

BMICA-INDEPENDENT COMPONENT ANALYSIS BASED ON B-SPLINE MUTUAL INFORMATION ESTIMATOR

Janett Walters-Williams^{1,2} and Yan Li²

¹ School of Computing and IT, University of Technology, Jamaica, 237 Old Hope Road,
Kingston 6, Jamaica

jwalters@utech.edu.jm

² Department of Mathematics and Computing, University of Southern Queensland,
Toowoomba, Australia

liyan@usq.edu.au

ABSTRACT

The information theoretic concept of mutual information provides a general framework to evaluate dependencies between variables. Its estimation however using B-Spline has not been used before in creating an approach for Independent Component Analysis. In this paper we present a B-Spline estimator for mutual information to find the independent components in mixed signals. Tested using electroencephalography (EEG) signals the resulting BMICA (B-Spline Mutual Information Independent Component Analysis) exhibits better performance than the standard Independent Component Analysis algorithms of FastICA, JADE, SOBI and EFICA in similar simulations. BMICA was found to be also more reliable than the "renown" FastICA.

KEYWORDS

B-Spline, Mutual Information, Independent Component Analysis, Reliability

1. INTRODUCTION

Blind Source Separation (BSS) focuses on solving the problem of estimating unobserved signals or sources from several observed mixtures without observing the source and knowing the mixing procedure. This lack of priori knowledge is compensated by the assumption that each source is statistically independent. BSS has wide applicability and has led to the development of Independent Component Analysis (ICA).

Herault and Jutten [25] were probably the first to informally use the concept of ICA in 1983. Comon presented a mathematical formulation a few years later defined as:

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

where A is an unknown $m \times m$ matrix called the mixing matrix, S is the pure signal and N is noise. The concept has led to the development of many algorithms such as JADE, AMUSE, SOBI, Radical, Infomax and FastICA. All these ICA algorithms identify the independent components (ICs) either simultaneously (symmetric) or one at a time (deflation). An important feature of most

ICA algorithms is a measure of the mutual independence of the extracted ICs. This measure is called the contrast function [12]. Many ICA algorithms are based on the minimization of this measure which have been based on

- (a) a quadratic “error” between probability densities [9],
- (b) moments of all orders [37],
- (c) Renyi’s entropy [38],
- (d) The mutual information of the estimated components [1,50].

Of the four, Mutual Information (MI) is considered the most appealing because it is:

- (i) A strict dependence measure
- (ii) Always non-negative
- (iii) Zero if only the estimated components are statistically independent
- (iv) Based on Shannon’s entropy
- (v) Insensitive to invertible transformations of the components

Hyvarinen *et al.* [23] considered MI to be the most natural and realistic approach to denoising, as it does not assume anything about the data; estimating it however raises difficulties. ICA literature presents mostly crude approximations to MI based on cumulant expansions which became popular because of their ease of use [29] and they have been very successful [12]. 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 MI where the difference between the marginal entropy and the joint entropy of different information sources was accumulated. The one difficulty of this method however is the estimation of marginal entropy. Although all these algorithms exist, Hyvarinen [21] stated that in their present use these algorithms are far from optimal as far as robustness and asymptotic variance are concerned. These algorithms are also sensitive to artifacts. Presently the newest MI estimators for ICA algorithms are using kernel density (KDE) and k-nearest neighbour (kNN) statistics [48].

Recently, B-Spline has been applied widely in the estimation of MI. Klien *et al.* [30], 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.* [47] presented MI schemes using B-Spline to help represent the deformation field. Daub *et al.* [13] 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 KDE are noted. They stated that B-Spline estimated MI outperforms all the other known algorithms for gene expression analysis. Rossi *et al.* [46] 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 –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 to use the recently introduced MI estimator based on B-Spline [13] to design a new ICA algorithm. The paper is organized as follows. In section 2 we discuss MI basic properties and present the MI estimator. The basic version of B-Spline Mutual Information ICA

(BMICA) is described in Section 3. In section 4 we present results of performance and a discussion of our method. Conclusions are drawn in the last section, Section 5.

2. MUTUAL INFORMATION

2.1. Properties

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 based on entropy as

$$\begin{aligned} I(X : Y) &= H(X) + H(Y) - H(X, Y) \\ &= H(X) - H(X | Y) \\ &= H(Y) - H(Y | X) \end{aligned} \quad (2)$$

and in terms of probabilities as

$$I(X : Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log \frac{p(x, y)}{f(x)g(y)} \quad (3)$$

where p is the joint probability distribution function of X and Y , and f and g are the marginal probability distribution functions of X and Y respectively. In the continuous case, it can be defined with a definite double integral:

$$I(X : Y) = \int_Y \int_X p(x, y) \log \frac{p(x, y)}{f(x)g(y)} dx dy \quad (4)$$

where p is now the joint probability density function (*pdf*) of X and Y , and f and g are the marginal probability density functions of X and Y respectively.

MI is considered to be very powerful yet it is difficult to estimate [16,29,52]. This is because MI requires the pdf of each variable and involves the integration of functions for other functions which can lead to high computational complexity – one problem [51]. To use the definition of entropy, the density has to be estimated which demands a duly large amount of data for an acceptable accuracy – another problem. Estimation can therefore be unreliable, noisy and even bias. These problems have severely restricted the use of MI in ICA estimation and many other applications. Joint density can be avoided however by merging Eq. (2) and Eq. (3) to give

$$\begin{aligned} H(X) &= - \sum_i p(x_i) \log p(x_i) \\ H(X, Y) &= - \sum_{i,j} p(x_i, y_j) \log p(x_i, y_j) \end{aligned} \quad (5)$$

Eq. (2) contains the term $-H(X, Y)$, which means that maximizing MI is related to minimizing joint entropy.

2.2. B-Spline based Estimator

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.

Krishnaveni *et al.* [29] found that a MI estimated using kNN distance outperforms many of the known ICA algorithms. B-Spline estimators according to our previous research [43] have been shown however to be one of the best nonparametric approaches, second to only wavelet density estimators thus better than Krishnaveni's kNN estimator.

Estimation of multidimensional entropies constitutes sufficient means to indirectly compute MI with Eq. (2) and Daub *et al.* [13] have presented a method for estimating such. They proposed a generalized binning method, which makes use of B-Spline functions to assign data points to bins. B-Spline is a flexible mathematical formulation for curve fitting due to number of desirable properties [44]. The shape of the B-Spline functions, in the method, is determined by their order k ($1..M$), which is a parameter of the method. With B-Spline order 1, each point is assigned to exactly one bin and the method is equivalent to simple equidistant binning. The proposed method is thus a fixed binning scheme extended with a preprocessing step designed to reduce the variance. A B-Spline function is defined with the help of a knot vector

$$t_i = \begin{cases} 0 & \text{if } i < k \\ i-k+1 & \text{if } k \leq i \leq M-1 \\ m-1-k+2 & \text{if } i > M-1 \end{cases} \quad (6)$$

where M is the total number of bins and i is an index into the knot vector. The B-Spline functions are defined and evaluated recursively, based on the Cox-de Boor Recursion formula, by

$$B_{i0}(z) = \begin{cases} 1 & \text{if } t_i \leq z < t_{i+1} \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

$$B_{i,k}(z) = B_{i,k-1}(z) \frac{z-t_i}{t_{i+k-1}-t_i} + B_{i+1,k-1}(z) \frac{t_{i+k}-z}{t_{i+k}-t_{i+1}} \quad (8)$$

where z is an element in the domain of the function i.e. $z \in [0, M-k+1]$. The probability of each bin is estimated by

$$p(x_i) = \frac{1}{N} \sum_{j=1}^N B_{i,k}(f_{M,k}(x_j)) \quad (9)$$

where $f_{M,k}(x)$ is a linear transformation which maps the values of x onto the domain of the B-Spline functions. $B_{i,k}$ is a B-Spline function of order k evaluated at bin i ; $\sim x_j$ is an appropriately scaled data sample mapping the values of x into the domain of the B-Spline functions. In two dimensions the joint *pdf* is computed as

$$p(x_i, y_j) = \frac{1}{N} \sum_{l=1}^N B_{i,k}(f_x(x_l)) B_{j,k}(f_y(y_l)) \quad (10)$$

MI, $I_{M,k}(X; Y)$ can then be determined as

$$I_{M,k}(X : Y) = H_{M,k}(X) + H_{M,k}(Y) - H_{M,k}(X, Y) \quad (11)$$

and each of the terms may be computed using the formulas in Eq. (9) & Eq. (11) together with Eq. (5). The notation $I_{M,k}(X;Y)$ and $H_{M,k}(X; Y)$ indicates that the method has two parameters: M , the number of bins and k , the order of the B-Spline.

3. NEWLY DESIGNED ICA

The aim of all ICA algorithms is to determine S from Eq (1). 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 :

$$u = W X = W A S , \quad (12)$$

ICA is therefore firstly concerned with finding W . In our algorithm we seek W such that it minimizes the MI estimator $I(Y)$ of the estimated components $Y = WX$.

Our ICA algorithm is a fixed point algorithm because these algorithms converge faster than other algorithms [44]. Unlike the gradient descent method, there is no need for the 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 [40]. Like all fixed-point algorithms we have a two-step approach – prewhitening and rotation of the observation vector.

3.1. PreWhitening

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 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. Here the matrix W is usually decomposed into two factors

$$W = R V \quad (13)$$

where the prewhitening V transform the covariance matrix into

$$C' = V C V^T = 1 \quad (14)$$

and R is a pure rotation. Whitening is done by multiplication with the transformation matrix P .

$$P = V C^{-1/2} V^T \quad (15)$$

$$\hat{W} = P \hat{W} \quad (16)$$

3.2. Rotation

Once Prewhitening is done the ICA problem is reduced to finding a suitable rotation for the prewhitened data. Our algorithm aims to minimize MI between the output signals under the pure rotation R . The performance of our algorithm depends on the nonquadratic nonlinear function G used. The four most used are *pow3*, *tanh*, *gaus*, and *skew*. Of the four, *tanh* has been described as a “good general purpose contrast function” [18] and one which is better at producing a more robust algorithm [33]. Our algorithm therefore uses *tanh*. As this is a symmetric algorithm,

independent sources are extracted simultaneously from the mixed signal. The algorithm is defined as follows, with X being the input data vector and u the output:

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

$$y = z' \times B \quad (17)$$

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

$$B = (zg(y)' / m - \sum (1 - g(y)^2) \times I) / m \quad (18)$$

where

$$g(y) = \tanh(y) \quad (19)$$

- (iv) Do a symmetric orthogonalization of B by

$$B = (BB^T)^{-1/2} B \quad (20)$$

4. Compute W using B
5. Determine u by Eq. (12)

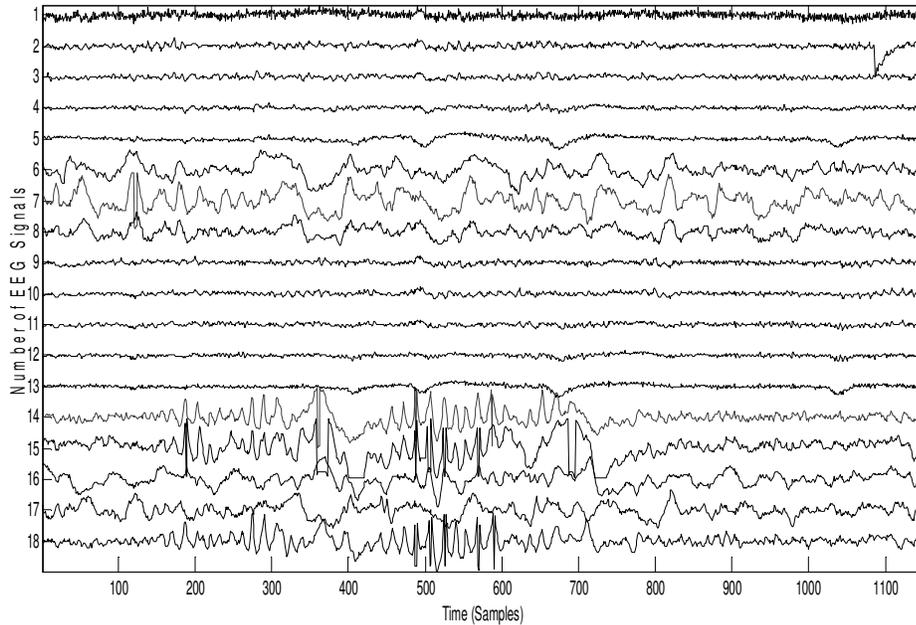


Figure 1 Sample of Raw EEG Signals

4. EXPERIMENTAL RESULTS

In this section we will investigate the BMICA algorithm to determine its validity. Data utilized in the performance tests were real comprised of electroencephalographic (EEG) signals from both human and animals collected from the following sites:

- (i) http://sccn.ucsd.edu/~arno/fam2data/publicly_available_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/databases.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. Experiments were conducted using the above mentioned signals, in Matrix Laboratory (MATLAB) 7.8.0 (R2009) on a laptop with AMD Athlon 64x2 Dual-core Processor 1.80GHz. Fig 1 shows one mixed EEG signal set where overlays in signals are in signals 2, 6-8 and 14-18. Fig 2 shows the same signal set after the application of BMICA showing that the overlays have been minimized – noise has been removed.

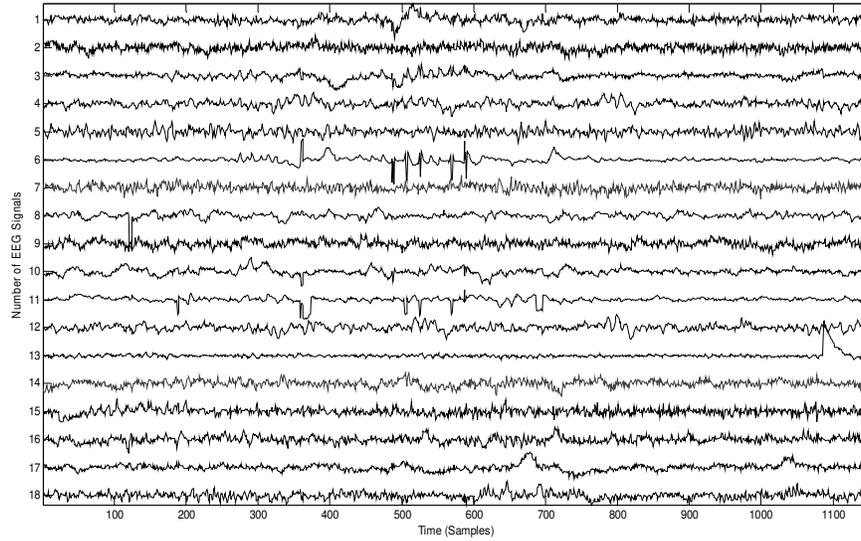


Figure 2. EEG Signals after denoised with BMICA algorithm

4.1. Performance Testing

Here we test the performance of the BMICA algorithm. Experiments were mainly aimed at assessing the method’s ability to perform ICA (extraction of ICs) and not blind source separation (recovery of original sources). The performance measures that will be used throughout are based on two categories of calculation:

1. Noise/Signal Measures - Mean Square Error (MSE), Peak Signal to Noise Ratio (PSNR), Signal to Noise Ratio (SNR) and
2. Separation Accuracy Measures - Signal to Distortion Ratio (SDR), Signal to Interference Ratio (SIR), Amari Performance Index.

Comparison with two categories of benchmark ICAs will be provided namely:

- (i) fixed-point- FastICA[22], and EFICA[28]
- (ii) non fixed-point - SOBI[6.], and JADE [10]

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

4.1.1 Noise/Signal Measures

These performance measures determine the relationship between the actual signal and the noise distortion which overlay them. MSE, defined as:

$$MSE = \frac{1}{N} \sum_{y=1}^N [I(x, y) - I'(x, y)]^2. \quad (21)$$

measures the average of the square of the “error” (the amount by which the estimator differs from the quantity to be estimated) and is the direct opposite of PSNR which is defined as:

$$PSNR = 10 \times \log_{10} \left(\frac{MAX^2}{MSE} \right). \quad (22)$$

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. For a perfect fit in Eq. (21), $I(x,y) = I'(x,y)$ 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. PSNR, unlike MSE which represents the cumulative squared error between the denoised and mixed signal, 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. Therefore an ICA which produces a small MSE

Table 1. MSE comparison for 15 EEG signals

BMICA	FASTICA	SOBI	JADE	EFICA
1.66E+03	1.67E+03	1.67E+03	1.66E+03	1.69E+03
1.27E+03	1.30E+03	1.29E+03	1.31E+03	1.28E+03
1.16E+03	1.17E+03	1.19E+03	1.22E+03	1.21E+03
1.81E+03	2.01E+03	1.01E+03	2.02E+03	2.00E+03
1.11E+03	1.12E+03	1.11E+03	1.08E+03	1.11E+03
1.17E+03	1.53E+03	1.54E+03	1.54E+03	1.55E+03
1.28E+04	1.29E+04	1.29E+04	1.29E+04	1.28E+04
4.91E+05	4.91E+05	4.91E+05	4.91E+05	4.92E+05
4.63E+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	3.30E+05
9.41E+02	9.63E+02	9.56E+02	9.43E+02	9.62E+02
8.79E+02	9.18E+02	9.18E+02	9.57E+02	9.82E+02
7.51E+02	7.73E+02	7.65E+02	7.58E+02	7.86E+02
7.09E+02	7.04E+02	7.16E+02	7.23E+02	7.17E+02
4.59E+02	4.62E+02	4.66E+02	4.70E+02	4.70E+02

and large PSNR is considered to be close to the actual signal. Table 1 shows sample MSE results and Table 2 shows sample PSNR for 15 EEG signals. The results for MSE show excellent performance for BMICA, having the lowest MSE 73% of the time. It was surpassed by the FastICA algorithm only 20% of the time while for the other times there was no significant difference in the MSE for any algorithm. For Table 2 it can be seen that BMICA has the highest PSNR 53% of the time. SOBI and JADE were marginally better than our algorithm 33% of the time.

Table 2. PSNR comparison for 15 EEG signals

BMICA	FASTICA	EFICA	SOBI	JADE
15.9207	15.8890	15.8526	15.9025	15.9205
17.0666	16.9961	17.0517	17.0139	16.9518
18.3936	18.2954	18.2982	18.3244	18.3857
18.6811	18.5034	18.2114	18.5020	18.3204
19.3730	19.2487	19.1739	19.2944	19.3362
17.4924	17.4529	17.2988	17.3896	17.3567
19.8679	19.9043	19.8811	19.9030	19.8743
19.6233	19.6568	19.5777	19.5816	19.5386
17.6643	17.6427	17.6769	17.6829	17.7787
13.1659	13.1872	13.2007	13.1822	13.1756
16.2423	16.2747	16.2377	16.2448	16.2435
20.3691	20.4044	20.4556	20.4020	20.4578
21.5157	21.4804	21.4104	21.4500	21.4110
30.0699	30.1499	30.1252	30.2565	30.1489
29.6448	29.7131	29.6117	29.7112	29.6404

Another signal/noise measure is SNR, defined mathematically as:

$$SNR(dB) = 20 \log_{10} \frac{\sum_{n=0}^N s^2(n)}{\sum_{n=0}^N x^2(n)} \quad (23)$$

It refers to how much signal and 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. For performance, the greater the ratio, evidenced by a larger number, the less noise and the more easily it can be filtered out. Examination of Table 3 shows BMICA to have the highest SNR clearly demonstrating that it has a less obtrusive background noise. BMICA has filtered out more noise than the other algorithms and therefore produces cleaner signals as its result.

Table 3. SNR comparison for 13 EEG signals

BMICA	FASTICA	SOBI	JADE	EFICA
2.26E-02	7.37E-02	1.80E-03	-9.30E-03	-3.43E-02
6.69E-02	-4.05E-02	8.86E-04	-2.12E-02	2.99E-02
9.60E-02	-7.84E-02	8.96E-04	1.18E-01	-5.81E-02
2.38E-01	2.00E-03	6.15E-04	-2.14E-01	-2.86E-01
1.16E-01	-5.47E-02	1.93E-02	2.21E-02	-1.25E-01
3.16E-02	2.80E-03	7.04E-05	7.50E-03	-2.76E-02
5.83E-02	9.86E-02	-3.02E-04	1.20E-03	1.92E-02
8.90E-03	2.85E-02	2.62E-04	-2.81E-02	4.80E-03
5.92E-05	1.66E-04	-9.32E-07	7.23E-05	-5.30E-03
1.05E-02	-2.50E-03	2.37E-04	3.20E-03	2.39E-02
1.35E-04	8.68E-04	2.30E-08	5.89E-04	3.27E-04
4.20E-03	4.47E-02	3.53E-04	7.60E-03	-2.46E-02
5.82E-02	1.37E-02	-7.14E-05	-4.68E-02	-3.53E-02

4.1.2 Separation Accuracy Measures

The most widely studied ICA situation is the non degenerate case where there are at least as many mixtures as there are sources (i.e., $n \geq m$). When that is the case the accuracy of an ICA algorithm can be assessed from its ability to estimate the mixing matrix. The most widely used measure for assessing the accuracy of the estimated mixing matrix is the Amari performance index P_{err} [2]:

$$P_{err} = \frac{1}{2m} \sum_{i,j=1}^m \left(\frac{|p_{ij}|}{\max_k |p_{ik}|} + \frac{|p_{ij}|}{\max_k |p_{kj}|} \right) - 1 \quad (24)$$

where $p_{ij} = (BA)_{ij}$. It assesses the quality of the de-mixing matrix W for separating observations generated by the 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 $m/2 - 1$. This is most likely the case when we try to separate the underlying brain sources from EEG recordings using a reduced set of electrodes. In the simulations it was found that the Amari indexes obtained for the different algorithms were similar even when different sample sizes were used. This is presented in Fig 3. This indicates that when the quality of A (determined from $W=A^{-1}$) is assessed BMICA has a similar performance to the other algorithms i.e. the performance of BMICA, as how well the demixing transformation W agrees with the true mixing matrix A , does not have any mark deviation from the behavior of the other algorithms.

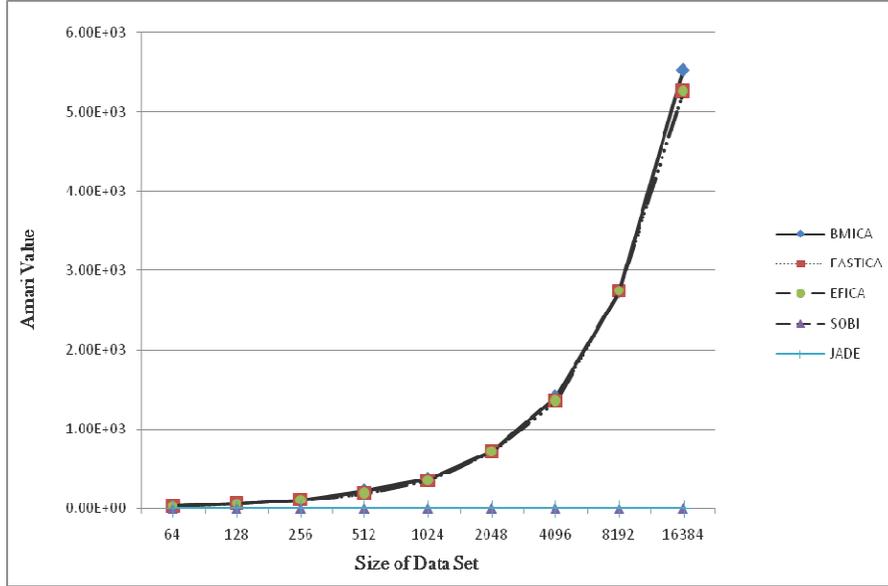


Figure 3. Amari Index Comparison with Benchmark algorithms

In degenerate demixing, the accuracy of an ICA algorithm cannot be described using only the estimated mixing matrix. In this case it becomes of particular importance to measure how well BSS algorithms estimate the sources with adequate criteria. The most commonly used index to assess the quality of the estimated sources is SIR [14]:

$$SIR(dB) = \frac{1}{n} \sum_{i=1}^n \left(\sum_j \frac{|p_{ij}|}{\max_k |p_{ij}|} - 1 \right) \quad (25)$$

The lower the SIR, the better the achieved separation therefore when the estimated source is orthogonal to the true source, the SIR is equal to zero. Since EEG signals are of the degenerate ICA problem to confirm the separation nature of BMICA Fig 4 shows its behaviour based on SIR. Overall, BMICA has the smallest SIR – the best separation of overlays and signals.

While SIR assesses the quality of the estimated sources, and the Amari Index assess the accuracy of the estimated mixing matrix, how accurate the separation of an ICA algorithm in terms of the signals is calculated by the total SDR defined as:

$$SDR(x_i, y_i) = \frac{\sum_{n=1}^L x_i(n)^2}{\sum_{n=1}^L (y_i(n) - x_i(n))^2} \quad i = 1, \dots, m, \quad (26)$$

where $x_i(n)$ is the original source signal and $y_i(n)$ is the reconstructed signal. The higher the SDR value, the better the separation of the signal from the noise. Examination of Table 4 shows that BMICA has an excellent performance rate i.e. its SDR is the highest 73% of the time. It is surpassed by only FastICA 20% of the time.

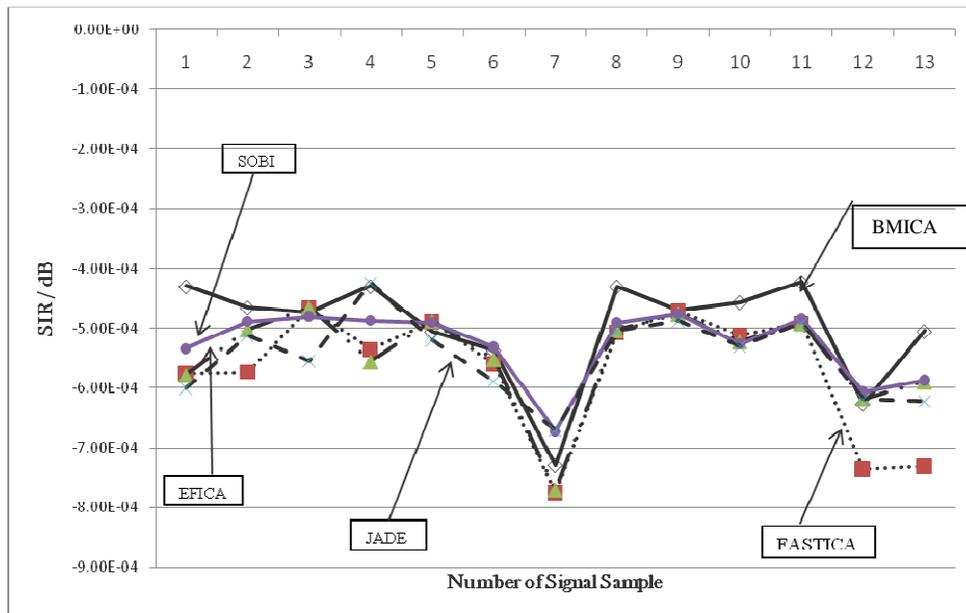


Figure 4. SIR comparison with Benchmark algorithms

4.2. Reliability Testing

Obtaining the ICs from a mixture is the first part on an ICA analysis. The next part is checking the actual dependencies between the obtained components although it is often ignored. Here the task is to check the reliability and robustness of the ICs.

Because ICA algorithms are iterative the results in most algorithms tend to differ when executed multiple times on the same dataset. The problem is that most ICA algorithms are based on methods related to gradient descent. The basic principle is to start in some initial point, and then make steps in a direction that decreases the objective function; until one finds a point in which the objective function is locally minimized. Depending on the point where the search was started (the "initial point"), the algorithm will find different local minima [17]. Results from each algorithm therefore have a level of uncertainty, which components should be considered seriously is not clear. This means that for real data it is not clear which components of the output should be interpreted meaningfully. Himberg and Hyvarinen [17] developed a software package for assessing both the algorithmic and statistical reliability of estimated independent components (ICs) called ICA repeatability analysis (ICASSO). The technique is based on running an algorithm many times with slightly different conditions and visualizing the clustering structure of the obtained components in the signal space. We used ICASSO to test the reliability of the estimated ICs from BMICA.

Table 4. SDR comparison for 15 EEG signals

BMICA	FASTICA	EFICA	SOBI	JADE
3.46E-01	1.02E-02	3.38E-01	1.43E-04	3.31E-01
3.24E-01	3.28E-01	3.21E-01	2.75E-04	3.26E-01
5.63E+00	5.68E+00	5.37E+00	3.67E-05	5.60E+00
5.46E+00	5.56E+00	5.55E+00	3.73E-05	5.75E+00
1.15E-01	1.12E-01	1.12E-01	9.68E-05	1.10E-01
7.56E-02	7.47E-02	7.33E-02	4.63E-05	7.42E-02
1.55E-01	1.43E-01	1.41E-01	6.61E-05	1.54E-01
1.86E-01	3.89E-04	1.63E-01	5.56E-05	1.64E-01
1.16E-01	1.10E-01	1.10E-01	1.83E-07	1.08E-01
1.75E-01	1.31E-01	1.30E-01	1.44E-07	1.27E-01
3.92E-01	3.87E-01	3.80E-01	4.19E-06	3.89E-01
1.92E-01	1.95E-01	1.95E-01	1.81E-04	1.93E-01
3.39E+00	4.47E+00	4.42E+00	3.60E-03	4.70E+00
1.63E-02	1.63E-02	1.63E-02	7.88E-06	1.63E-02
3.10E-01	3.07E-01	3.01E-01	-1.16E-05	-2.40E-02

ICASSO used cluster analysis to reveal results where the clustering of the estimated components is expected to yield information on the reliability (robustness) of estimation. Fig. 5 shows the 2D projections of the clustered IC estimates for BMICA where each IC cluster is prescribed by a convex hull with the black dots representing the individual IC estimates. In the figure the lightly coloured lines connect estimates whose similarity is larger than a threshold and the darker lines indicate that there are stronger similarities. A compact, tight cluster emerges when a similar component repeatedly comes up despite the randomization. From Fig 5 the smaller clusters are indicated by smaller, tighter convex hulls and the tightness of a cluster corresponds to a component's reliability. It can be seen that the clusters for BMICA are relatively small. Cluster analysis also shows that if an IC is reliable then every run of the algorithm produces one point that is very close to the real component thus producing a tight cluster. The tighter cluster, the smaller it becomes resulting in a single point. This is an ideal cluster [18]. There are ideal clusters in #1, #2 and #3. ICASSO also tested reliability using a stability (quality) index (I_q) which indicates ideal cluster when equal to one. The I_q of each cluster from BMICA was calculated resulting in stability indices of most of the clusters ranging from 0.45 to 0.7, shown in Fig 6. Further analysis of the I_q also reveals that 67% of ICs have I_q of 0.5 indicating that the clusters were getting more compact thereby indicating that the estimates are more consistent thus more reliable. FastICA has been established over time to be a reliable algorithm. BMICA was therefore compared with this algorithm to see its level of reliability. Investigations using the same signal sets show that the I_q for BMICA was better as can be seen when Fig 6 is compared with Fig 7. Fig 7 shows a sample I_q set for FastICA where the I_q increases from 0.35 to 0.6. Further analysis reveals that only 20 of the signals have I_q greater or equal to 0.5 compared to 67% for BMICA. Smaller I_q indicates that the clusters are wide and mix up; more compact clusters have high I_q .

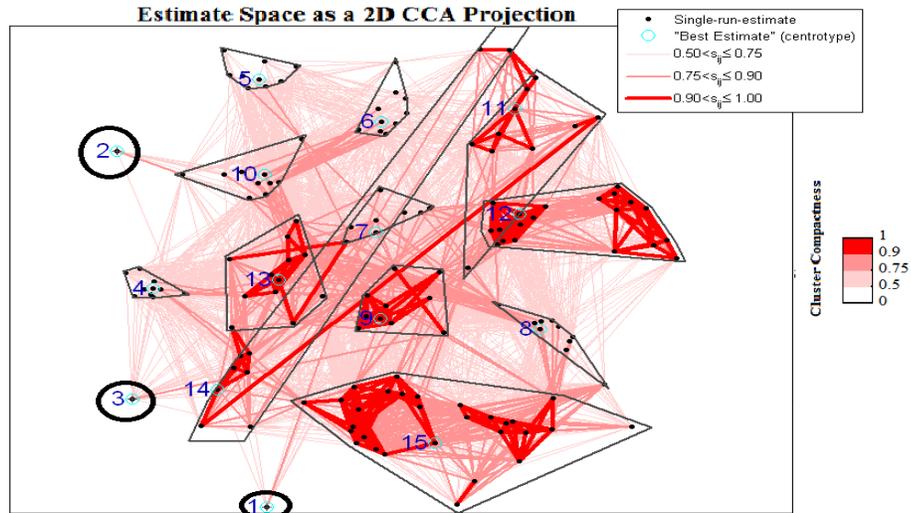


Figure 5: Cluster Analysis of BMICA using ICASSO

In [54] investigations were done on all eight linearity of FastICA and compared to BMICA. Research concluded that the clusters in BMICA are more compact than those of FastICA thus the ICs are more reliable for all linearity. ICASSO therefore showed that BMICA produces ICs which can be interpreted meaningfully.

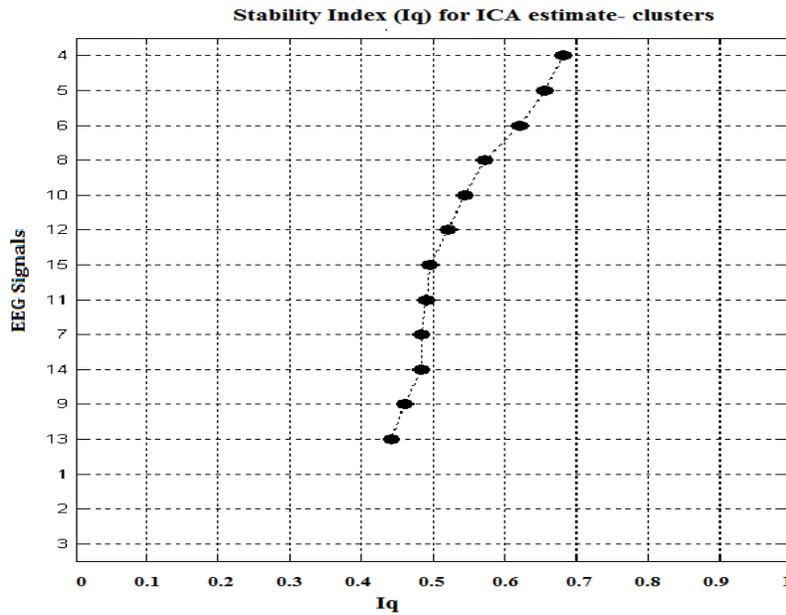


Figure 6. Similarity (I_q) Analysis of BMICA using ICASSO

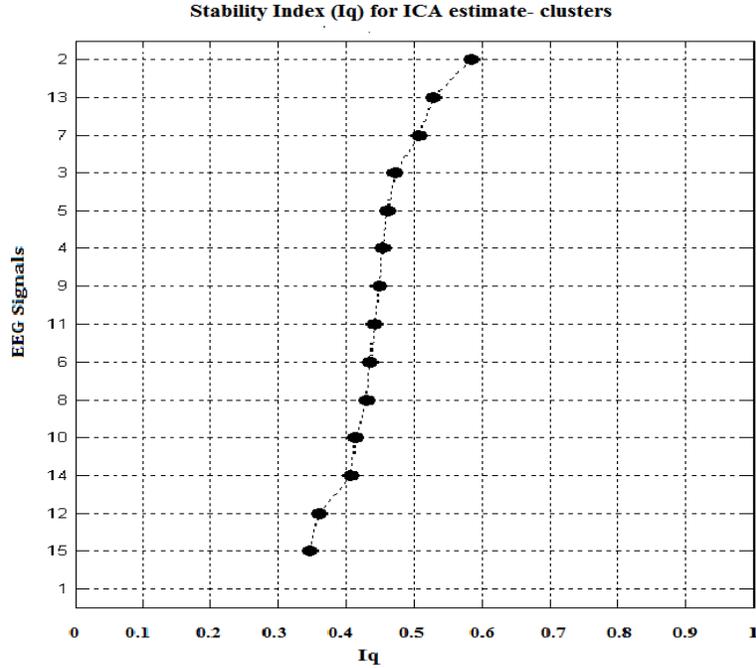


Figure 7. Similarity (Iq) Analysis of FastICA using ICASSO

4.3. 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 various ICA approaches from a numeral complexity viewpoint. Here we determine the computational complexity of BMICA and compare it to benchmarked ICA algorithms such as FastICA, and JADE.

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 $O(N)$
- Run 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 has a complexity of $O(M) * O(N^2) * O(N/2)$ resulting in an overall complexity of $O(N^2M)$. When compared to other ICA algorithms it was found that

- (i) FastICA has a complexity on the order of $O(N^3M)$ [51],
- (ii) JADE algorithm is on the order of $O(N^4M)$ [51], and
- (iii) EFICA has a computational complexity only slightly (about three times) higher than that of the standard symmetric FastICA[28],

Investigations show that BMICA has the smallest overall complexity.

4.4. Convergence Speed

It has been shown experimentally that FastICA has outperformed most of the commonly used ICA algorithms [41] and its convergence speed has been the topic of many research [11,41,42]. We have therefore compared the convergence speed of BMICA with FastICA. FastICA is

quadratic in general and cubic for symmetric distributions [20]. Investigations have shown that our algorithm has a slower speed than FastICA. Comparison of the running time has also shown that FastICA completes its analysis much faster than BMICA. In both instances we conclude that the presence of the B-Spline iterative steps influences the speed of the BMICA as on average B-Spline calculates slowly.

4.5. Discussion

The quality of the BMICA algorithm was assessed based on performance, reliability, computational complexity and convergence speed. Since it was designed as a fixed-point algorithm its performance as an ICA algorithm was compared to the most widely used – FastICA and EFICA which was created as an improvement of FastICA [27]. It was also compared to SOBI and JADE two well used nonfixed-point algorithms to show overall performance. Assessment shows that BMICA has out performed both categories of algorithms having the smallest signal/noise ratio and producing the cleanest results.

All ICA algorithms will find some meaningful solution, however these solutions have a level of uncertainty, which components should be considered seriously is not clear. To address this problem, algorithms should be assessed to determine their fitness to the underlying ICA model. The better the model fits the reality, the more reliable the components derived from the ICA algorithm. We have assessed the reliability of BMICA using the ICASSO technique and have, after executing the algorithm many times and administering cluster analysis on the real world data, found that the I_q for BMICA tended to 1 more than the “reliable” FastICA (67% to 20%), and the clusters were mostly compact. This indicates that BMICA produces ICs which can be interpreted meaningfully – BMICA is reliable.

Investigations of the computational complexity of BMICA, when compared to the chosen benchmark algorithms of JADE, EFICA and FastICA, finds BMICA having the smallest complexity. Hence, our algorithm is computationally more intensive. It has been found however that it converges slowly due to the presence of B-Spline.

5. CONCLUSIONS

In the first part of this paper we discussed the ICA model and introduced a new algorithm based on using MI designed using B-Spline functions. We have also shown that entropy can be easily estimated using B-Spline functions. Our algorithm was tested on real world data since it works with actual dependencies between reconstructed sources.

Comparisons with benchmark algorithms FastICA, JADE, EFICA and SOBI showed that the B-Spline based MI ICA is extremely rewarding; its superior performance over the other methods very encouraging. In simulations, including those reported in this paper BMICA has

- (i) The best Signal/Noise Ratio having
 - the highest PSNR and SNR and
 - the lowest MSE
- (ii) The best Separation Accuracy having
 - The highest SDR
 - the lowest SIR and
 - similar Amari Performance Index to the other four algorithms

Comparison with FastICA has shown that it is more reliable however it is slower in speed than FastICA. These results show that BMICA can perform comparable to the benchmark algorithms.

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

REFERENCES

- [1] Almeida L.B. (2000) "Simultaneous Mi-based Estimations of Independent Components and their Distributions", *Proceedings of 2nd International Workshop on Independent Component Analysis and Signal*, pp 169-74.
- [2] Amari, S., Chichocki, A., and Yang, H. H., (1996) "A new learning algorithm for blind source separation", *Advances in Neural Information Processing (NIPS'95)*, pp. 757–763.
- [3] Bach, F. and M. Jordan, (2002) "Kernel independent component analysis", *Journal of Machine Learning Research* 3: vol. 1 no. 48.
- [4] Bar-Joseph, Z., G.K. Gerber, D.K., Gifford, T.S. Jaakkola, and I. Simon, (2003) "Continuous representations of time-series gene expression data", *Journal of Computational Biology*, vol. 10 no. 3-4, pp. 341–356.
- [5] Bell, A.J. and T.J. Sejnowski, (1995) "An Information-Maximization Approach to Blind Separation and Blind Deconvolution", *Neural Computation*, vol. 7 no. 6, pp. 1129—1159.
- [6] Belouchrani, A. K.A., Meraim, J.F. Cardoso, and E. Moulines, (1993) "Second-order blind separation of correlated sources" *International Conference on Digital Signal Processing*, Cyprus, pp. 346–351.
- [7] Bhasi, K., A. Forrest, and M. Ramanathan, (2005) "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.
- [8] Boscolo, R., H. Pan and V.P. Roychowdhury, (2004) "Independent component analysis based on nonparametric density estimation", *IEEE Transactions on Neural Networks*, vol. 15, no. 1.
- [9] Burel G. (1992) "Blind Separation of Sources: A nonlinear neural algorithm", *Neural Networks* vol. 5 no. 6, pp937-947.
- [10] Cardoso, J.F. (1998) Blind signal separation: statistical principles. *IEEE* vol. 86 no. 10, pp. 2009-2025.
- [11] Chen, X., L. Wang, and Y. Xu, (2009) "A Symmetric Orthogonal FastICA Algorithm and Applications in EEG," *5th International Conference on Natural Computation*, 2009, vol. 2, pp.504-508.
- [12] Comon, P. (1994) "Independent component analysis, a new concept?" *Signal Processing* vol. 36 no. 3, pp. 287-314.
- [13] Daub C.O., R. Steuer, J. Selbig, and S. Kloska, (2004) "Estimating mutual information using B-Spline functions—an improved similarity measure for analysing gene expression data," In *BMC Bioinformatics*, vol. 5, no. 1, p. 118.
- [14] Gribonval, R., Vincent, E., and Févotte, C., (2003) "Proposals for performance measurement in source separation," *4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003)*, Nara, Japan, pp. 763–768.
- [15] Hastie, T. and R. Tibshirani, (2002) "Independent component analysis through product density estimation, Technical report, Stanford University.
- [16] He, W. (2002) "A spline function approach for detecting differentially expressed genes in microarray data analysis," *Bioinformatics*, vol. 20, no. 17, pp. 2954–2963.
- [17] Himberg, J. and A. Hyvarinen, (2003) "ICASSO: Software for Investigating the reliability of ICA estimates by Clustering and Visualization", *IEEE Workshop on Neural Networks for Signal Processing (NNSP2003)*, pp 259-268.

- [18] Himberg, J., A. Hyvarinen, F. Esposito, (2004) "Validating the independent components of neuroimaging time-series via clustering and visualization", *Neuroimage*, vol 22 no. 3, pp1214-1222
- [19] Hoffman, S. and M. Falkenstein, (2008) "The Correction of Eye Blink Artefacts in the EEG: A Comparison of a Two Prominent Methods", *PLoS One* vol. 3 no. 8, p. e3004.
- [20] Hyvarinen, A. (1999) "Fast and robust fixed-point algorithms for independent component analysis, *IEEE Transactions on Neural Networks*, vol 10 no 3 pp 626 – 634.
- [21] Hyvärinen, A. (1999) "Survey on Independent Component Analysis", *Neural Computing Surveys* vol 2 pp 94-128.
- [22] Hyvärinen A. and E. Oja, (1997) "A Fast Fixed-Point Algorithm for Independent Component Analysis", *Neural Computation*, vol. 9 no. 7, pp. 1483-1492.
- [23] Hyvarinen, A. J. Karhunen and E. Oja, (2001) *Independent Component Analysis*, eds. Wiley & Sons.
- [24] Imoto, S. T. Goto, and S. Miyano, (2002) "Estimation of genetic networks and functional structures between genes by using Bayesian networks and nonparametric regression," *Pacific Symposium on Biocomputing*, pp. 175–186.
- [25] Josien, P., W. Pluim, J.B. Antoine Maintz and M.A. Viergever, (2003) "Mutual information Based Registration of Medical Images: A Survey", *IEEE Transactions on Medical Imaging*, vol. XX, no. Y.
- [26] Kachenource, A., L. Albera, L. Senhadji, P. Comon, (2008) "ICA: A potential Tool for BCI Systems", *IEEE Signal Processing Magazine*, pp 57-68
- [27] Karvanen, J., J.Eriksson, and K.V. Pearson, (2000) "System Based Method for Blind Separation", *2nd International Workshop on Independent Component Analysis and Blind Signal Separation*, pp. 585—590.
- [28] Koldovský Z. and P. Tichavský, (2007) "Time-Domain Blind Audio Source Separation Using Advanced ICA Methods", *8th Annual Conference of the International Speech Communication Association (Interspeech 2007)*, pp. 846-849.
- [29] Krishnaveni, V., S. Jayaraman P.M. Manoj Kumar, K. Shivakumar, and K. Ramadoss, (2005) "Comparison of Independent Component Analysis Algorithms for Removal of Ocular Artifacts from Electroencephalogram", *Measurement Science Review, Volume 5*, Section 2.
- [30] Klein, S., M. Marius Staring, and P.W. Pluim, Josien, (2005) "Comparison of gradient approximation techniques for optimisation of mutual information in nonrigid registration" *SPIE, (Medical Imaging 2005)* vol.5747, 192.
- [31] Lee, T.W., M. Girolami, and T. Sejnowski, (1999) "Independent component analysis using an extended informax algorithm for mixed subgaussian and supergaussian sources". *Neural Computation* vol. 11 no. 2, pp.417-441.
- [32] Lee, J.H., H.Y. Jung, T.W. Lee, and S.Y. Lee, (2000) "Speech feature extraction using independent component analysis", *e25th International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 3, pp. 1631-1634
- [33] Li, Y., T. Adalı, and V.D. Calhoun, (2007) "Estimating the number of independent components for functional magnetic resonance imaging data", *Human Brain Mapping* vol. 28 issue 11 pp1251–1266.
- [34] Luan Y., and H. Li, (2003) "Clustering of time-course gene expression data using a mixed-effects model with B-splines," *Bioinformatics*, vol. 19, no. 4, pp. 474–482.
- [35] Ma, P., C.I. Castillo-Davis, W. Zhong, and J.S. Liu, (2006) "A datadriven clustering method for time course gene expression data" *Nucleic Acids Research*, vol. 34 no. 4, pp. 1261–1269.
- [36] Makeig, S., J. Anthony, A.J., Bell, T. Jung, and T.J. Sejnowski, (1996) "Independent Component Analysis of Electroencephalographic data", *Advances in Neural Information Processing Systems* 8.
- [37] Marques G.C., and L.B. Almeida, (1996) "An objective function for independence", *International Conference on Neural Network*, pp 453-457.

- [38] Marques G.C., and L.B. Almeida, (1999) "Separation of Nonlinear Mixtures Using Pattern Repulsion", *1st International Workshop on Independent Component Analysis and Signal Separation* pp 277-282.
- [39] Miller, E., and J Fisher, (2003) "ICA Using Spacings Estimates Of Entropy", *Journal of Machine Learning Research* 4, pp. 1271-1295.
- [40] Mitianoudis N., and M. Davies, (2001) "New Fixed-Point ICA Algorithms for Convolved Mixtures", *3rd International Conference on Independent Component Analysis and Blind Source Separation*, San Diego, California.
- [41] Oja, E., and Z. Yuan, (2006) "The FastICA Algorithm Revisited: Convergence Analysis", *IEEE Transactions on Neural Networks* vol. 17, no. 6.
- [42] Oja, E. (2002) "Convergence of the symmetrical FastICA algorithm", *9th International Conference on Neural Information Processing (ICONIP '02)* vol. 3, pp1368 – 1372.
- [43] Pham, D.T., and P. Garrat, (1997) "Blind separation of mixture of independent sources through a quasi-maximum likelihood approach". *IEEE Transaction on Signal Processing* vol. 45 no. 7, pp 1712-1725.
- [44] Prasad, R., H. Saruwatari, A. Lee, K. Shikano, (2003) A Fixed-Point ICA for Convolved Speech Signal Separation, In the *Proceeding of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003)*, pp 579-584.
- [45] Prautzsch, H., W. Boehm, and M. Paluszny, (2002) *B´ezier and B-Spline Techniques*, Springer, Berlin, Germany.
- [46] Rossi, F., D. Fracois, V. Wertz, M. Meurens, and M. Verleysen, (2007) "Fast Selection of Spectral Variables with B-Spline Compression". *Chemo metrics and Intelligent Laboratory Systems* vol 86, pp. 208-218.
- [47] Rueckert D., A.F. Frangi, and J.A. Schnabel, (2003) "Automatic Constructions of 3D Statistical Deformation Models of the Brain using non-rigid Registration". *IEEE Transaction on Medical Imaging*, vol 22 no. 8 pp 1014-1025.
- [48] H. Stogbauer, A. Kraskov, S.A. Astakhov, P. Grassberger, (2004) "Least Dependent Component Analysis based on Mutual Information" *Physical Review E* 70 (6) 066123.
- [49] Storey, J.D., W. Xiao, J.T. Leek, R.G. Tompkins, and R.W. Davis, (2005) "Significance analysis of time course microarray experiments," *National Academy of Sciences of the United States of America*, vol. 102, no. 36, pp. 12837–12842.
- [50] Taleb, A., C. Jutten, (1997) "Entropy optimization application to blind separation of sources, 7th International Conference on Artificial Neural Networks(ICANN) pp 529-534
- [51] Torkkola, K. (2003) "Feature Extraction by Non parametric Mutual Information Maximization", *Journal of Machine Learning Research*, vol. 3.
- [52] M.M. Van Hulle, (2008) "Sequential Fixed-Point ICA Based on Mutual Information Minimization" *Neural Computation*, vol. 20 no 5, pp. 1344-1365.
- [53] Walters-Williams, J. and Y. Li. (2009) "Estimation of Mutual Information: A Survey". *4th International Conference on Rough Set and Knowledge Technology (RSKT2009)*, pp. 389-396.
- [54] Walters-Williams J. and Y. Li, Reliability Testing of B-Spline Mutual Information Independent Component Analysis, Submitted to EURASIP Journal on Advances in Signal Processing March 2011
- [55] Xu, D., J. Principe, J. Fisher and H.C. Wu, (1998) "A novel measure for independent component analysis (ICA)", *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 2 pp1161-1164.
- [56] Zhou W. and J. Gotman, (2004) "Removal of EMG and ECG Artifacts from EEG based on Wavelet Transform and ICA". *26th Annual International Conference of the IEEE EMBS* pp 392-395.

Authors

Janett Walters-Williams received the B.S. and M.S. degrees, from the University of the West Indies in 1994 and 2001, respectively. She is presently a Doctoral student at the University of South-ern Queensland. After working as an assistant lecturer (from 1995), in the Dept. of Computer Studies, in the University of Technology, she has been a lecturer in the School of Computing & Information Technology, since 2001. Her research inter-est includes Independent Component Analysis, Neural Network Applications, signal processing, bioinformatics and artificial intelligence.



Yan Li received the B.E., M. E., and Dr. Eng. degrees from Hiroshima Univ. in 1982, 1984, and 1990, respectively. She has been an associate professor at the University of Queensland since 2008. She is the winner of the 2008 Queensland Smart Woman-Smart State Awards in ICT as well as one of the Head of Department awardees for re-search publications in 2006 and 2008. She is an Australian Reader to assess Australia Research Council Discovery and Linkage Project Proposals an has organized the RSKT 2009 and CME 2010 international conferences. Her research interest includes signal/image processing, independent component analysis, Biomedical Engineering, Blind Signal Separation and artificial intelligence.

