7th WSEAS Int. Conf. on MATHEMATICAL METHODS and COMPUTATIONAL TECHNIQUES IN ELECTRICAL ENGINEERING, Sofia, 27-29/10/05 (pp324-329)

Comparative Evaluation of Decomposition Algorithms based on Frequency Domain Blind Source Separation of Biomedical Signals

Matteo Milanesi[1], Nicola Vanello[1], Vincenzo Positano[2], Maria Filomena Santarelli[2],
Danilo De Rossi[3], Luigi Landini[3]
[1]Department of Electrical Systems and Automation, Faculty of Engineering, University of Pisa, Italy,
[2]CNR Institute of Clinical Physiology, Pisa, Italy, [3]Interdepartmental Research Center "E. Piaggio",
Faculty of Engineering, University of Pisa, Italy.

Abstract. In this paper we compare the performance of different algorithms employed in solving frequency domain blind source separation of convolutive mixtures. The convolutive model is an extension of the instantaneous one and it allows to relax the hypothesis of a linear mixing process in which all the sources are supposed to reach the electrodes at the same time. This test is carried out in the frequency domain, where the algorithms developed for independent component analysis can be employed with minor modifications. The decomposition performance of such algorithms is evaluated on simulated dataset of convultive mixtures of biomedical signals.

Key-Words: Independent component analysis, frequency domain, decomposition algorithms, biomedical signals.

1 Introduction

Biomedical signals are means that transport information regarding one or more biological systems under study. In our data it is necessary to discriminate signals of interest from others that must be considered artifacts: these can have physiological origin, as muscular activity signals superimposed on electrocardiographic registration, or due to acquisition conditions, as movements related artifacts, caused by the displacement of the electrodes.

To recover the signals of interest, the most known methods are linear and nonlinear filtering techniques [1] adaptive signal processing [2] and wavelets based methods [3]. Other techniques, as principal component analysis (PCA) [4] and independent component analysis (ICA) [5] take advantage from multichannel data acquisition. While PCA looks for linearly independent components, ICA starts the search for sources hidden in the data under the hypothesis of statistical independence among them.

The ICA-based model assumes that each electrode measures an instantaneous mixture of signals and both the mixing process and the sources are unknown. Applications in removing artifacts from biomedical signals have been presented in several publications: Barros [6] utilized ICA for removing artifacts from ECG signals; Wachowiak [7] showed how ICA is able to separate artifacts from EMG signal while Jung [8] compared ICA and PCA for EEG artifacts removal. The linear ICA model does not account for the influence, on the mixing process, of the different paths from the signal sources to the sensors, and of the spatio-temporal dynamics of some signals as Anemuller [9] hypothesized for EEG data.

Thus, we used blind separation of convolutive mixtures by means of independent component analysis [5]. We choose to solve this problem in the frequency domain [9, 10] as this allows us to use fast algorithms developed for the instantaneous mixing problem. After decomposition, the signals of interest can be reconstructed back in the original observation space, avoiding some ambiguities introduced by the independent components [11, 12].

Different algorithms were compared throughout the evaluation of a performance index on simulated convolutive mixtures

2 Methods

2.1 Instantaneous ICA model

The basic ICA model assumes that a set of measurements $x_i(t)$ are originated by a linear mixing process of some latent sources $s_j(t)$. By neglecting any signal delay in the mixing process, we can introduce the instantaneous ICA model, mathematically expressed by:

$$x_i(t) = a_{i1}s_1(t) + \ldots + a_{ij}s_j(t) + \ldots + a_{in}s_n(t) \qquad (1)$$

with $i=1,2,...m$, $j=1,2,...,n$ and $t=1,2,...,T$ as we operate with discrete time signals. If we use a vector representation of $\mathbf{x}(t)=[x_1(t),...,x_m(t)]^T$ and $\mathbf{s}(t)=[s_1(t),...,s_n(t)]^T$, we can express equation (1) in matrix notation $\boldsymbol{x(t)=As(t)}$ where $\boldsymbol{A}$ is called the mixing matrix. Both the sources $s_i(t)$ and the mixing process, $\boldsymbol{A}$, are unknown. The hypothesis used in order to extract the original sources is that they are statistically independent. The goal is to estimate a matrix $\boldsymbol{W}$ called the unmixing matrix, such that $\boldsymbol{y(t)=Wx(t)}$ is an estimate of the original sources $\mathbf{s}(t)$. In the following we assume the number of sources equals the number of acquired signals, thus $n=m$.

2.2 Convolutive ICA model in the frequency domain

The basic ICA model (1) assumes that the mixing process is instantaneous, meaning that every single component produced by original sources reaches each sensors at the same time. In some applications this seems to be a too strong assumption since the paths of the signals to each sensor can be different and the finite propagation speed in the medium can generate different time delays.

These considerations lead to introduce a convolution process between the original sources $s_j(t)$ and the elements of the mixing matrix $\boldsymbol{A}$, that becomes the coefficient of unknown *FIR* filters, with impulse response $a_{ij}(t)$ of length L:

$$x_i(t)=\sum_{j=1}^{n}\sum_{k=1}^{L}a_{ij}(k)s_j(t-k) \quad for \quad i=1,2,...,n \qquad (2)$$

Fig.1 shows a scheme of the convolutive mixing process.

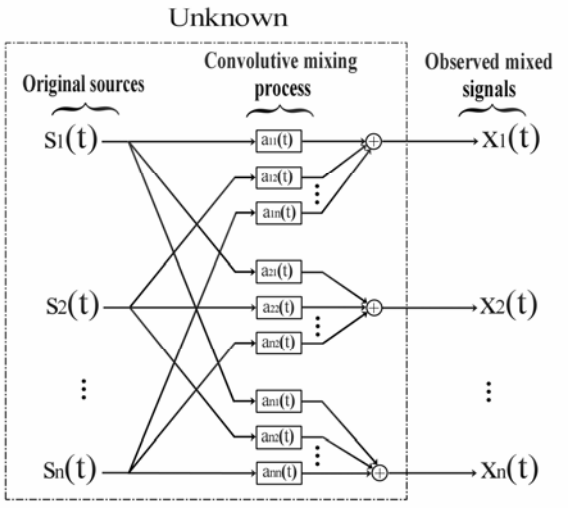

Fig.1. Convolutive ICA model.

The problem can be solved in the time domain using natural gradient methods, or using a frequency domain approach [9, 10, 11, 12]. Since a convolution in time domain is a product in the frequency domain, it is possible to transform the convolutive mixture model into an instantaneous linear mixing model for each frequency bin.

The convolutive model is solved splitting the frequency domain in intervals, i.e. frequency bins, and applying the basic ICA model independently for each frequency bin. A short time Fourier transform (*STFT*) is used to collect a number of observation of mixtures for each bin. The one-dimensional sequence $x_i(t)$, which is a function of a single discrete variable t, is then converted into a two-dimensional function $X_i(f,t)$ of the time variable t and a frequency variable f, both discrete.

The window length, used in the *STFT*, must be greater than the maximum delay occurring in the convolutive process, that is related to the maximum order of the *FIR* filters. After this transformation, we get the following expression:

$$X_i(f,t)=\sum_{j=1}^{n}A_{ij}(f)S_j(f,t) \qquad (3)$$

Now we have as many temporal observations of the signal frequency content as the number of windows we choose. It is possible to split the frequency domain into a number of bins and use the basic ICA instantaneous model for each frequency bin. In this case we assume the statistical independence among temporal observations of the sources frequency contents.

Several approaches have been proposed in order to solve the instantaneous ICA model, like nonlinear decorrelation, maximization of nongaussianity, maximum likelihood estimation methods, infomax principle, minimization of mutual information or some tensorial methods. See [5] for a review.

As pre-processing step, before performing ICA both a removal of the mean value and a whitening operation using PCA is performed. This operation simplifies the estimation of the unmixing matrix $\boldsymbol{W}$ that becomes orthogonal with only $n(n-1)/2$ degrees of freedom instead of n^2.

2.3 Methods for Independent Component Analysis

2.3.1 Maximum Likelihood Approach

Maximum likelihood estimation is used to find the parameters of a model given the data: the computation of the likelihood is easy when the probability density functions are known a priori.

The likelihood of the observed data, given the model ***m*** described in equation (1) and the sources described by parameters ***θ***, can be written as:

$$\boldsymbol{L(m,\theta)} = \prod_{\boldsymbol{j}=1}^{\boldsymbol{T}} \boldsymbol{p(x(j)\,|\,m,\theta)} \tag{4}$$

where T is the number of observations. It can be found that the likelihood is a function of the unmixing matrix $\boldsymbol{W}$ and can be written as

$$\boldsymbol{\log L(W)} = \sum_{\boldsymbol{j}=1}^{\boldsymbol{T}} \sum_{\boldsymbol{i}=1}^{\boldsymbol{n}} \boldsymbol{p_i(w_i^T x)} + \boldsymbol{T}\log(|det(\boldsymbol{W})|) \tag{5}$$

where $\boldsymbol{w_i}$ are the columns of $\boldsymbol{W}$ so that $\boldsymbol{w_i^T x}$ is an estimate of s_i.

In order to maximize the log-likelihood function Bell and Sejnowski [13] proposed a gradient approach such that at each iteration t the unmixing matrix $\boldsymbol{W}$, starting from an initial random guess, is updated as follows

$$\mathbf{W}(t) = \mathbf{W}(t-1) + \mu \frac{1}{T} \frac{\partial \log L(\mathbf{W})}{\partial \mathbf{W}} \tag{6}$$

where μ is the learning rate: the unmixing matrix at each step is changed to maximize the cost function, i.e. the log-likelihood of the data. The derivative in the right side can be written as

$$\frac{1}{T} \frac{\partial \log L(\mathbf{W})}{\partial \mathbf{W}} = \left[\mathbf{W}^T\right]^{-1} + f(\mathbf{W}\mathbf{x})\mathbf{x}^T \tag{7}$$

where the nonlinear function $\boldsymbol{f(\cdot)}$ is used to parameterize the probability density functions of the unknown sources: in particular $f(s) = (f_1(s_1),...,f_n(s_n))$ is a vector of functions whose elements are $f_i(s_i) = \frac{\partial}{\partial s_i} \log p_i(s_i)$.

2.3.2 Natural gradient

Amari et al [14] introduced an approach related to the principle of relative gradient that simplifies the likelihood maximization approach. This method brings to the Natural Gradient algorithm that can be obtained by multiplying the right hand side of equation (7) by $\boldsymbol{W^TW}$. It is expressed by the following learning rule :

$$\mathbf{W} \leftarrow \mathbf{W} + \mu\left[\mathbf{Z} + f(\mathbf{y})\mathbf{y}^T\right]\mathbf{W} \tag{8}$$

where μ is the learning rate, $\boldsymbol{Z}$ can be $\boldsymbol{I}$ or $diag(f(\mathbf{y})\mathbf{y}^T)$ and $f(.)$ is a nonlinear function related to the probability density function of the sources we are interested in. This procedure must be repeated until convergence is reached and it can be implemented in on line and batch versions. The nonlinearities proposed for the natural gradient algorithm are $f(\mathbf{y})=-2\tanh(\mathbf{y})$ for supergaussian components and $f(\mathbf{y})=\tanh(\mathbf{y})-\mathbf{y}$ or $f(\mathbf{y})=-\mathbf{y}^3$ for subgaussian distributed ones.

2.3.3 Maximization of Nongaussianity

Another approach to find the independent components is based on the maximization of nongaussianity. Negentropy, defined in equation (9), is employed as measure of non gaussianity.

$$J(\mathbf{y}) = H(\mathbf{y}_{gauss}) - H(\mathbf{y}) \tag{9}$$

where $H(\mathbf{y})$ is the entropy of the $\boldsymbol{y}$ variable and $\boldsymbol{y}_{gauss}$ is a Gaussian variable with the same covariance as $\boldsymbol{y}$. In order to evaluate negentropy, higher order cumulants can be approximated by means of expectation of non quadratic functions $F(\mathbf{y})$. With this approximations equation (9) can be written as $J(\mathbf{y}) \propto [E\{F(\mathbf{y})\} - E\{F(\mathbf{v})\}]^2$ where ν is a gaussian distributed variable with the same variance as $\boldsymbol{y}$ and $\boldsymbol{E}\{\cdot\}$ is the expectation operator. Hyvarinen [15] proposed a fixed point algorithm for performing ICA of instantaneous mixtures, known as FastICA.

The learning rule employed by FastICA for the research of the independent components is:

$$w_i \leftarrow E\left\{\mathbf{x}f\left(w_i^T\mathbf{x}\right)\right\} - E\left\{f'\left(w_i^T\mathbf{x}\right)\right\}\mathbf{w} \tag{10}$$

where $\boldsymbol{f(\cdot)}$ is a nonlinear function used in order to take into account higher order statistics of the data. This learning rule is applied, at each algorithmic step, to each column of $\boldsymbol{W}$ and is followed by a symmetric Gram-Schmidt orthogonalization of $\boldsymbol{W}$.

The nonlinear function can be chosen among $f(\mathbf{y}) = \tanh(a_1\mathbf{y})$, $f(\mathbf{y}) = \mathbf{y}\exp(-\mathbf{y}^2/2)$ or $f(\mathbf{y}) = \mathbf{y}^3$. The FastICA algorithm can separate components belonging to different probability density functions, like both super- and sub-gaussian with the same nonlinear functions.

The fixed point scheme can be applied also to maximize the likelihood. From equation (10), it is possible to derive another symmetric fixed-point algorithm:

$$\mathbf{W} \leftarrow \mathbf{W} + \mathbf{D}\left[diag(-\alpha_i) + E\left\{f(\mathbf{y})\mathbf{y}^T\right\}\right]\mathbf{W} \tag{11}$$

where $\alpha_i = -E\{y_i f(y_i)\}$ and $\mathbf{D} = -1/(\beta_i + E\{f'(y_i)\})$ with $\beta_i = -E\{y_i f(y_i)\}$.

This last learning rule seems very similar to the natural gradient algorithm but, instead of a constant learning rate μ, an optimal step size ***D*** is applied. Moreover the term α_i accounts for the super or sub-gaussianity of the independent components.

We have to point out that all the above algorithms must be modified since we are working in frequency domain and we are dealing with complex numbers: this means that the transposition operator must be changed into an Hermitian operator and the nonlinearity function must be adapted to the frequency domain too. For example $f_1(\mathbf{y})=tanh(\mathbf{y})$ and $f_2(\mathbf{y})= tanh(\mathbf{y})-\mathbf{y}$ become in the complex domain [10]:

$$\begin{aligned} f_1(\mathbf{y}) &= \tanh(\mathrm{Re}(\mathbf{y})) + j\tanh(\mathrm{Im}(\mathbf{y})) \\ f_2(\mathbf{y}) &= \tanh(\mathrm{Re}(\mathbf{y})) + j\tanh(\mathrm{Im}(\mathbf{y})) - \mathbf{y} \end{aligned} \qquad (12)$$

The centering and prewhitening operations are not altered working with complex numbers.

3 Simulation experiments

Three frequency domain ICA algorithms were tested on simulated convolutive mixtures of biomedical signals downloaded from the PhysioNet database [16], which is standard for testing ECG algorithms. The methods under analysis were the ones described in equations (8), (10), (11) discussed in paragraphs 2.3.2 and 2.3.3. For each method the nonlinearities seen in equation (12) were used. The sources vector $\mathbf{s}(t)$ is composed by real and noise-free ECG, $s_1(t)$, and EMG, $s_2(t)$. The EMG recording is the surface registration of the chin muscles activity and the ECG signal is the precordial lead V5. Both the signals were sampled at 250 Hz. The FIR filters $a_{ij}(k)$ of the mixing matrix ***A***(t), introduced in equation (2) represent the effects produced by each source $s_j(t)$ in the detected signals $x_i(t)$. The elements of the mixing matrix were designed to take into account possible time delays that may occur from the source origins to the electrode and the acquisition specifics of each channel. Hence, $a_{11}(t)$ was a 20 coefficient low pass filter with cut off frequency $f_1=50Hz$ followed by 10 zeros, while $a_{12}(t)$ was a time shifted version of $a_{11}(t)$ with lower gain. In the same way $a_{22}(t)$ was a high pass filter with cut off frequency $f_2=10Hz$ and $a_{12}(t)$ was a time shifted version of $a_{22}(t)$ with higher gain.

Fig.2 shows a frame of 4 seconds of the original sources $\mathbf{s}(t)$, and the detected signals $\mathbf{x}(t)$, obtained by applying equation 2 to $\mathbf{s}(t)$ with the filters described above.

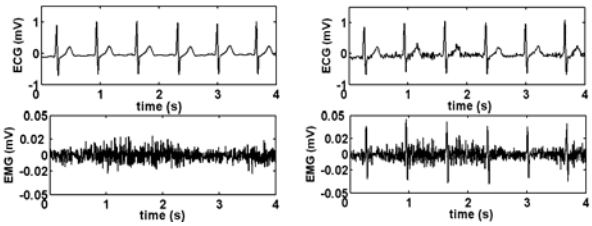

Fig.2. (a) Source signals: noise-free ECG and EMG recording. (b) Simulated convolutive mixtures

The convolutive mixtures were analyzed by the frequency domain approach. A STFT was applied to the data matrices $\mathbf{x}(t)$, using a Hamming window of the same length of the FIR filter while the overlap between each window was 90%. This allows the algorithm to have a resolution $\Delta f = 1/(N * T_c)$, where N is the number of points of the FIR filters and T_c the sampling period. The ICA analysis was carried out in the frequency bins included in the interval between f_1 and f_2, that is were the two signals really overlap. Outside this interval the simulated acquired channels are left unchanged. The total number of frequency bins included in the analysis, with a frequency resolution of Δf is six.

In general, the accuracy of time domain ICA algorithms can be measured using the performance index expressed in [5] for a *n*x*n* instantaneous mixing matrix ***A***:

$$E = \sum_{i=1}^{n}\left(\sum_{j=1}^{n}\frac{|p_{ij}|}{\max_k |p_{ik}|} - 1\right) + \sum_{j=1}^{n}\left(\sum_{i=1}^{n}\frac{|p_{ij}|}{\max_k |p_{kj}|} - 1\right) \qquad (13)$$

where p_{ij} is the *ij-th* element of the matrix ***P***=***WA***.

P would be a permutation matrix in the ideal case of perfectly separated sources. The value of E is always positive and it increases as statistical performance of a separation method grows worse. The minimum is zero and is achieved when ***P*** is a permutation matrix.

The presence of a normalization factor, together with the absolute value operator, assures that there are not any scale and phase indeterminacies: thus this error index can be extended to the complex domain without any modifications. This error index was employed to evaluate the separation capability of different ICA algorithms in each frequency bin where the analysis was carried out. We obtained a number of matrices ***P*** equal to the number of the points in which the discrete time Fourier transform

of the matrix ***A*** was represented, that was the same of the original FIR filters length.

Before starting with independent components research, a whitening step was performed as explained in the previous section. After each iteration, a symmetric Gram-Schmidt orthogonalization was applied to the unmixing matrix.

Since ICA algorithms look for the unmixing matrix starting from initial random guesses, the results may be depend upon this initial value: in order to have statistical reliability, the experiment was repeated 30 times for each bin, changing every time the initial value of the unmixing matrix. Whiskers graphs of the accuracy were realized, as depicted in Fig.3. The rectangular boxes have lines at the lower, median, and upper quartile values. The whiskers extends to the most extreme data value within 1.5 times the height of the rectangle of the box. Outliers are data with values beyond the ends of the whiskers and are denoted by small crosses. The number of iterations was fixed to 50 for each trial.

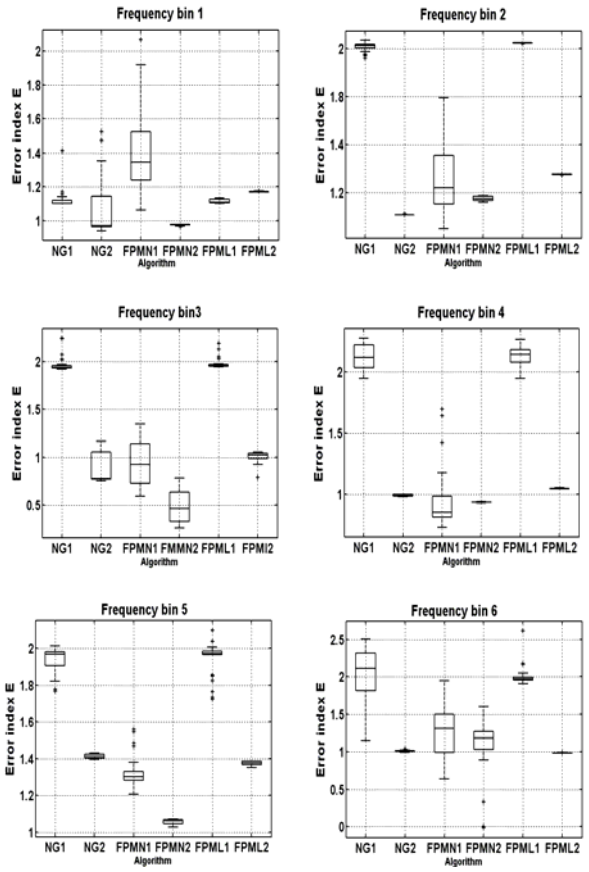

Fig.3. Error indexes of the algorithms in a set of frequency bins. The abbreviation NG stands for natural gradient algorithm, equation (8); FPMN for fixed point algorithm for maximization of nongaussianity, equation (10); FPML for fixed point for maximum likelihood estimation, equation (11). The suffix 1 means that we are employing $f_1(\mathbf{y})$ as non linearity; while the suffix 2 stands for $f_2(\mathbf{y})$.

4 Discussion

The performance index expressed in equation (13) and estimated for the different algorithms is related to the accuracy in solving the convolutive mixtures model. The algorithms performances were tested in each frequency bin taking into account both the mean value of the index and its variability achieved from different trials on the same dataset: this step is necessary because of the random initial guess of the algorithms.

The algorithm, that showed the best stability in the results, was the fixed point for the estimation of the maximum likelihood, independently from the nonlinearity used.

The best performance, in almost every frequency bins, were exhibited by the fixed point algorithm for maximization of nongaussianity, using the non linearity $f_2(\mathbf{y})$: this algorithm seems to slightly outperform the fixed point for the estimation of the maximum likelihood and the natural gradient one, both with non linearity given by $f_2(\mathbf{y})$. The worst results were shown by the natural gradient algorithm and fixed point for maximum likelihood estimation with nonlinearity given by $f_1(\mathbf{y})$.

Note that the source signals under examination were an ECG and an EMG recording. The results seem to indicate that the STFT of the underlying sources are well separated using a nonlinearity suited for subgaussian distributed components, that is $f_2(y)$. The only exception is the fixed point algorithm for maximization of nongaussianity that works quite well independently from the nonlinearity, that can be almost any smooth function, as stated in [5].

5 Conclusions

An algorithm for blind source separation of convolutive mixtures of biomedical signals was introduced: the algorithm works in the frequency domain exploiting the ICA instantaneous model in each frequency bin.

Several algorithms for instantaneous ICA were tested to individuate the one giving the best performances within the convolutive mixtures separation approach proposed. The source signals under examination were an ECG and an EMG recording. An index of accuracy for the different algorithms was suggested. The results seems to indicate that best performance was achieved by the fixed point algorithm for maximization of nongaussianity. Both the methods for maximum likelihood estimation given by Amari's natural

gradient algorithm and by Hyvarinen's fast fixed point approach, give very similar result, if a nonlinearity for estimation of subgaussian components is employed.

Acknowledgments

This work was supported by E.U. project MyHeart-IST-2002-507816.

References

[1] N. B Eugene, Biomedical signal processing and signal modelling, Wiley-Interscience, 2001.

[2] V. Almebar, A. Albiol, A New Adaptive Scheme for ECG Enhancement, *Signal Processing,* Vol. 27, 1999, pp. 253-263.

[3] S. Kadambe, R. Murray, G.F. Bourdeaux-Bartels, Wavelet Transform-based QRS Complex Detector, *IEEE Transactions on Biomedical. Engineering*, Vol 46, 1999, pp. 838–847.

[4] H. Hotelling, Analysis of a complex of statistical variables into principal component analysis, *J. Edu. Psychol*, Vol. 24, 1933, pp. 417-421.

[5] A. Hyvarinen, J. Karhunen, E. Oja, *Independent component analysis*, John Wiley & Sons, 2001.

[6] A. K. Barros, A. Mansour, N. Ohnishi, Removing artifacts from electrocardiographic signals using independent component analysis, *Neurocmputing*, Vol 22, 1998, pp. 173-186.

[7] M. Wachowiak, R. Smolikovà, G. D. Tourassi, A. S. Elmaghraby, Separation of Cardiac Artifacts from EMG Signals with Independent Component Analysis, *Biosignal* 2002.

[8] T. P. Jung., C. Humphries, T. W. Lee, S.Makeig, M.J. McKeown, V. Iragui, T.J. Sejnowski, Removing Electroencephalographic Artifacts: Comparison between ICA and PCA, *Proceeding of the 1998 IEEE signal processing society workshop,* 1998.

[9] J. Anemuller, T.J. Sejnowski, S Makeig, Complex Spectral-domain Independent Component Analysis of Electroencephalographic Data, *Neural Neworks*, vol. 16, 2003, pp. 1311-1323.

[10] P. Smaragdis, Blind Separation of Convolved Mixtures in the Frequency Domain, *Neurocomputing*, Vol. 22, 1998, pp.21-31.

[11] N. Murata, S. Ikeda, A. Ziehe, An Approach to Blind Source Separation based on Temporal Structure of Speech Signals, *Neurocomputing*, Vol. 41, 2001, pp. 1-24.

[12] M. Milanesi, N. Vanello, V. Postano, M.F. Santarelli, D. De Rossi, L. Landini, An Autoamtic Method for Separation and identification of Biomedical Signal from Convolutive Mixtures by Independent Component Analysis in the Frequency Domain, Proceeding of the 5th WSEAS Int. Conf. on SSIP, 2005, pp. 74-79.

[13] A. J. Bell, T. J. Sejnowski, An Information-Maximization Approach to Blind Separation and Blind Deconvolution, *Neural Comput*., Vol 7, 1995, pp. 1129–1159.

[14] S. I. Amari, A, Cichocki, H.H. Yang, A new learning algorithm for blind source separation, *Advances in Neural Information Processing Systems*, Vol. 8, 1996, pp. 757-763.

[15] A. Hyvärinen, A Fast Fixed-Point Algorithm for Iindependent Component Analysis, *Neural Computation,* Vol.9, 1997, pp. 1483-1492.

[16] A.L. Goldberger, L.A.N. Amaral, L. Glass, J.M. Hausdorff, P.Ch. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.K. Peng, H.E. Stanley, PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals, *Circulation,* Vol. 101, pp. 215-220.