OCULAR ARTIFACTS REMOVAL IN SCALP EEG: COMBINING ICA AND WAVELET DENOISING

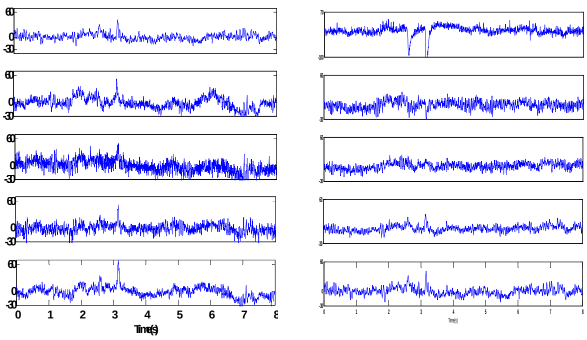
Rebeca Romo-Vázquez, Radu Ranta, Valérie Louis-Dorr, Didier Maquin
CRAN - UMR 7039, Nancy - University, CNRS
2 Avenue de la Forêt de Haye, 54506, Nancy, France
rebeca.romo-vazquez@ensem.inpl-nancy.fr

ABSTRACT

In this communication we present the first results of a project whose goal is to remove artifacts from electroencephalographic epileptic signals. More precisely the present objective is to remove ocular (blinking) artifacts in simulated and real EEG signal using Independent Component Analysis and wavelet denoising algorithms.

1.INTRODUCTION

Epilepsy is one of the most common brain disorders. It is characterized by repeated seizures, which range from the shortest lapse in attention to severe, frequent convulsions. They can occur from several times a day to once every few months. The seizures are caused by bursts of excessive electrical activity in the brain. This electrical activity is measured by the electroencephalogram (EEG), a recording of potential changes on the scalp caused by brain activity. The EEG is the main method of putting in evidence the epileptic activity of the brain.



(a) (b)

Fig 1. Example of EEG

In EEG recordings the sensors are placed on the scalp within a few centimetres of each other. Therefore, the sensors record overlapped brain activity transmitted by volume conduction from different dynamic neocortical processes. An example of normal EEG (5 channels in average reference montage) is shown in figure 1(a).

EEG channels also record other signals, such as noise or artifacts, supposed independent from brain processes. These perturbations overlap with neural brain activity and may be present in all sensors. Obviously they increase the difficulty of EEG interpretation. Therefore, a useful tool would be a method able to remove noise and external artifacts as eye or muscle activities.

One current hypothesis is that this artifacts are independent from brain activity, either normal or pathologic. Under this hypothesis and considering these signals non Gaussian, a frequently used method is blind source separation (BSS) by independent component analysis (ICA) [1,2]. Using ICA as a tool to blindly separate overlapping EEG signals and artifacts into independent components, one can eliminate artifact sources and reconstruct "proper" EEG recordings, more easily used for further analysis.

This paper is structured as follows: after this brief introduction, the second section is dedicated for the presentation of the signal processing methods, we have used: the first subsection presents source separation algorithms (more detailed in section on the algorithm was chosen), the second one presents wavelet denoising techniques and the last subsection introduces the evolution crieteria we have used to test the selected algorithms.

The third section of the paper presents our results, mainly on simulated signals, and justifies the choices we have made concerning the algorithms. This section ends with an example of real EEG processing. The last section concludes this paper and indicates same possible perspectives for the future work.

2. METHOD

2.1 Source separation

The goal of blind source separation is to recover independent sources given only sensor observations. This sensor observation are modelled as linear mixtures of independent source signals. The term blind indicates that both the source signals and the way the signals were mixed are unknown. Several algorithms for BSS were developed in the last 15 years. Some of the most important are described in [2] and implemented in ICALAB [14] toolbox under MATLAB.

2.1.1 Independent Component Analysis

Independent Component Analysis (ICA) is a method for solving the blind source separation problem. It is a way to find a linear transformation of the measured sensor signals such that the resulting source signals are as statistically independent from each other as possible. ICA not only decorrelates the signals (2^{nd} order statistics) but also reduces higher – order statistical dependencies [1].

The mixing ICA model can be represented as:

$$\mathbf{x(k)=As(k)+n(k)} \qquad (1)$$

where $\mathbf{x}$ is an N-dimensional vector containing the mixed signals (sensors), $\mathbf{A}\in\mathbf{R}^{NxN}$ is the unknown nonsingular mixing matrix, $\mathbf{s}$ is an N-dimensional vector of independent source signals. The vector $\mathbf{n}$ is an additive noise assumed to be zero mean, temporally white and independent from the source signals (k being the time index after sampling).

The goal of ICA is to find a linear transformation $\mathbf{W}$ of the dependent sensor signals $\mathbf{x}$ that makes the outputs as independent as possible:

$$\mathbf{y(k)=Wx(k)=WAs(k)} \qquad (2)$$

where **y** is an estimate of the sources. The sources are exactly recovered when **W** is the inverse of **A**.

As it was pointed out by different authors ([1, 2]), obtaining the exact inverse of the **A** matrix is in most of the cases impossible, therefore source separation algorithms aim to find a **W** matrix such as the product **WA** should be a permuted diagonal and scaled matrix. Consequently, sources can be recovered up to their order (permutation) and their amplitude (scale).
Different types of algorithms were proposed in the last 10 to 12 years. Most of them suppose that the sources are stationary and are based explicitly or implicitly on high order statistics (HOS) computation. Therefore, Gaussian sources cannot be separated, as they don't have higher than 2 statistic moments. Another type of algorithms does not make the stationarity hypothesis, and uses the non stationary structure of the signals (i.e. their time or frequency structure) to separate them. These methods use second order statistics (SOS) only, and they are called SOS algorithms. As EEG signals are highly non stationary, this type of algorithms is the most widely used. We briefly introduce in the next section these algorithms and we describe more precisely the specific method that we have used.

2.1.2 Second Order Statistics Algorithms

Temporal, spatial and spatio-temporal decorrelations play important roles in EEG/MEG data analysis. These techniques are based only on second-order statistics (SOS). They are the basis for modern subspace methods of spectrum analysis and array processing and are often used in a preprocessing stage in order to improve convergence properties of adaptive systems, to eliminate redundancy or to reduce noise [9].
The simplest SOS algorithm is spatial decorrelation or whitening, which is often considered a necessary condition for stronger stochastic independence criteria. In fact, whitening (or data sphering) is an important pre-processing step in a variety of BSS methods. After whitening, the BSS or ICA tasks usually become somewhat easier and well-posed, because the subsequent separating (unmixing) system is described by an orthogonal matrix for real-valued signals and a unitary matrix for complex-valued signals and weights.
Based on the same second-order statistic, for non-stationary signals one can compute different whitening transformations. Using these different transforms, one can obtain spatio-temporal and time-delayed decorrelation, which can be used to identify the mixing matrix and to perform blind source separation of coloured sources [2,11].
On the other hand, conventional whitening exploits the equal-time correlation matrix of the data **x**, so that the effect of additive noise can not be removed. A more robust whitening method lies in utilizing time-delayed correlation matrices that are not sensitive to the white noise [2,9,13].

*2.1.3 Robust SOB*I

Recall the model introduced by (1). In a second order statistics framework, source signals **s** are assumed to be mutually uncorrelated and temporally correlated (instead of independents). Computing a separating matrix on this model can be difficult because of the noise, which influences the correlation between the signals. Contrary to the sources, no assumptions are made on the distribution or the spatial correlation properties of the noise vector **n**, which is nevertheless considered white (individual components of the vector are not autocorrelated). Hence, its covariance matrix at lag 0 $\mathbf{R_n(0)= E[n(k)n(k)^T]}$, can be a full matrix which is generally unknown, while any time delayed correlation matrix $\mathbf{R_n(i)=E[n(k)n(k\text{-}i)^T]}$ will be null. Given the above assumptions, the correlation matrices of the observation have the following structure [2,12]:

$$\mathbf{R_x(0) = E[x(k)x(k)^T] = AR_s(0)A^T + R_n} \qquad (3)$$

$$\mathbf{R_x(i) = E[x(k)x(k-i)^T] = AR_s(i)A^T} \quad \forall i \qquad (4)$$

The robust SOBI (SOBI-RO) algorithm combines robust whitening and time-delayed decorrelation. This algorithm, introduced in [12], improves the classical SOBI method [11] by integrating robust whitening [2,12,13] instead of simple whitening, the main objective being the elimination of the influence of the white noise.

The first step (robust whitening) consists, in the general case, in finding a matrix **Q** that decorrelates the signals in **x** for several (small) time lags. This method is described in detail in [2,12,13]. In our case, we have utilized the ICALAB [14] implementation, which exploits equation (4) for a single time lag *i*=1. Then, the matrix $\mathbf{R_x(1)}$ is diagonalized by an eigen-decomposition:

$$\mathbf{R_x(1) = U_c diag}[\lambda_1^2 ... \lambda_N^2]\mathbf{U_c^T} \qquad (5)$$

The whitening matrix **Q** will be obtained from the eigen-vectors matrix $\mathbf{U_c}$ and a the diagonal eigen-values matrix:

$$\mathbf{Q = diag}[\lambda_1 ... \lambda_n]^{-1}\mathbf{U_c^T} \qquad (6)$$

Using this **Q** matrix, one can compute whitened signal **z(k-i)=Qx(k-i)** for different time lags (the default option under ICALAB is 100 time lags).

The second step of SOBI-RO is the same as in classical SOBI, namely an approximate joint diagonalization of the different $\mathbf{R_z(i)}$ matrices, computed according to equation (4). This *approximate* diagonalization aims to minimize the sum of the squared off-diagonal elements of these matrices [11,13]. The result of this operation will be an orthonormal matrix **V**, and the final estimation for the demixing matrix W will be:

$$\mathbf{W = V^T Q} \qquad (7)$$

2.2 Wavelet denoising

Besides ocular or muscular artefacts (for which we intent to apply BSS methods), real EEG recordings are contaminated with noise. The previously presented robust whitening partially deals with this noise. A nowadays classical solution for noise removal from non-stationary signals is wavelet denoising, which we considered for improving the separation results.
The basic idea is simple: by decomposing the signal on a wavelet basis (discrete wavelet transform DWT), we obtain a representation of the signal that concentrates most of its energy in few wavelet coefficients having large absolute values. On the contrary, noise energy distribution doesn't change (for noises modelled as random uncorrelated processes), which means that its energy will not be retained by large value coefficients.

Consequently, performing a partial reconstruction of the signal using only these large coefficients (by inverse DWT) leads to an almost noise-free version of the signal (see [3,15] for a broader presentation of wavelet denoising framework). The main problem is choosing the threshold, which means responding to the question: where to fix the frontier between small and large wavelet coefficients?

Tens of algorithms have been proposed in the last years, the most well known being Donoho's universal thresholding [5]. This algorithm computes a threshold which, asymptotically, ensures that no gaussian noise will be left in the denoised signal. The first consequence is an apparently noise-free signal, visually very satisfactory (hence the VisuShrink name given to the algorithm), but a sometimes important drawback is the elimination of possibly informative parts of the signal.

In EEG case, it is of great importance not to loose information potentially useful to medical diagnosis. Moreover, EEG informative signals often have small amplitudes, and their wavelet coefficients can have rather low values. Therefore, a high threshold as proposed by VisuShrink is not appropriate.

Two other methods seemed adapted to our case: SURE denoising [5] and minimal iterative denoising [3], the two of them offering low thresholds (and thus preserving most of the signal but eliminating less noise).

The first one (as the universal thresholding also) propose a one step thresholding, the value of the threshold being computed considering a Gaussian noise hypothesis for which a robust estimation of variance is made. The SURE (Stein Unbiased Risk Estimator) method has the important property that it adapts itself to the signal (i.e. the threshold depends on the signal, not only on the estimated noise, as for universal thresholding).

Iterative minimal denoising does not make any assumptions on the noise and it adapts itself to the distribution of the wavelet coefficients, modelled as a generalized Gaussian. Unlike the previous methods, the threshold is obtained by iterative outlier detection, the convergence of the algorithm being assured and depending on the shape of the wavelet coefficients distribution. Its main utility was proved for sparse signals, which is not always the case in EEG.

These methods were tested on the same simulated signals (for completeness, universal thresholding was also tested, but the results are not reported here). The evaluation of their results was done by two methods: the first one oriented strictly on the denoising aspect, the second one considering the whole processing chain (denoising plus source separation). The evaluation criteria are detailed in the next section.

2.3 The evaluation criteria

For denoising algorithms evaluation, we have used the classical criterion of the mean squared error (*MSE*) between the original signals and their denoised versions, for the x_i components of the vector $\mathbf{x}$ (i^{th} sensors, i=1…N), we have :

$$MSE_i = \frac{1}{M}\sum_{k=1}^{M}\left(x_i(k) - \hat{x}_i(k)\right)^2 \tag{8}$$

where $\hat{x}_i(k)$ is the denoised signal and M is the length of the signal. As we are in a multi-channel set-up (multiple simulated sources mixed to obtain the same number of recordings), the denoising quality criterion was the average *MSE* for the N signals (MSE_{avg}).

To validate the separation method, we have chosen two criteria:

- a first one computed on the signals: the mean correlation (r) between the simulated sources ($s(k)$) and the estimate independent components ($y(k)$).

$$r_{sy} = \frac{\mathrm{cov}(s,y)}{\sigma_s \sigma_y} \tag{9}$$

For this criterion, we have computed the correlation between each estimated source and all of the original sources, and we have chosen the maximum value. In this way, each estimated source corresponds to one original source, and the value of the correlation coefficient between the two of them is retained. The mean value (over the N sources) of the retained correlation coefficients gives the average correlation coefficient r_{avg}.

- a second one computed on the mixing-demixing system: the index of separability *IS* [2]. The index is computed form the $N \times N$ transfert matrix $\mathbf{G}$ between the original sources and the estimated ones after separation:

$$\mathbf{G} = \mathbf{WA} \tag{10}$$

In order to obtain the *IS* is necessary to take the absolute value of elements of $\mathbf{G}$ and to normalize the lines $\mathbf{g_i}$ by dividing each element by the maximum absolute value of the line. The lines of the resulting matrix $\mathbf{G'}$ will be:

$$g_i^{'} = \frac{|g_i|}{\mathbf{max}|g_i|} \tag{11}$$

The separability index is obtained by:

$$IS = \frac{\sum_{j=1}^{N}\left(\sum_{i=1}^{N}\left(\mathbf{G}'(i,j)\right) - 1\right)}{\left[N(N-1)\right]} \tag{12}$$

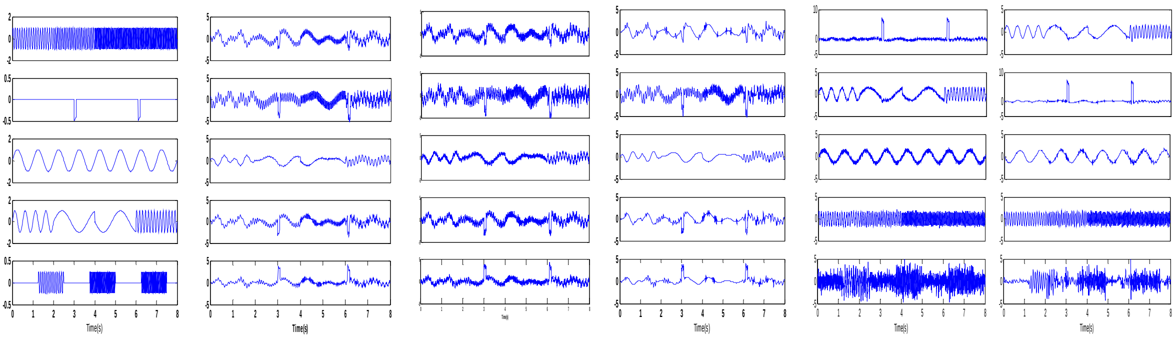
For perfect separation, the two criteria should have the values r_{avg}=1 and *IS*=0.

3. RESULTS

3.1 Simulated EEG

To choose the most appropriate algorithm for our application we decided to test the algorithms with simulated signals in order to compare the results with known reference sources. We created 4 simulated sources having frequencies close to the real brain signals and one ocular blinking artefacts source (fig. 2(a)). These sources were mixed using 5 random matrices to obtain signals similar to the EEG, in frequency and form (fig.2(b)). Three types of test were performed:

1. The first one works directly on the simulated EEG signals (figure 2(b)). The goal of this test was to choose a source separation algorithm in ideal conditions (no noise). Twenty ICA algorithms were tested. The best 10 results (considering the previous defined separation criteria) are presented table I.



(a) (b) (c) (d) (e) (f)

Fig 2. Simulated signals and results

Algorithms	r_{avg}	IS
AMUSE	0,996032	0,06
Evd2	0,99378	0,07
Evd 24	0,97204	0,18
SOBI	0,989832	0,07
SOBI - RO	**0,998324**	**0,0098**
SOBI-BPF	0,985576	0,02
Jade td	0,99434	0,13
FASTICA tan	0,996772	0,088
SANG	0,998512	0,08
ThinICA	0,99816	0,09

Table I. Comparison of algorithms in simulated signals

From the results of the table I, we decided to use the SOBI-RO algorithm to remove the ocular artifacts. This result confirms our bibliographical research, which indicate that second order statistics ICA algorithms perform well on non-stationary EEG signals [4,6,7].

2. In order to better approximate real EEG signals, we added different types of noise (Gaussian and Uniform white noise with signals to noise ratios from 20 dB to 0 dB).

Besides the two separation criteria (r_{avg} and *IS*), we have also considered the denoising quality criterion (MSE_{avg}). Several simulations were performed for different random noises. The mean results (over all simulations) are presented table II and table III. The figure 2(c) show the simulated noisy signals and 2(e) the independent components without denoising.

	r_{avg}	IS	MSE_{avg}
20dB	0,8904	0,0725	0.5931
15dB	0,8243	0,0814	0.1875
10dB	0,7295	0,1086	0.0593
5dB	0,6014	0,1473	0.0188
0dB	0,4552	0,1757	0.0059

Table II. Separation of noisy signals (Gaussian noise)

	r_{avg}	IS	MSE_{avg}
20dB	0,8927	0,0919	0.0492
15dB	0,8106	0,1064	0.0156
10dB	0,7302	0,1074	0.0049
5dB	0,6088	0,1339	0.0016
0dB	0,4531	0,1733	0.0005

Table III. Separation of noisy signals (Uniform noise)

3. Finally, we introduced a wavelet denoising step before the ICA. Two denoising methods were tested, SURE [5] (the results are presented table IV and table V), and iterative minimal denoising (table Vl and table Vll) [3]. Both of them propose low denoising thresholds, and therefore ensure a minimal distortion of the informative signal. An example independent components after wavelet denoising is presented figure 2(f).

	r_{avg}	IS	MSE_{avg}
20dB	0,9137	0,0837	0.1952
15dB	0,8705	0,1267	0.0730
10dB	0,7981	0,1569	0.0287
5dB	0,7223	0,1729	0.0107
0dB	0,6029	0,1805	0.0037

Table IV. Separation of denoised signals (SURE, Gaussian noise)

	r_{avg}	IS	MSE_{avg}
20dB	0,9241	0,0880	0.0186
15dB	0,8683	0,1376	0.0073
10dB	0,8339	0,1444	0.0025
5dB	0,7707	0,1656	0.0008
0dB	0,6542	0,1922	0.0003

Table V. Separation of denoised signals (SURE,Uniform noise)

	r_{avg}	IS	MSE_{avg}
20dB	0,9122	0,1062	0.1786
15dB	0,8708	0,1298	0.1015
10dB	0,7787	0,2008	0.0447
5dB	0,6781	0,2113	0.0166
0dB	0,5724	0,2252	0.0088

Table Vl. Separation of denoised signals (Minimal, Gaussian noise)

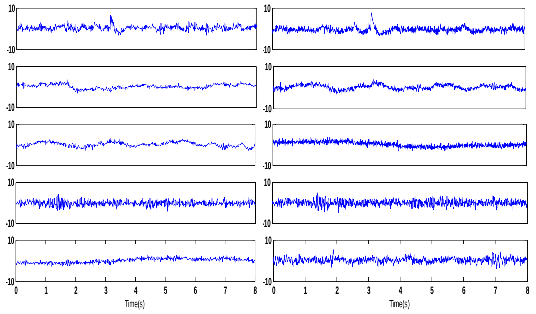
	r_{avg}	IS	MSE_{avg}
20dB	0,9117	0,1058	0.0368
15dB	0,8727	0,1327	0.0147
10dB	0,7832	0,1918	0.0080
5dB	0,6860	0,2089	0.0044
0dB	0,5918	0,2270	0.0030

Table Vll. Separation of denoised signals (Minimal, Uniform noise)

3.2. Application on real EEG

EEG signals were recorded using 10/20 international system and average reference montage[16]. We considered 8 seconds of EEG, sampled at 256 Hz. Source separation using SOBI-RO algorithm was performed on the 24 channels, both with and without wavelet denoising.

The 24 “proper” (i.e. without blinking artifact) EEG channels were reconstructed by multiplying the estimated sources with the inverse of the separation matrix. Figure 1(b) shows independent components real EEG signals before denoising. Figure 3, presents the same estimated source signals using a denoising step before the separation: iterative minimal denoising (a) and SURE denoising (b).



(a) (b)

Fig 3. Separated sources on denoised EEG

4. CONCLUSION AND PERSPECTIVES

We can conclude that the algorithm SOBI-RO works well in EEG source separation and particularly for identifying eye artifacts. This conclusion is supported by the bibliography, by the simulated results and by our first tests on real EEGs.

It is more difficult to evaluate the role of the denoising step: our first criterion (correlation with the original sources) shows better performance, but the index of separability is lower. Further tests are needed to decide if this step is useful, both on simulated signals and mainly on real EEGs. Obviously, medical validation is necessary in this latter case Considering the denoising methods, the SURE algorithm has better performances than the iterative minimal denoising. This conclusion is supported by both the *MSE* criterion and the two separation criteria.

This work has several perspectives: first of all, we intent to test other methods of denoising, using different types of simulated signals and noises (coloured, other distributions).

The joint diagonalization that we have used in this paper considers a fixed number of time-lags, but another way of using it is to compute correlation matrices on different windows instead of time lags (including ictal periods). Moreover, this type of method can be applied considering decorrelation in different frequency bands instead of time lags or windows.

Another possibility is to perform ICA to separate the wavelet transforms of the recorded signals instead of the actual signals.

On real EEG signals, our goal is to facilitate analysis algorithms (using correlation and coherence) developed in our team for lateralization and classification [8]. The results of these 2 methods will serve as a validation for our source separation and denoising algorithms. Of course, the final objective is to obtain medical validation.

5. REFERENCES

[1] Te-Won Lee, Independent component Analysis Theory and Applications, Kluwer Academic Publisher, Boston, 1998.

[2] A. Cichocki, Shun-ichi Amari, Adaptative blind Signal and Image Processing Learning Algorithms and Applications, John Wiley & Sons, ltd, 2002.

[3] R. Ranta, V. Louis-Dorr, Christian Heinrich, D. Wolf, “Iterative wavelet-based denoising methods and robust outlier detection ” IEEE Signal Processing Letters pages 557-560 vol. 12 num. 8 - august 2005.

[4] Joep J. M. Kierkels, Geert J. M. Van Botel, and Leo L. M. Vogten. ‘A Model-Based Objective Evaluation of Eye Movement Correction in EEG Recordings’IEEE, Transactions on biomedical engineering, vol, 53, No. 2, February 2006.

[5] D. Donoho and I. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," J. Amer. Statist. Assoc., vol. 90, pp. 1200--1224, Dec. 1995.

[6] C.W. Hesse and C.J. James, “Seizure tracking and detection in ictal EEG using time-structure based blind source separation methods and prior spatial topographical information” Proceedings of the IFMBE, EMBEC’05, Prague, Czech Republic, November 20-25, 2005.

[7] Sutherland, M.T., and Tang A.C. “ Blind source separation can recover systematically distributed neuronal sources from “resting” EEG”, Proceedings of the Second International Symposium on Communications, Control, and Signal Processing (ISCCSP 2006), Marrakech, Morocco, March 13-15.

[8] M. Caparos, V. Louis-Dorr, F. Wendling, L. Maillard, D. Wolf “Automatic lateralization of temporal lobe epilepsy based on scalp EEG”, Clinical Neurophysiology, 2006, IN PRESS.

[9] S. Choi, A. Cichocki, H. Park, S. Lee, blind Source “Separation and Independent Component Analysis : A Review”, Neural Information Processing – Letters and Reviews, Vol. 6, no. 1, January 2005.

[10] J.F. Cardoso, “Eigen-Strucutre of the fourth-order cumulant Tensor with Application to the blind source separation problem” in Ent. Conference on Acoustics speech and Signal Processing, pages 2655-2658, Albuquerque, NM, USA, April 3-6, 1990.

[11] A. Belouchrani, K. Abed-Meraim, J.F. Cardoso, E. Moulines, “A blind source separation technique using second order statistics”, IEEE Transactions on signal processing, vol. 45, No. 2, February, 1997.

[12] A. Belouchrani, A. Cichocki, “Robust whitening procedure in blind source separation context” Electronics Letters, vol. 36, No. 24, 2000, pp. 2050-2053.

[13] S. Choi, A. Cichocki, A. Belouchrani, “Second order nonstationary source separation”, Journal of VLSI Signal Processing, vol. 32, No. 1-2, pp. 93-104, 2002.

[14] A. Cichocki, S. Amari, K. Siwek, T. Tanaka et al., ICALAB Toolboxes, http://www.bsp.brain.riken.jp/ICALAB.

[15] A. Antoniadis, J. Bigot, T. Sapatinas, “Wavelet estimators in nonparametric regression: a comparative simulation study”, Journal of Statistical Software, vol. 6, no. 6, pp. 1-83 , 2001.

[16] Gil – Nagel, Manual of electroencephalography, McGRAW-HILL, 2002.