Data were recorded for 10 seconds during each task and each task was repeated five times per session. Subjects attended two sessions recorded on separate weeks, resulting in a total of ten trials for each task. With a 250 Hz sampling rate, each 10 second trial produces 2,500 samples per channel.

Trail- 1



Trail-2

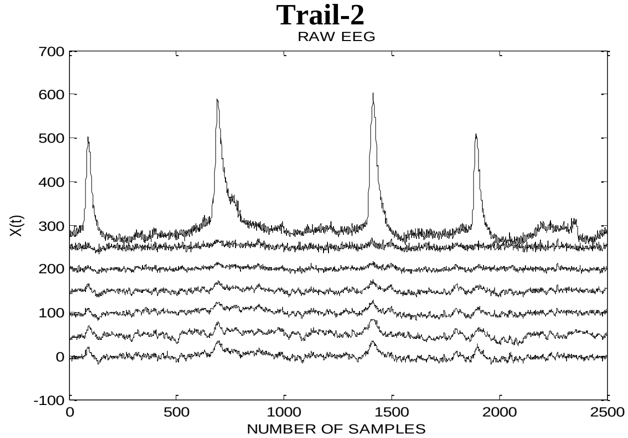


Fig.3 EEG Recording of a Subject during Mental Multiplication

5.1 Results of Fast ICA Algorithm

Parameters:

Nonlinearity: log(cosh(y))

No. of iterations: 100

Max. weight change: 10e-300

Fig 4. shows the independent components obtained using the Fast ICA algorithm from the EEG data mixed with EOG which is shown Fig.3. In Fig.3, the EOG recording appears in all channels of original EEG data. But from Fig.4, it can find the EOG artifact is concentrated in component 7(order from top to bottom) and no more appears in other independent components.

Execution time in seconds: Trail1: 1.59, Trail 2: 1.61

Trail- 1 &Trail-2

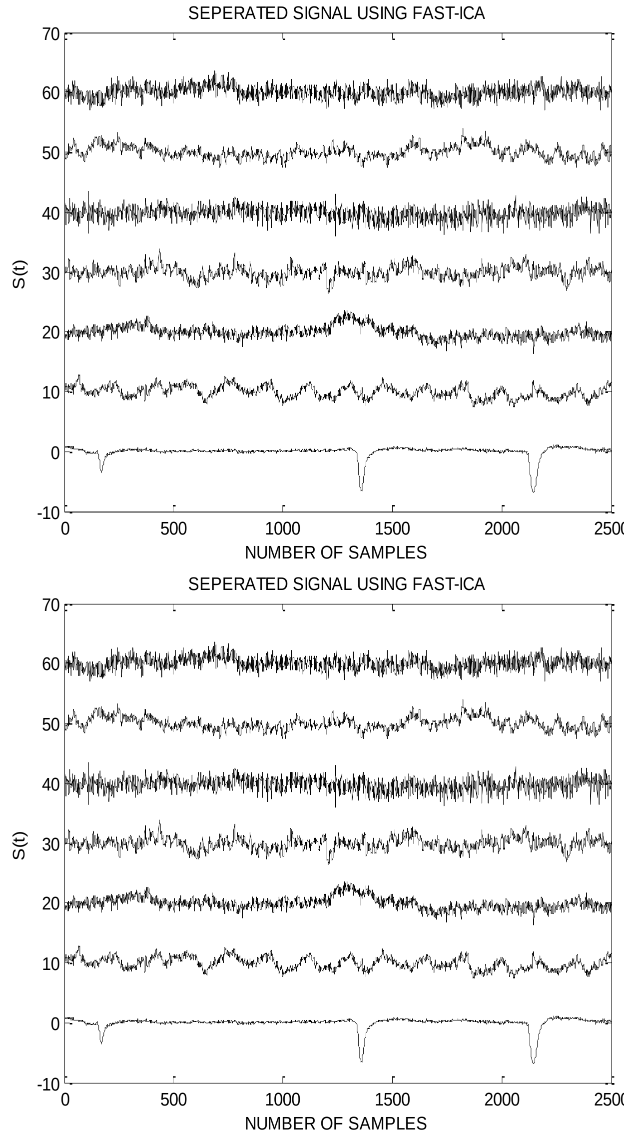


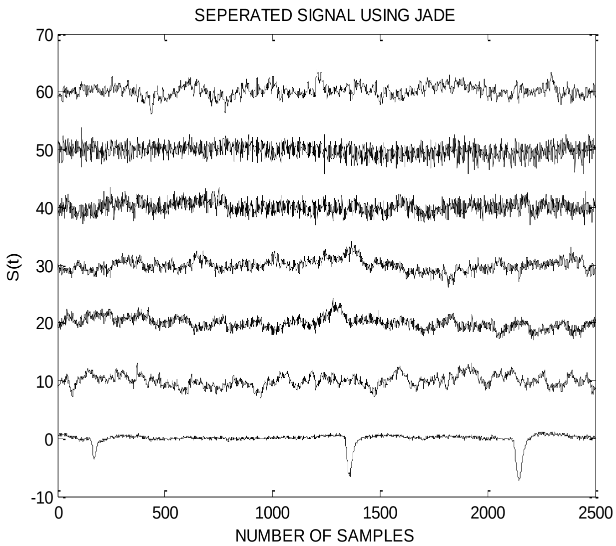
Fig 4.Independent Component s obtained using Fast ICA

5.2 Results of JADE Algorithm:

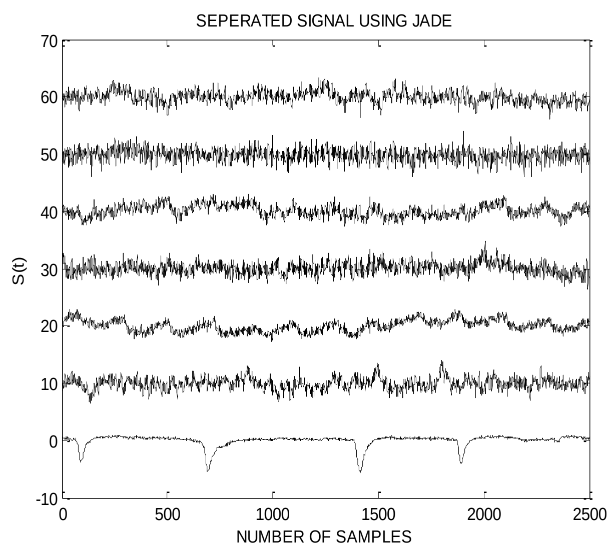
No adjustable parameters.

Fig 5. shows the independent components obtained using the JADE algorithm for the EEG data of Fig.3..

Trail-1



Trail 2



(b)

Fig.5. Independent Components Obtained using JADE.
Execution time in seconds: Trail1: 0.99, Trail 2: 1.01

5.3Results of Infomax:

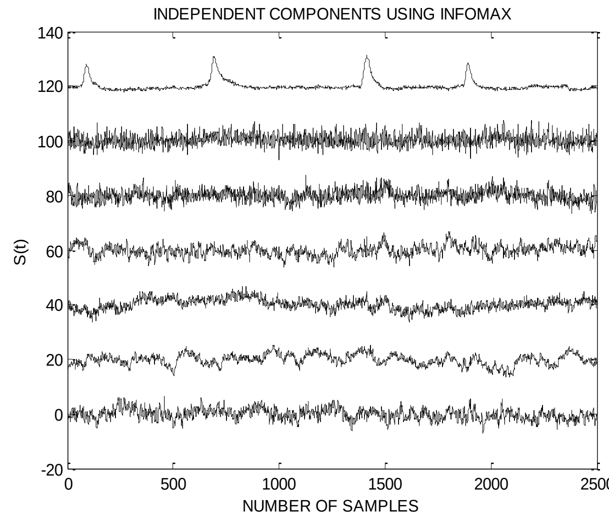
Parameters: learning rate= 0.1;
Max. Change in weight =1e-3
Transformation function =logistic sigmoid

$$= \frac{1}{1+e^{-u}}$$

Number of iterations: 5
Bias weight =0
Initial weight =Identity Matrix

Fig.6shows the independent components obtained using the Infomax algorithm for the EEG data of Fig.3.

Trail1



Trail 2

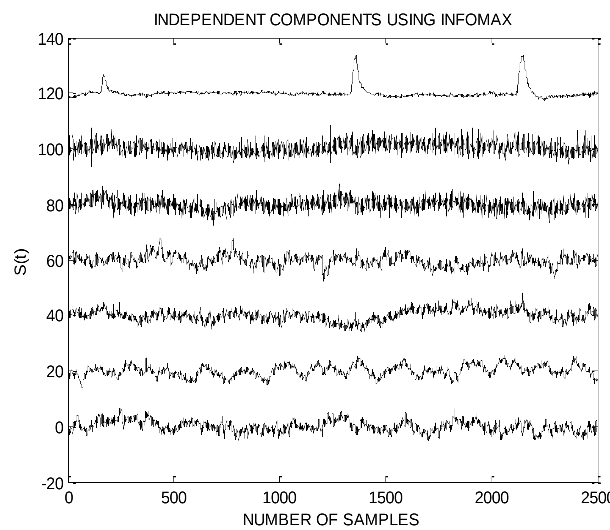


Fig.6 Independent Components Obtained using Infomax

Execution time in seconds:
Trail 1: 2.20 Trail 2: 2.52

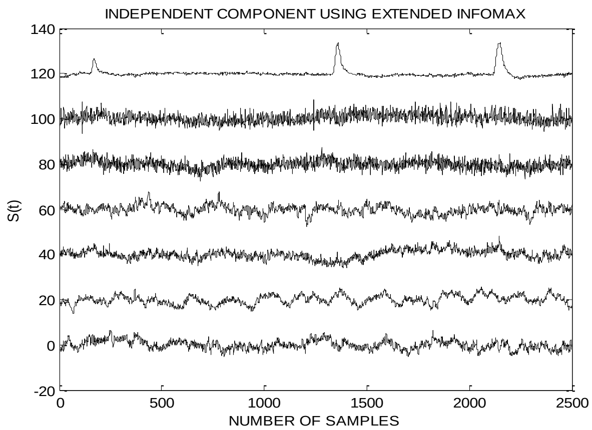
5.4 Results of Extended Infomax

Parameters:

Min. Weight-change: 1e-3
Number of iterations=512
Signs: -1: Subgaussian
1: Supergaussian
Initial weight= Identity matrix,
Nonlinearity = tanh(u),
Bias =0
Learning rate =0.1

Fig.7 shows the independent components obtained using the Extended Infomax algorithm for the EEG data of Fig.3.

Trail-1



Trail-2

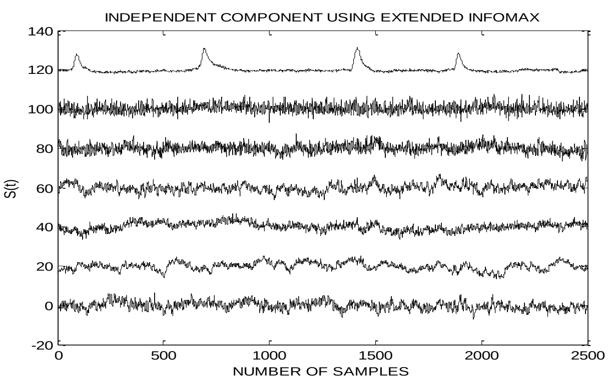


Fig.7.Independent Components Obtained using extended infomax

Execution time in seconds: Trail 1: 3.07 Trail 2: 3.03

5.5Reconstruction:

Fig 8.shows the reconstructed EEg without EOG,by using scalp map topography.

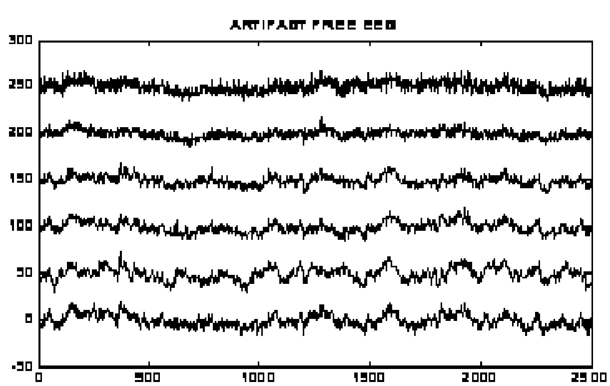


Fig.8 Clean EEG after removing EOG

6.Performance Comparision of Algorithms:

Computation Time: The computation time i.e., the timetaken by algorithm to separate the EEG signal is measured.For comparison of algorithms mental multiplication task EEG which is measured for five trails is used. The computation time in seconds is tabulated in Table 1

Entropy: Entropy of original mental multiplication EEG signals of five trails is given in Table **2. Entropy** of the separated signal by different ICA algorithms is given Table3.

Table 1 Computation Time for ICA Algorithms in Seconds

Fast ICA	JADE	Infomax	Extended Infomax
1.59	0.99	2.20	3.07
1.61	1.01	2.52	3.03
1.78	1.12	2.71	3.44
1.70	1.01	2.29	3.14
1.50	0.98	2.43	3.32

Table 2 Entropy of Original EEG

Trail 1	Trail 2	Trail 3	Trail 4	Trail 5
1.646	1.664	1.759	1.787	1.705

Table 3 Entropy of EEG after Removal of EOG

Algorithm Mult	Fast ICA	JADE	Infomax	Extended Infomax
Trail 1	4.409	4.468	4.001	2.996
Trail 2	3.781	4.570	3.518	2.858
Trail 3	3.955	4.124	3.878	2.899
Trail 4	4.293	4.222	3.650	3.139
Trail 5	4.092	4.335	3.91	3.050

7. Conclusion

ICA plays a vital role in removing of artifacts in EEG signals .It maintains the similarity in their patterns when subject is performing the mental task. BCI systems using EEG as control signal suffers from the artifact problem. The traditional methods applied for remove artifacts can only compromise between eliminating artifacts and protecting useful signals so that the result is not very satisfying. However, ICA method can protect the useful signals as well as obviously weaken even entirely remove the artifacts in multi channel EEG signals, this characteristic of ICA is the key to get stable EEG patterns which can be used for mental task classification.

8. Future Enhancement

Usage of ICA method for pre-processing EEG in BCI systems is still in research stage only. ICA algorithms use different tuning parameters and contrast functions during the separation which are to be adjusted continuously depending upon type of data. Though, ICA separates artifacts from EEG efficiently, as BCI system is real time system, the usage of ICA for BCI with adjustable parameters makes them non-realistic. So, a method can be found for all type s of data to fix the parameters. This paper can also be extended to classify the mental tasks with the improved efficiency.

Reference

[1] M.M.Teplan,"Fundamentals of EEG measurement," Measu rement and science review, vol.2, pp.1-11, 2002.

[2] R.D.Bickford,"Electroencephalography,"In: Adelman G.ed Encyclopedia of Neuroscience, Brikhauser, Cambridge (USA).

[3] Scott Makeig,"Developments in Neural Human-Systems InterfaceDesign,"downloadedfrom ftp://www/cnl.salk.edu/pup/

[4] G.Gratton M.G.Coles, and E. Donchin,"A new method for off-line removal of ocular artifacts, "Electroencep halographyandClinical neurophysiology, vol.55, pp.468-484,

[5] S.A.Hillyard and R.Galambos,"Eye-movement artifact in The CNV, "Electroencephalography and linicalneuro physiology,vol.28, pp.173-182, 1970.

[6] R. Verleger and T.Gasser,"Correction of EOG artifacts in event related potentials of EEG: Aspects of reliability and validity,"Psychophysiology, vol.19, pp.472-480,1982

[7] P.Berg and M.Scherg,"Dipole model of eye activity and itsapplication to the removal of eye artifacts from the EEG and MEG," Clinical Physics and Physiological Measurements, vol.12, pp.4954,1991.

[8] A.Cichoki and S.Vorobyov,"Application of ICA for automatic noise and interference cancellation in multisensory biomedical signals,", In Proceedings of Second International Workshop on ICA and BSS, pp.621-626, June 2000.

[9] S.Verobyov and A.Cichocki,"Blind noise reduction from multisensory signals using using ICA and subspace filtering, with application to EEG analysis," Biological Cybernetics, vol.86, pp.293-303, 2002.

[10] L.Vigon, and M.R.Saatchi,"Quantitative evaluation of techniques for ocular artifacts removal," IEEE Proceedings-Science and Measurement technology, vol.147 September 2000.

[11] T.P.Jung and C.Humphries,"removing electroencephalogr aphic artifacts by blind source seperation," Psychophysiology, vol.37, pp.163-178, 2000.

[12] Aapo Hyvarinen.,"Independent Component Analysis: Algorithms and Applications," IEEE transactions on Neural Networks, vol.13, pp.411-430, 2000.

[13] T.W.Lee and T.J.sejnowski, "A Unifying Information-theoretic Framework for Independent component analysis," Comp.math.Appl, vol.31, pp.1-21, April 2003,.

[14] M.Laubach and M.Shuler, "Independent Component Analysis for quantifying neuronal ensemble interactions," Journal of Neuroscience Methods, vol.94, pp.141-154,1999.

[15] Jane lehtonen, "EEG-based Brain Computer Interfaces,"Helsinki University of Technology, Finland, May 2002.

[16] Pierre Comon, "Independent Component Analysis: a new concept,"ELSEVIER, Signal Processing, vol.36, pp.287-314, 1992.

[17] Hyvarinen., "Survey on Independent component analysis,

[18] Pierre Comon., "Independent Component Analysis: a new concept," ELSEVIER, Signal Processing, vol.36, pp.287-314, 1992.

[19] D.P.Acharya,"Development of novel independent component analysis techniques and their applications,"National Institute Technology, Rourkela, india.

[20] Aapo Hyvarinen.,"fast and Robust fixed point algorithm for independent component analysis," IEEE transactions on Neural Networks, Vol.10, pp.626-634, 1999.

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

[22] 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.

[23] T.W.Lee and T.J.Sejnowski,"Independent component analysis for sub-gaussian and super-gaussian mixtures,"Proceedings of 4th joint symposiums on Neural computations, vol.7, pp.132-139, 1997.

[24] Sarah M. Hosni, and Mahmoud E.gadallah., "Classification of EEG signals using different feature extraction techniques for mental-task BCI," Ain Shams university, Cairo, Egypt.