B-Spline Mutual Information Independent Component Analysis

Janett Walters-Williams†, and Yan Li††

†School of Computing and IT, University of Technology, Jamaica, 237 Old Hope Road, Kingston 6, Jamaica
††Department of Mathematics and Computing, University of Southern Queensland, Toowoomba, Australia

Summary

Mutual Information is one of the most natural criteria when developing independent component analysis (ICA). Although utilized to some level it has always been difficult to calculate. We present a new algorithm which utilizes a contrast function related to Mutual Information based on B-Spline functions. We compared this algorithm with benchmarked ICA algorithms such as FastICA, Infomax and JADE and found it to be very favourable with them in performance.

Key words:
B-Spline, Mutual Information, Independent Component Analysis, Electroencephalogram

1. Introduction

Of all the body's organs the brain is the most mysterious. Main studies of this organ lay in the electrical activity of the firing neurons which cannot be directly investigated by any Magnetic Resonance Imaging (MRI) procedure. Analysis of the brain is now an increasingly important area of research for understanding and modeling it for medical diagnosis and treatment, especially for developing automated patient monitoring and computer-aided diagnosis. The Independent Component Analysis (ICA) approach is one exploratory method which has proven to be reasonably fit for the underlying assumption in Electroencephalogram (EEG), Event Related Potentials (ERP), Magnetoencephalography (MEG), Positron Emission Tomography (PET), functional Magnetic Resonance Imaging (fMRI) and Single Photon Emission Computed Tomography (SPECT). It is also effective in removing artifacts due to volume conduction through cerebrospinal fluid, skull, scalp and experimental imperfections.

ICA is a powerful technique, closely related to blind source separation (BSS), which aims to recover blind sources since the 1980s [16]. It can be described as the problem of recovering a latent random vector $S = [S_1 ... S_m]^T$ of independently distributed components from the observation vector $X = [X_1 ... X_m]^T$ modeled as:

$$X = AS + N, \qquad (1)$$

where A is an unknown m x m matrix called the mixing matrix and N is noise. The problem is to determine A and recover the independent components S knowing only X, as there is no knowledge on the sources distribution and the mixing matrix (this is why the separation method is called blind). The approach estimates A using the separation/demixing matrix W which is the inverse of A, i.e. $W = A^{-1}$ resulting in the equation below which produces the independent components (ICs), u:

$$u = WX = WAS, \qquad (2)$$

ICA is therefore firstly concerned with finding W.

Since ICA was motivated by neurophysiological problems in the early 1980s [16] there have been many methods proposed to estimate W. Most of them are based on estimating equations deduced from some contrast functions, such as maximum likelihood estimate (MLE) [23,30], minimizing mutual information of WX [8] by parametrizing each distribution definitely, minimizing higher-order correlation between WX' components [7], using entropy based calculation [3,13] and maximizing the non-gaussianity of WX's components [13]. Recently, some nonparametric methods to estimate W have appeared. For example, Bach and Jordan [1] minimized a kernel canonical correlation (KCCA), Hastie and Tibshirani [10] proposed a MLE by using Spline-based density approximations and Miller and Fisher [28] proposed using a neighbourhood density estimator.

ICA is a viable tool for analyzing the activity of EEG signals producing outputs which are as independent as possible. In this methodology therefore there is a need to exploit an independence measure. Mutual Information (MI) is such a measure and is considered to be the best choice to measure the independence of the estimated sources [21,35] and a good contrast function [8,14]. MI, however, is 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 [21] and have been very successful [8]. One of the main differences among the various MI-based ICA methods is the way in which this estimation is dealt with. For example the ICA method using minimum mutual information (MMI) was

constructed by Shannon's mutual information where the difference between the marginal entropy and the joint entropy of different information sources was accumulated. The one difficulty of this method however was the estimation of marginal entropy. Comon approximated the output marginal probability density function by applying the truncated polynomial expansion [8]. Another MMI method was proposed by Xu *et al.* [37] which prevented the polynomial expansion through approximating the Kullback-Leibler divergence using the Cauchy-Schwartz inequality. The ICA estimation was performed by using the Parzen window based distribution. Boscolo *et al.* [6] also proposed an ICA algorithm where the MI between the reconstructed signals was minimized. Using nonparametric kernel density technique, this algorithm was carried out by estimating the unknown probability density functions of the source signals and finding the unknown mixing matrix. Although all these algorithms existed, Hyvarinen [14] stated that in its present use these algorithms were far from optimal as far as robustness and asymptotic variance were concerned. These algorithms were also sensitive to artifacts.

Recently, B-Spline has been widely used in the estimation of MI. Klien *et al.* [22] in their research found that the maximisation of MI, in combination with a deformation field parameterised by cubic B-Spline, has been shown to be robust and accurate in many applications. In 2003 Rueckert *et al.* [34] presented MI schemes using B-Spline to help represent the deformation field. Daub *et al.* [9] went on to actually estimate MI using B-Spline. They found that since MI is defined in "terms of discrete variables" B-Spline can be used to perform a numerical estimation to give more accurate estimation of probabilities. Their algorithm avoided the time-consuming numerical integration steps for which kernel density estimators (KDE) are noted. They stated that B-Spline estimated MI outperforms all the other known algorithms for gene expression analysed. Rossi *et al.* [33] stated that B-Spline estimated MI reduces feature selection. It is a good choice as it is non-parametric and model-independent. The other newest form of estimating MI – K nearest neighbor (KNN) has a total complexity of $O(N^3P^2)$ while B-Spline worst-case complexity is still less at $O(N^3P)$ thus having a smaller computation time. They also stated that B-Spline does not require samples that grow exponentially to provide accurate estimations when estimating joint densities, unlike other estimation methods.

In this paper, we propose a new method for denoising EEG signal ICA methodology. The basic idea is to use B-Spline functions to define a MI contrast function to be utilized in an ICA. Thus, this method is called B-Spline Mutual Information ICA (BMICA). Being an ICA method BMICA will not only decorrelate signals but also reduce higher-order statistical dependencies [33]. The method will overcome (i) estimating joint densities dependent on samples that grow exponentially to provide accurate estimations and (ii) the choice-of-origin problem by smoothing the effect of transition of data points between bins due to shifts in origin.

The outline of this paper is as follows. In Section 2 we discuss the signals for which this algorithm is designed and how it relates to the ICA model. In Section 3 we discuss the mutual information estimator. Description of the proposed ICA is in Section 4. Section 5 has an experimental study on the algorithm and concluding remarks are given in Section 6.

2. EEG Signals

The nervous system communicates by trains of electric impulses. When the neurons of the brain process information they do so by changing the flow of electrical current across their membranes. These changing currents (potential) generate electric fields that can be recorded from the scalp. Researchers have been interested in these electrical potentials but they can only be received by direct measurement. This requires a patient to under-go surgery for electrodes to be placed inside the head. This is not acceptable because of the risk to the patient. Another possibility for measurement is to record the potentials on the scalp using an electroencephalograph. Here the potentials are collected from tens or hundreds of electrodes, positioned in pairs, on different locations on the surface of the head. These potentials are simultaneously tested through individuals' amplifiers or channels. Recordings from anyone channel does not represent total discharge from a single underlying segment of the brain but represent the difference in potential between two (2) areas under each pair of electrodes. These recordings are called electroencephalogram (EEG) signals.

EEG signals have been collected so that researchers can try to understand the brain. These signals are being used for clinical and research purposes. In neurology EEG is used to:

(i) Diagnose epilepsy and see what type of seizures is occurring.
(ii) Produce the most useful and important test in confirming a diagnosis of epilepsy.
(iii) Check for problems with loss of consciousness or dementia.
(iv) Help find out a person's chance of recovery after a change in consciousness.
(v) Find out if a person who is in a coma is brain-dead.
(vi) Study sleep disorders, such as narcolepsy.
(vii) Watch brain activity while a person is receiving general anesthesia during brain surgery.
(viii) Detect brain tumors or sensory deficits

In cognitive neuroscience it is used to investigate the neural correlates of mental activities from low-level perceptual and motor processes to high-order cognition

(attention, memory, reading). These signals must therefore present a true and clear picture about brain activities. EEG signals however are highly attenuated and mixed, since they originate from the activity of thousands of neurons, which passes through different tissue layers before reaching the recording electrodes. These neurons may be outside the brain as they also communicate using electrical impulses. These non-cerebral impulses are produced from:

(i) Eye movements and blinking - Electrooculogram (EOG)
(ii) Cardiac Movements - Cardiograph (ECG/ EKG)
(iii) Muscle Movements - Electromyogram (EMG) and
(iv) Chewing and Sucking Movement – Glossokinetic

The EEG signals can also be a mixture including non-biological impulses from:

(i) The electroencephalography which can generate electrode pops
(ii) Poor grounding of power lines and
(iii) Intravenous (IV) drips

These non-cerebral impulses or artifacts (noise) contaminate the EEG signals making detection more difficult because they introduce spikes which can be confused with neurological rhythms. They also mimic EEG signals, overlaying these signals resulting in signal distortion. The recorded EEG signals can therefore be described mathematically as:

$$E(t) = S(t) + N(t), \qquad (3)$$

where S is pure EEG signal, N is the noise and E represents the recorded signal. Correct analysis is almost impossible, resulting in misdiagnosis in the case of some patients. Noise ($N(t)$) must be eliminated or attenuated leaving only the pure EEG signals (Fig 1). Nowadays, there are investigations on how to remove the noise and Independent Component Analysis is one such method.
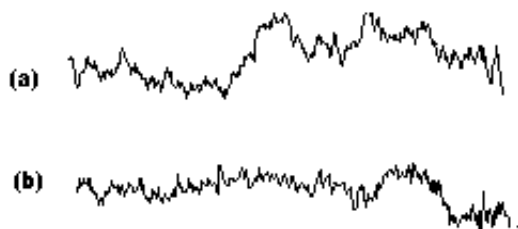

Fig 1 (a) EEG with noise
(b) EEG without noise

Eq (3) can be equated to the ICA definition in Eq. (1) where the rows of the input matrix X are the EEG signals recorded at different electrodes, the rows of the output data matrix (where u is the estimated ICs, W the separation matrix, and X the observation vector from Eq (2)) are time courses of activation of the ICA components, and the columns of the inverse matrix W^{-1} give the projection strengths of the respective components onto the scalp sensors. The scalp topographies of the components provide information about the location of the sources e.g. eye activity should project mainly to frontal sites. "Corrected" EEG signals can then be derived as:

$$X' = (W)^{-1} u', \qquad (4)$$

where u' is the matrix of activation waveforms, u, with rows representing artifactual components set to zero. The rank of corrected EEG data is less than that of the original data. For this solution to work however the assumption is made that the components are statistically independent, while the mixture is not. This is plausible since biological areas are spatially distinct and generate a specific activation; they however correlate in their flow of information [12]. ICA algorithms are suitable for denoising EEG signals because:

(i) the signals recorded are the combination of temporal ICs arising from spatially fixed sources and
(ii) the signals tend to be transient (localized in time), restricted to certain ranges of temporal and spatial frequencies (localized in scale) and prominent over certain scalp regions (localized in space) [27].

3. Mutual Information

Mutual Information (MI), also known as the archaic term transinformation, was first introduced in classical information theory by Shannon in 1948. It is considered to be a non parametric measure of relevance that measures the mutual dependence of two variables i.e. it looks at the amount of uncertainty that is lost from one variable when the other is known. MI, represented as $I(X:Y)$, in truth measures the reduction in uncertainty in X which results from knowing Y i.e. it indicates how much information Y conveys about X and is defined as:

$$I(X,Y) = \sum_{i,j} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i) p(y_j)} \qquad (5)$$

Hyvarinen et al. [16] stated that the use of MI produces a very realistic approach to denoising, as it does not assume anything about the data. It defines "ICA as a linear

decomposition that minimizes that dependence measure" with respect to the separating matrix *W*. MI however is unknown, so in practice it must be substituted by an estimator. The estimation of MI requires the estimation of the joint density in Eq. (5), which demands an duly large amount of data for an acceptable accuracy – a problem. Joint density can be avoided however by expressing MI in term of entropy as:

$$I(X,Y) = H(X) + H(Y) - H(X,Y) \qquad (6)$$

where

$$
\begin{aligned}
H(X) &= -\sum_i p(x_i)\log p(x_i) \\
H(X,Y) &= -\sum_{i,j} p(x_i,y_j)\log p(x_i,y_j)
\end{aligned}
\qquad (7)
$$

Eq. (6) contains the term $-H(X,\ Y)$, which means that maximizing MI is related to minimizing joint entropy. MI is better than joint entropy however because it includes the marginal entropies $H(X)$ and $H(Y)$ [18]. When using a definition of MI based on entropy different definitions of entropy can be chosen. These have resulted in two basic categories:

(i) Parametric which include Bayesian, Edgeworth, maximum likelihood (ML), and least square estimators and

(ii) Nonparametric which includes histogram based, adaptive partitioning of the XY plane, kernel density, B-Spline, nearest neighbour and wavelet density estimators.

There have been many MI estimators in ICA literature which are very powerful yet difficult to estimate resulting in unreliable, noisy and even bias estimation. Most of these algorithms have been based on cumulant expansions because of their ease of use. [21]. Krishnaveni *et al.* found that a MI estimated using k-nearest neighbor distance outperforms many of the known ICA algorithms. B-Spline estimators according to our previous research [36] have been shown to be one of the best nonparametric approaches, second to only wavelet density estimators thus better than Krishnaveni's nearest neighbor (KNN) estimator.

3.1 B-Spline

B-Spline is a flexible mathematical formulation for curve fitting due to a number of desirable properties [30]. Under the smoothness constraint, B-Spline gives the "optimal" curve fitting in terms of minimum mean-square error [31,26]. A 2D B-spline curve can be defined mathematically as:

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} f(t) \\ g(t) \end{pmatrix} = \sum_{i=1}^{m-1} B_{j,k}(t) \begin{pmatrix} \tilde{x}_i \\ \tilde{y}_i \end{pmatrix}, \quad t_{\min} \le t < t_{\max} \qquad (8)$$

were $\left\{ \begin{pmatrix} \tilde{x}_i \\ \tilde{y}_i \end{pmatrix} \; i=1,2, \; . \; . \; . \; m\text{-}1 \right\}$ are *m-1* control points assigned from data samples. *t* is a parameter and is in the range of maximum and minimum values of the element in a knot vector. A knot vector, $t_1, t_2, \ldots, t_{k+(m-1)}$, is specified for giving a number of control points *m-1* and B-spline order *k*. It is necessary that $t_i \le t_{i+1}$, for all *i*. For an open curve, open uniform knot vector defined as:

$$t_i = \begin{cases} 0 & if & i < k \\ i-k+1 & if & k \le i \le m-1 \\ m-1-k+2 & if & i > m-1 \end{cases} \qquad (9)$$

is used. The $B_{j,k}(t)$ basis functions are of order ($2 \le k \le m\text{-}1$), depending only on the value of *k* and the values in the knot vector and defined recursively as:

$$B_{i0}(t) = \begin{cases} 1 & if & t_i \le z \prec t_{i+1} \\ 0 & & otherwise \end{cases} \qquad (10)$$

$$B_{i,k}(t) = B_{i,k-1}(t)\frac{t-t_i}{t_{i+k-1}-t_i} + B_{i+1,k-1}(t)\frac{t_{i+k+1}-t}{t_{i+k+1}-t_{i+1}} \qquad (11)$$

Given a pair of signals *sx* and *sy* with expression values $\{(xi, yi), i = 1, \ldots, m\}$, *m-1* control points $\{(\tilde{x}_i\ \tilde{y}_i)\ i = 1, \ldots, m\text{-}1$ selected from $\{(xi, yi), j = 1, \ldots, m\}$, a knot vector, $t1, t2, \ldots, tk+(m+1)$, and the order of *k*, the plotted pattern can be modeled by Eq. (8). In Eq. (8), $f(t)$ and $g(t)$ are the *x* and *y* components of a point on the curve, *t* is a parameter in the parametric representation of the curve.

Recently, B-Spline has been widely used in microarray data analysis, including inference of genetic networks, estimation of MI, and modeling of time-series gene expression data [2,5,9,11,16,25,26,32]. In numerical estimation of MI from continuous microarray data [9], a generalized indicator function based on B-Spline has been proposed to get more accurate estimation of probabilities.

3.2 MI Estimator

Since MI estimation using joint density is a problem MI can be defined as seen in Eq. (6) and Eq. (7) using entropy, *H(X),* defined according to Shannon. Here we calculate the entropy of the sequence by using probability distribution functions (pdfs). In our design we have defined our pdf using a B-Spline calculation resulting in the algorithm below.

Algorithm 1: Entropy for variable x

Input: Data vector x

1.   Generate $B(x)$
(a)   Determine the validity of variable x
(b)   Calculate D_1 based on Cheney and Kincaid (1994)
(c)   Determine D_i with

$$s(x_i) = D_i B_{i-2}(x_i) + D_{i+1} B_{i-1}(x_i) \qquad (12)$$

where

$$y_i = D_i \frac{h_i}{h_i + h_{i-1}} + D_{i+1} \frac{h_{i-1}}{h_i + h_{i-1}} \qquad (13)$$

(d)   Determine data interval for x
(e)   Calculate $B(x)$ with

$$B(x) = \sum_{i=1}^{n+1} D_i B_{i-k}^k(x) \qquad (14)$$

2.   Sum all $B(x)$ and determine $P(x)$ from

$$p(x_i) = \frac{1}{N} \sum_{u=1}^{N} \tilde{B}_{i,k}(x_u) \qquad (15)$$

3.   Determine entropy $H(x)$ according to Eq (7).

Output: H(x)

Joint entropy is calculated:

$$H(X,Y) = \sum_{i=1, j=1}^{M_X, M_Y} p(x_i, y_j) \log p(x_i, y_j) \qquad (16)$$

where $p(x_i, y_i)$ is the joint probability defined as:

$$p(x_i, y_j) = \frac{1}{N} \sum_{u=1}^{N} B(x_u) \times B(y_u) \qquad (17)$$

Joint entropy can be linked to Eq. (4) as:

$$H(u) = H(X) + \log |\det W| \qquad (18)$$

MI is then determined according to Eq. (6)

4. ICA Algorithm

In this section we describe our approach to ICA development based on the methodology for BSpline MI estimation given in Daub *et al.* [9]. We start with the preprocessing procedure

4.1 Preprocessing

Prewhitening is a popularly used preprocessing technique in ICA literature which speeds algorithms up substantially. For example many famous ICA algorithms such as FastICA, and JADE, have used this pre-processing technique. It is really the actual whitening of a signal ahead of some processing i.e. removing bias and unwanted autocorrelations derived from both internal and external processes, so that all parts of the signal enter the next stage of processing on a level playing field. This amounts to a principal component analysis (PCA) of the observations. The removal of these autocorrelations is necessary to the interpretation of other potential relationships. This technique is done before estimating W from Eq. (2) and can be preformed as below:

1.   Subtract mean value $E[X]$ from observed signal

$$\tilde{X} = X - E[X] \qquad (19)$$

2.   Whiten results via eigenvalue decomposition of the covariance matrix:

$$VDV^T = E[\hat{X}\hat{X}^T] \qquad (20)$$

where V is the matrix of orthogonal eigenvectors and D is a diagonal matrix with the corresponding eigenvaluves. Whitening is done by multiplication with the transformation matrix P.

$$P = VD^{-\frac{1}{2}}V^T \qquad (21)$$

$$\tilde{X} = P\hat{X} \qquad (22)$$

The matrix for extracting the independent components from $\tilde{X}$ is $\tilde{W}$ where $W = \tilde{W}P$.

4.2 Algorithm

The most important step is separating the prewhitened signal into ICs. In our algorithm the ICA is performed using a B-Spline defined MI contrast function. Our ICA algorithm is a fixed point algorithm because these algorithms allow for fast convergence of the nongaussianity criterion. Unlike the gradient descent method, there is no need for adjustment of learning steps or other adjustable parameters and the rate of convergence is therefore fixed without regard to the changing environment. Fixed-point algorithms also tend to be much more stable than other algorithms [29]. The algorithm is defined as:

Algorithm 2: ICA algorithm

Input: Data vector X

1.   Preprocess X to produce z
2.   Choose an initial random separating matrix B
3.   For $i = 1,...$ *until convergence* :
   (i)   Determine the whitened signals based on:

$$y = z' \times B \qquad (23)$$

   (ii)   Determine I running Algorithm 1 on y
   (iii)   Update B using:

$$B = (zg(y)'/m - \sum(1 - g(y)^2)' \times I)/m \qquad (24)$$

where

$$g(y) = \tanh(y) \qquad (25)$$

   (iv)   Do a symmetric orthogonalization of B by
$$B = (BB^T)^{-\frac{1}{2}} B \qquad (26)$$

4.   Compute W using B
5.   Determine u by Eq (2)

   Output: *u*

5. Experiment and Discussion

5.1 Experimental Data

In this section we have applied the proposed method to actual EEG data in order to confirm the practical effectiveness of the method. The data have been collected from to sites:
   (i)   http://sccn.ucsd.edu/~arno/fam2data/publicly_ava ilable_EEG_data.html. All data are real comprised of EEG signals from both human and animals. Data were of different types.
      (a)   Data set acquired is a collection of 32-channel data from one male subject who performed a visual task. Fig. 5 shows 10 signals from this dataset.
      (b)   Human data based on five disabled and four healthy subjects. The disabled subjects (1-5) were all wheelchair-bound but had varying communication and limb muscle control abilities. The four healthy subjects (6-9) were all male PhD students, age 30 who had no known neurological deficits. Signals were recorded at 2048 Hz sampling rate from 32 electrodes placed at the standard positions of the 10-20 international system.
      (c)   Data set is a collection of 32-channel data from 14 subjects (7 males, 7 females) who performed a go-nogo categorization task and a go-no recognition task on natural photographs presented very briefly (20 ms). Each subject responded to a total of 2500 trials. The data is CZ referenced and is sampled at 1000 Hz.
      (d)   Five data sets containing quasi-stationary, noise-free EEG signals both in normal and epileptic subjects. Each data set contains 100 single channel EEG segments of 23.6 sec duration.
   (ii)   http://www.cs.tut.fi/~gomezher/projects/eeg/datab ases.htm. Data here contains
      (a)   Two EEG recordings (linked-mastoids reference) from a healthy 27-year-old male in which the subject was asked to intentionally generate artifacts in the EEG
      (b)   Two 35 years-old males, where the data was collected from 21 scalp electrodes placed according to the international 10-20 System with addition electrodes T1 and T2 on the temporal region. The sampling frequency was 250 Hz and an average reference montage was used. The electrocardiogram (ECG) for each patient was also simultaneously acquired and is available in channel 22 of each recording.

These two sites produce real signals of different sizes however all were 2D signals. The length of all signals, N, was truncated to a length equivalent to powers of twos i.e. 2^x.

5.2 Results

In the previous section we have described an ICA algorithm where the contrast function is motivated by B-Spline functions. In this section we investigate its performance. There are different means to access the separation quality performed by ICA methods; however the performance measures used throughout this section will be:
   (i)   the Mean Square Error (MSE),
   (ii)   the Peak Signal to Noise Ratio (PSNR),
   (iii)   the Signal to Distortion Ratio (SDR),
   (iv)   the Signal to Noise Ratio (SNR),
   (v)   the Signal to Interference Ratio (SIR) and
   (vi)   the Amari Performance Index
Comparison with two categories of benchmark ICAs will be provided namely:

(i) fixed-point - FastICA[15], EFICA[20] and Pearson_ICA [19]

(ii) non fixed-point - Infomax[3], SOBI[4], and JADE [7]

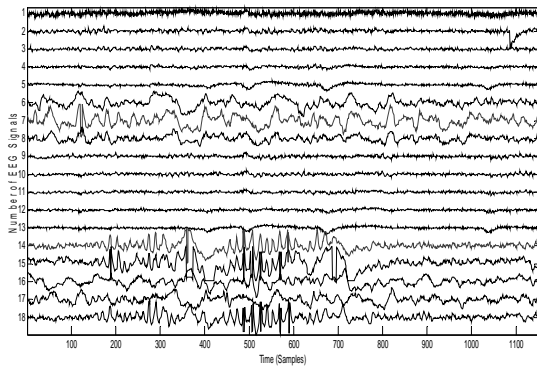For these algorithms, we used the publicly available Matlab codes.

Fig 2: Sample of Raw EEG Signals

Experiments were conducted using the above mentioned signals, in Matlab 7.8.0 (R2009) on a laptop with AMD Athlon 64x2 Dual-core Processor 1.80GHz. Fig 2 shows one mixed EEG signal set where there are overlays in signals 2, 6-8 and 14-18. Fig 3 shows the same signal set after applying our algorithm showing that the overlays have been minimized – noise has been removed.
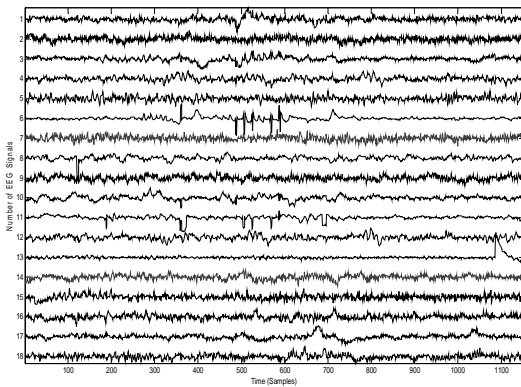

Fig 3: EEG Signals after denoised with New ICA algorithm

5.2.1 Noise/Signal Measures

The MSE measures the average of the square of the "error" which is the amount by which the estimator differs from the quantity to be estimated. Mathematically it is defined as:

$$MSE = \frac{1}{N} \sum_{y=1}^{N} [I(x,y) - I'(x,y)]^2 . \tag{27}$$

The difference occurs because of randomness or because the estimator doesn't account for information that could produce a more accurate estimate. For a perfect fit, $I(x,y) = I'(x,y)$ and $MSE = 0$; so, the MSE index ranges from 0 to infinity, with 0 corresponding to the ideal. The smaller the MSE therefore the closer the estimator is to the actual data. BMICA was compared to both categories in Table 1 and

Table 1: MSE comparison with fixed-point algorithms

BMICA	FASTICA	PEARSON	EFICA
1.66E+03	1.67E+03	1.68E+03	1.69E+03
1.27E+03	1.30E+03	1.27E+03	1.28E+03
1.16E+03	1.17E+03	1.21E+03	1.21E+03
1.81E+03	2.01E+03	2.02E+03	2.00E+03
1.11E+03	1.12E+03	1.12E+03	1.11E+03
1.17E+03	1.53E+03	1.55E+03	1.55E+03
3.14E+03	3.12E+03	3.11E+03	3.11E+03
1.28E+04	1.29E+04	1.29E+04	1.28E+04
4.91E+05	4.91E+05	4.91E+05	4.92E+05
4.63E+05	4.63E+05	4.63E+05	4.63E+05
3.30E+05	3.30E+05	3.30E+05	3.30E+05
9.41E+02	9.63E+02	9.22E+02	9.62E+02
8.79E+02	9.18E+02	9.52E+02	9.82E+02
7.51E+02	7.73E+02	7.57E+02	7.86E+02
6.70E+02	6.65E+02	6.68E+02	6.68E+02
7.09E+02	7.04E+02	7.25E+02	7.17E+02
5.97E+02	5.92E+02	5.95E+02	5.85E+02
4.59E+02	4.62E+02	4.70E+02	4.70E+02
7.30E+04	**7.30E+04**	**7.30E+04**	**7.30E+04**

Table 2. Table 1 shows that on average all tested fixed-point algorithms have similar MSE. Further investigations show that BMICA has the lowest MSE 75% of the time when there are differences in the MSE. Examination of Table 2 shows Infomax to have the lowest MSE on average. BMICA performed best in 10 of 15 experiments of the other algorithms

Table 2: MSE comparison with non fixed-point algorithms

BMICA	SOBI	INFOMAX	JADE
1.66E+03	1.67E+03	1.61E+03	1.66E+03
1.27E+03	1.29E+03	1.26E+03	1.31E+03
1.16E+03	1.19E+03	1.12E+03	1.22E+03
1.81E+03	1.01E+03	2.40E+03	2.02E+03
1.11E+03	1.11E+03	1.06E+03	1.08E+03
1.17E+03	1.54E+03	1.49E+03	1.54E+03
3.14E+03	2.13E+03	3.03E+03	3.13E+03
1.28E+04	1.29E+04	1.27E+04	1.29E+04
4.91E+05	4.91E+05	4.91E+05	4.91E+05
4.63E+05	4.63E+05	4.62E+05	4.63E+05
3.30E+05	3.30E+05	3.29E+05	3.30E+05
9.41E+02	9.56E+02	8.96E+02	9.43E+02
8.79E+02	9.18E+02	8.60E+02	9.57E+02
7.51E+02	7.65E+02	7.21E+02	7.58E+02
6.70E+02	6.65E+02	6.22E+02	6.69E+02
7.09E+02	7.16E+02	6.75E+02	7.23E+02
5.97E+02	5.93E+02	5.56E+02	5.85E+02
4.59E+02	4.66E+02	4.32E+02	4.70E+02
7.30E+04	**7.29E+04**	**7.28E+04**	**7.30E+04**

PSNR is the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. Mathematically it is defined as:

$$PSNR = 10 \times \log_{10}(\frac{MAX^2}{MSE}).$$ (28)

Because many signals have a very wide dynamic range, PSNR is usually expressed in terms of the logarithmic decibel scale. In this research MAX takes the value of 255. Unlike MSE which represents the cumulative squared error between the denoised and mixed signal, PSNR represents a measure of the peak error i.e. when the two signals are identical the MSE will be equal to zero, resulting in an infinite PSNR.

Table 3: First PSNR comparison with non fixed-point algorithms

NEWICA	SOBI	INFOMAX	JADE
-8.7827	-8.7834	-8.7770	-8.7831
-8.5257	-8.5209	-8.5146	-8.5212
7.0480	7.0340	7.0968	7.0289
-7.0496	-7.0498	-7.0444	-7.0489

The higher the PSNR therefore, the better the quality of the reconstructed signal i.e. a higher PSNR indicates that the reconstruction is of a higher quality and therefore the algorithm is considered good. Examination of Table 3, and Table 4 show that Infomax is the algorithm that has the highest PSNR then BMICA. Table 5 shows BMICA presenting more PSNR values which are higher than the others. BMICA therefore presents more signal than noise in its denoised results than SOBI, JADE FastICA, EFICA and Pearson_ICA resulting in the second best performance. This follows as this is the same behavior with the MSE investigations.

Table 4: Second PSNR comparison with non fixed-point algorithms

BMICA	SOBI	INFOMAX	JADE
15.9207	15.9025	16.0672	15.9205
17.0666	17.0139	17.1141	16.9518
18.3936	18.3244	18.6072	18.3857
18.6811	18.502	18.7848	18.3204
19.3730	19.2944	19.5525	19.3362
17.4924	17.3896	17.6247	17.3567
19.8679	19.9030	20.1919	19.8743
19.6233	19.5816	19.8367	19.5386
17.6643	17.6829	17.8611	17.7787
13.1659	13.1822	13.3197	13.1756
16.2423	16.2448	16.4067	16.2435
20.3691	20.4020	20.6802	20.4578
21.5157	21.4500	21.7757	21.4110
30.0699	30.2565	30.8013	30.1489
29.6448	29.7112	30.1369	29.6404

SNR, normally expressed in decibels, is refers to how much signal and how much noise is present regarding just about anything and everything i.e. the ratio compares the level of a desired signal to the level of background noise. It is expressed mathematically as:

$$SNR(dB) = 20\log_{10} \frac{\sum_{n=0}^{N} s^2(n)}{\sum_{n=0}^{N} x^2(n)}$$ (29)

The greater the ratio, evidenced by a larger number, the less noise and the more easily it can be filtered out. A SNR of 0 however means that noise and signal levels are the same. Although signals contain non-random intelligence and can be isolated and separated, with a 0 SNR, it would be extremely difficult to isolate the signal in real time. On average in Table 6 BMICA has the second highest SNR on average.

Table 5:   PSNR comparison with fixed-point algorithms

BMICA	FASTICA	PEARSON	EFICA
15.9207	15.8890	15.8898	15.8526
17.0666	16.9961	17.1005	17.0517
18.3936	18.2954	18.4831	18.2982
18.6811	18.5034	18.3444	18.2114
19.3730	19.2487	19.3375	19.1739
17.4924	17.4529	17.2981	17.2988
19.8679	19.9043	19.8846	19.8811
45.5542	45.0919	45.0751	45.1186
19.6233	19.6568	19.5265	19.5777
17.6643	17.6427	17.6560	17.6769
13.1659	13.1872	13.1976	13.2007
16.2423	16.2747	16.2351	16.2377
20.3691	20.4044	20.3862	20.4556
21.5157	21.4804	21.4129	21.4104
30.0699	30.1499	30.2253	30.1252
29.6448	29.7131	29.6213	29.6117

Examination of the table shows that BMICA when compared to the other fixed-point algorithms has the highest SNR. This shows that of the seven algorithms only Infomax has a higher SNR thus less obtrusive background noise and better signal performance.

5.2.2 Separation Accuracy Measures

The most widely used measure for assessing the accuracy of the estimated mixing matrix is the Amari performance index defined as:

$$P_{err} = \frac{1}{2m} \sum_{i,j=1}^{m} \left(\frac{|p_{ij}|}{\max_k |p_{ik}|} + \frac{|p_{ij}|}{\max_k |p_{kj}|} \right) - 1 \qquad (31)$$

where $p_{ij} = (BA)_{ij}$. It assesses the quality of the de-mixing matrix **W** for separating observations generated by the

Table 6:   SNR comparison with non fixed-point algorithms

BMICA	SOBI	INFOMAX	JADE
2.26E-02	1.80E-03	1.14E-01	-9.30E-03
6.69E-02	8.86E-04	1.03E-01	-2.12E-02
9.60E-02	8.96E-04	1.93E-01	1.18E-01
2.38E-01	6.15E-04	2.00E-01	-2.14E-01
1.16E-01	1.93E-02	1.88E-01	2.21E-02
3.16E-02	7.04E-05	1.25E-01	7.50E-03
5.83E-02	-3.02E-04	1.60E-01	1.20E-03
8.90E-03	2.62E-04	8.60E-02	-2.81E-02
5.92E-05	-9.32E-07	5.30E-03	7.23E-05
1.05E-02	2.37E-04	3.86E-02	3.20E-03
1.35E-04	2.30E-08	3.60E-03	5.89E-04
4.20E-03	3.53E-04	8.74E-02	7.60E-03
5.82E-02	-7.14E-05	1.77E-01	-4.68E-02
5.47E-02	**1.85E-03**	**1.14E-01**	**-1.22E-02**

mixing matrix **A.** When the separation is perfect, the Amari index is equal to zero. In the worst case, i.e. when the estimated sources contain the same proportion of each original source signal, the Amari index is equal to $m/2-1$. The Amari indexes obtained for the different algorithms and for different sample sizes are presented in Fig 4 and Fig 5.   From observation it can be seen that BMICA has an Amari separation pattern similar to all the other algorithms, i.e. all algorithms behave the same as size of sample increase or decrease. Because of the Amari results other methods to determine separation accuracy was used, resulting in SDR and SIR
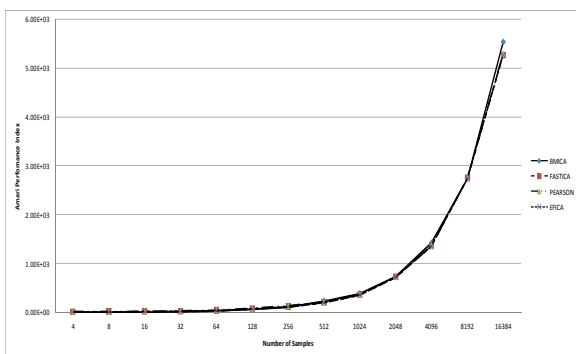

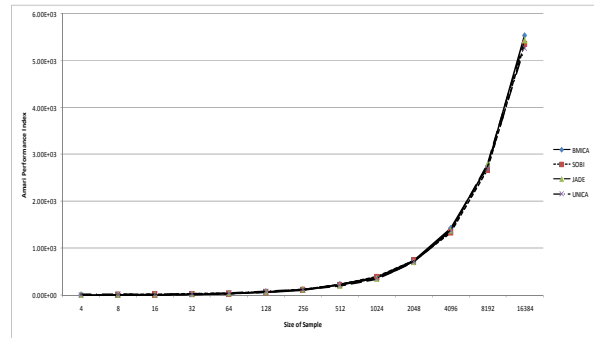Fig 4:   Amari Index for fixed-point algorithms

.

Fig 5:   Amari Index for non fixed-point algorithms

How accurate the separation of an ICA algorithm in terms of the signals can be calculated by the total SDR which is defined as:

$$SDR(x_i, y_i) = \frac{\sum_{n=1}^{L} x_i(n)^2}{\sum_{n=1}^{L} (y_i(n) - x_i(n))^2} \qquad i = 1,...m, \qquad (30)$$

Table 7:   SNR comparison with fixed-point algorithms

BMICA	FASTICA	PEARSON	EFICA
2.26E-02	7.37E-02	-7.73E-02	-3.43E-02
6.69E-02	-4.05E-02	8.92E-02	2.99E-02
9.60E-02	-7.84E-02	1.25E-01	-5.81E-02
2.38E-01	2.00E-03	-1.43E-01	-2.86E-01
1.16E-01	-5.47E-02	7.69E-02	-1.25E-01
-1.24E-02	6.94E-02	-7.17E-02	7.82E-02
3.16E-02	2.80E-03	-2.56E-02	-2.76E-02
5.83E-02	9.86E-02	-7.51E-02	1.92E-02
-4.01E-02	-1.58E-02	-1.90E-03	1.32E-02
8.90E-03	2.85E-02	3.87E-02	4.80E-03
5.92E-05	1.66E-04	-6.12E-05	-5.30E-03
-4.90E-03	-1.70E-03	9.82E-07	-2.72E-05
1.05E-02	-2.50E-03	-6.20E-03	2.39E-02
1.35E-04	8.68E-04	-8.42E-05	3.27E-04
4.20E-03	4.47E-02	1.76E-02	-2.46E-02
-3.61E-02	2.50E-03	-1.20E-02	9.87E-02
5.82E-02	1.37E-02	8.97E-04	-3.53E-02
3.63E-02	**8.43E-03**	**-3.82E-03**	**-1.92E-02**

where $x_i(n)$ is the original source signal and $y_i(n)$ is the reconstructed signal. Consider Table 8; this shows that of the four fixed-point algorithms BMICA has the highest average SDR indicating that BMICA performed

Table 8:   SDR comparison with fixed-point algorithms

BMICA ▼	FASTICA ▼	PEARSON ▼	EFICA ▼
3.46E-01	1.02E-02	3.19E-01	3.38E-01
3.24E-01	3.28E-01	3.19E-01	3.21E-01
5.63E+00	5.68E+00	5.57E+00	5.37E+00
5.46E+00	5.56E+00	5.49E+00	5.55E+00
1.15E-01	1.12E-01	1.14E-01	1.12E-01
7.56E-02	7.47E-02	7.53E-02	7.33E-02
1.55E-01	1.43E-01	1.49E-01	1.41E-01
1.86E-01	3.89E-04	1.69E-01	1.63E-01
1.16E-01	1.10E-01	1.14E-01	1.10E-01
1.75E-01	1.31E-01	1.29E-01	1.30E-01
3.92E-01	3.87E-01	3.79E-01	3.80E-01
1.92E-01	1.95E-01	1.96E-01	1.95E-01
1.63E-02	1.63E-02	1.63E-02	1.63E-02
3.10E-01	3.07E-01	3.05E-01	3.01E-01
9.64E-01	**9.32E-01**	**9.53E-01**	**9.43E-01**

Fig 6:   SIR comparison with fixed-point algorithms

the best at separating the EEG from the noise. When BMICA is compared with the three non fixed-point algorithms it was seen that the SDR for our algorithm was superior to the others as Table 9 shows.

Table 9:   SDR comparison with non fixed-point algorithms

BMICA ▼	SOBI ▼	INFOMAX ▼	JADE ▼
3.46E-01	1.43E-04	3.37E-01	3.31E-01
3.24E-01	2.75E-04	3.32E-01	3.26E-01
1.15E-01	9.68E-05	1.14E-01	1.10E-01
7.56E-02	4.63E-05	5.43E-02	7.42E-02
1.55E-01	6.61E-05	1.50E-01	1.54E-01
1.86E-01	5.56E-05	1.78E-01	1.64E-01
1.16E-01	1.83E-07	1.13E-01	1.08E-01
1.75E-01	1.44E-07	1.33E-01	1.27E-01
3.92E-01	4.19E-06	3.85E-01	3.89E-01
1.92E-01	1.81E-04	2.01E-01	1.93E-01
1.63E-02	7.88E-06	1.63E-02	1.63E-02
3.10E-01	-1.16E-05	-2.40E-02	-2.40E-02
2.00E-01	**7.20E-05**	**1.66E-01**	**1.64E-01**

In degenerate demixing, the accuracy of an algorithm cannot be described using only the estimated mixing matrix. In this case it becomes of particular importance to measure how well algorithms estimate the sources with adequate criteria. The most commonly used index to assess the quality of the estimated sources is the SIR calculation using:

$$SIR(dB) = \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{j}^{n} \frac{|p_{ij}|}{\max_{k} |p_{ij}|} - 1 \right) \qquad (32)$$
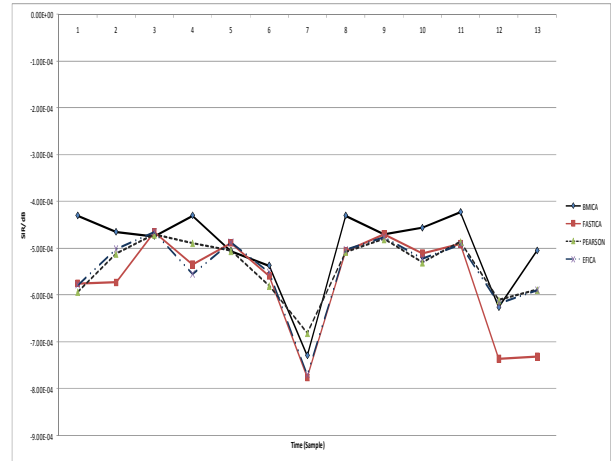
The lower the SIR, the better the achieved separation and a SIR index of 0 implies a perfect separation. Examination of the algorithms' SIR shows that of the seven algorithms BMICA displays the SIR index nearest to 0, implying a good separation as seen in Fig 6 and Fig 7.
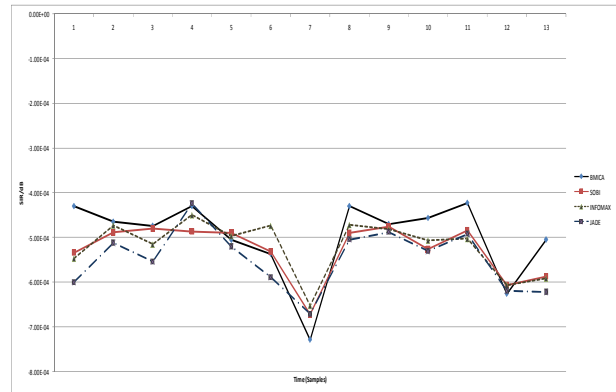

Fig 7:   SIR comparison with non fixed-point algorithms

5.3 Computational Complexity

Although the ultimate goal of a signal separation approach is the quality of such a separation, reflected on the estimated source signals, it is interesting to relate the various ICA approaches from a numeral complexity viewpoint. Here we determine the computational complexity of BMICA and compare it to the other algorithms.

Let N denote the number of samples, and m denote the number of sources. M is the maximum number of iteration. We assume that $m \leq N$.

- Performing preprocessing is O(N)
- Running the Iterations for algorithm is O(M)

- Determining the contrast function is $O(N^2)$ – calculating the loop for the matrix containing N signals is $O(N)$ and determine the MI for each signal of *m* size is $O(N)$
- Determining the matrix to calculate *W* is $O(N/2)$

BMICA therefore has a complexity of $O(M) * O(N^2) * O(N/2)$ resulting in an overall complexity of $O(N^2M)$. This is in line with research [36] where the worst case complexity for B-Spline MI estimators is $O(N^3P)$. When compared to other ICA algorithms it was found that

(i) FastICA and Infomax, both have a complexity on the order of $O(N^3M)$ [34], (worst case B-Spline MI)
(ii) JADE algorithm is on the order of $O(N^4M)$ [34], (greater than worst case) and
(iii) EFICA has a computational complexity only slightly (about three times) higher than that of the standard symmetric FastICA [20],

This shows that BMICA has the best complexity.

6. Conclusions

In this paper, we have presented a new algorithm, BMICA for independent component analysis. Our approach is based on maximizing entropy in the probability distribution functions (pdf) estimation step utilizing B-Spline functions. The commonly used whitening-rotation topology is borrowed from the literature, whereas the criterion used, minimum output mutual information, is considered to be the natural information theoretic measure for ICA.

We have shown the accuracy of our algorithm by comparing it with benchmark ICA algorithms showing that BMICA has

(i) The best computational complexity of $O(N^2M)$,
(ii) The best Separation Accuracy as it has
- the highest SDR
- the lowest SIR and
- similar Amari Performance Index to the other six algorithms
(iii) Relatively good Noise/Signal ratio as it has
- The highest SNR for fixed point algorithms and third overall next to Infomax and SOBI
- the highest PSNR for fixed point algorithms and second overall to Infomax
- the lowest MSE for fixed point algorithms and second overall to Infomax

While these initial results are promising there is room for improvement. Our future work is to optimize parameters for better performance.

References

[1] F. Bach, and M. Jordan, "Kernel independent component analysis, *Journal of Machine Learning Research* 3 vol. 1 no. 48, 2002.

[2] Z. Bar-Joseph, G.K. Gerber, D.K., Gifford, T.S. Jaakkola, and I. Simon, "Continuous representations of time-series gene expression data," *Journal of Computational Biology*, vol. 10 no. 3-4, pp. 341–356. 2003.

[3] A.J. Bell, and T.J. Sejnowski, "An Information-Maximization Approach to Blind Separation and Blind Deconvolution", *Neural Computation,* vol. 7 no. 6, pp. 1129—1159, 1995.

[4] A. Belouchrani, K.A., Meraim, J.F. Cardoso, and E. Moulines,"Second-order blind separation of correlated sources". In *Proceedings International Conference on Digital Signal Processing,* Cyprus, pp. 346–351 1993.

[5] K. Bhasi, A. Forrest, and M. Ramanathan, "SPLINDID: a semiparametric, model-based method for obtaining transcription rates and gene regulation parameters from genomic and proteomic expression profiles," *Bioinformatics*, vol. 21, no. 20, pp. 3873–3879, 2005.

[6] R. Boscolo, H. Pan and V.P. Roychowdhury, "Independent component analysis based on nonparametric density estimation", *IEEE Transactions on Neural Networks,* vol. 15, no. 1, 2004.

[7] J.F. Cardoso, Blind signal separation: statistical principles. *Proceedings of the IEEE vol.* 86 no. 10, pp. 2009-2025, 1998.

[8] P. Comon, "Independent component analysis, a new concept?", *Signal Processing* vol. 36 no. 3, pp. 287-314, 1994.

[9] C.O. Daub R. Steuer, J. Selbig, and S. Kloska, "Estimating mutual information using B-Spline functions—an improved similarity measure for analysing gene expression data," *BMC Bioinformatics*, vol. 5, no. 1, p. 118, 2004.

[10] T. Hastie, and R. Tibshirani, "Independent component analysis through product density estimation, Technical report, Stanford University, 2002.

[11] W. He, "A spline function approach for detecting differentially expressed genes in microarray data analysis," *Bioinformatics*, vol. 20, no. 17, pp. 2954–2963. 2002.

[12] S. Hoffman, and M. Falkenstien, "The Correction of Eye Blink Artefacts in the EEG: A Comparison of a Two Prominent Methods", *PLoS One* vol. 3 no. 8, p. e3004, 2008.

[13] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis *IEEE Transactions on Neural Networks, May 1999 10 (3):* 626 – 634, 1999.

[14] A. Hyvärinen, "Survey on Independent Component Analysis". In *Neural Computing Surveys 2*: 94-128,

1999

[15] A. Hyvärinen and E. Oja, "A Fast Fixed-Point Algorithm for Independent Component Analysis", *Neural Computation,* vol. 9 no. 7, pp. 1483-1492, 1997.

[16] A. Hyvarinen, J. Karhunen and E. Oja, *Independent Component Analysis*, eds. Wiley & Sons. 2001.

[17] S. Imoto, T. Goto, and S. Miyano, "Estimation of genetic networks and functional structures between genes by using Bayesian networks and nonparametric regression," *Pacific Symposium on Biocomputing*, pp. 175–186, 2002.

[18] P. Josien W.. Pluim, J.B. Antoine Maintz and M.A. Viergever, "Mutual information based registration of medical images: a survey", In *IEEE Transactions on Medical Imaging,* vol. Xx, no. Y, 2003.

[19] J. Karvanen, J.,Eriksson, and K.V. Pearson, "System Based Method for Blind Separation", *Proceedings of Second International Workshop on Independent Component Analysis and Blind Signal Separation, Helsinki 2000*, pp. 585—590, 2000.

[20] Z. Koldovský and P. Tichavský, P., "Time-Domain Blind Audio Source Separation Using Advanced ICA Methods", *Proceedings of 8th Annual Conference of the International Speech Communication Association (Interspeech 2007)*, pp. 846-849, 2007.

[21] V. Krishnaveni, S. Jayaraman P.M. Manoj Kumar, K. Shivakumar, and K. Ramadoss, "Comparison of Independent Component Analysis Algorithms for Removal of Ocular Artifacts from Electroencephalogram", *Measurement Science Review, Volume 5*, Section 2, 2005.

[22] S. Klein, M. Marius Staring, and P.W. Pluim, Josien, "Comparison of gradient approximation techniques for optimisation of mutual information in nonrigid registration" *Proc. SPIE* vol. 5747, 192, 2005.

[23] T.W. Lee, M. Girolami, and T. Sejnowski, "Independent component analysis using an extended informax algorithm for mixed subgaussian and supergaussian sources". *Neural Computation* vol. 11 no. 2, pp. 417-441, 1999.

[24] J.H. Lee, H.Y. Jung, T.W. Lee, and S.Y. Lee, "Speech feature extraction using independent component analysis", *Proc. of ICASSP, vol. 3,* pp. 1631-1634, 2000.

[25] Y. Luan and H. Li, "Clustering of time-course gene expression data using amixed-effects model with B-splines," *Bioinformatics*, vol. 19, no. 4, pp. 474–482. 2003.

[26] P. Ma, C.I. Castillo-Davis, W. Zhong, and J.S. Liu, "A datadriven clustering method for time course gene expression data," *Nucleic Acids Research*, vol. 34 no. 4, pp. 1261–1269. 2006.

[27] S. Makeig, J. Anthony, A.J., Bell, T. Jung, and T.J. Sejnowski, "Independent Component Analysis of Electroencephalographic data", *Advances in Neural Information Processing Systems 8*. 1996.

[28] E. Miller, and, J Fisher, "ICA using spacings estimates of entropy". *Journal of Machine Learning Research 4,* pp. 1271-1295, 2003

[29] N. Mitianoudis, and M. Davies, "New Fixed-Point ICA Algorithms For Convolved Mixtures", In Proceedings of the *3rd International Conference on Independent Component Analysis and Blind Source Separation*, San Diego, California, December 2001.

[30] D.T. Pham, and P. Garrat, "Blind separation of mixture of independent sources through a quasi-maximum likelihood approach". *IEEE Trans. on Signal Processing 45(7)* 1712-1725. 1997.

[31] H. Prautzsch,W. Boehm, and M. Paluszny, *B´ezier and B-Spline Techniques*, Springer, Berlin, Germany. 2002.

[32] J.D. Storey, W. Xiao, J.T. Leek, R.G. Tompkins, and R.W. Davis, "Significance analysis of time course microarray experiments," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 36, pp. 12837–12842, 2005.

[33] F. Rossi, D. Fracois, V. Wertz, M. Meurens, and M. Verleysen, "Fast Selection of Spectral Variables with B-Spline Compression". In *Chemo metrics and Intelligent Laboratory Systems* vol 86, pp. 208-218, 2007.

[34] D. Rueckert A.F. Frangi, and J.A. Schnabel, "Automatic Constructions of 3D Statistical Deformation Models of the Brain using non-rigid Registration". In *IEEE Transaction on Medical Imaging, Vo.l 22 No. 8*:1014-1025, 2003

[35] M.M. Van Hulle, "Sequential fixed-point ICA based on mutual information minimization" *Neural Computation*, vol. 20 no 5, pp. 1344-1365, 2008.

[36] J. Walters-Williams, and Y. Li. "Estimation of Mutual Information: A Survey". In the *Proceedings of the 4th International Conference on Rough Set and Knowledge Technology (RSKT2009),* pp. 389-396, 2009.

[37] D. Xu, J. Principle, J. Fisher and H.C. Wu, "A novel measure for independent component analysis (ICA)", In the *Proceedings. of ICASSP* vol. 2, pp. 1161-1164, 1998.

[38] W. Zhou and J. Gotman, "Removal of EMG and ECG Artifacts from EEG based on Wavelet Transform and ICA". In the *Proceedings of the 26th Annual International Conference of the IEEE EMBS*. 2004.

Janett Walters-Williams received the B.S. and M.S. degrees, from the University of the West Indies. in 194 and 2001, respectively.   She is presently a Doctoral student at the University of Southern Queensland. After working as an assistant lecturer (from 1995), in the Dept. of Computer Studies, the University of Technology, she has been a lecturer in the School of Computing & Information Technology, since 2001. Her research interest includes independent component analysis, Neural Network Applications, signal processing, biometrics and artificial intelligence.

Yan Li received her PhD degree from the Flinders University of South Australia, Australia in March 2003. She is currently an Associate Professor in the Department of Mathematics and Computing at the University of Southern Queensland, Australia. Her research interests lie in the areas of artificial intelligent, Neural Networks, Computer Communications and Internet Technologies, Blind Signal Separation, Signal/Image Processing technologies and their applications on medical bioinformatics etc.