APPLICATION OF ICA FOR AUTOMATIC NOISE AND INTERFERENCE CANCELLATION IN MULTISENSORY BIOMEDICAL SIGNALS

Andrzej Cichocki and Sergiy Vorobyov

Lab. for Advanced Brain Signal Processing, Brain Science Institute,
Institute of Physical and Chemical Research (RIKEN),
2-1, Hirosawa, Wako-shi, Saitama, 351-0198, Japan
E-mail: cia(svor)@brain.riken.go.jp

ABSTRACT

Independent Component Analysis (ICA) and related methods like Adaptive Factor Analysis (AFA) are promising novel approaches for elimination of artifacts and noise from biomedical signals, especially EEG/MEG data. However, most of the methods require manual detection and classification of interference components. Main objective of this paper is to detect and eliminate noise and some artifacts automatically by computer using criteria for classification, ordering and detection of noisy and random signals. The automatic detection and on-line elimination of noise and other interferences is especially important for long recordings, e.g. EEG/MEG recording during sleep. In this paper we focus mainly on the problem of 'cleaning' or enhancement of noisy EEG/MEG data from noise and undesired interferences using several techniques: ICA and HOS measure of Gaussianity (to detect and eliminate Gaussian noise), linear predictor (to detect i.i.d. sources and classify temporally structured sources) and Hurst exponent (to detect randomness in independent components and classify independent signals). Preliminary extensive computer simulation confirmed potential usefulness of proposed methods for wide class of applications, especially in area of analysis and processing of EEG/MEG data.

1. INTRODUCTION AND PROBLEM DETAILED ELABORATION

The nervous systems of humans and animals must encode and process sensory information in the context of noise and interference, and the signals which are encoded (the images, sounds, etc.) have very specific statistical properties. One of the challenging task is how to reliably detect, enhance and localize very weak, non-stationary and corrupted by noise brain source signals (e.g., evoked and event related potentials EP/ERP) using EEG/MEG data.

Independent Component Analysis (ICA) and related methods like Adaptive Factor Analysis (AFA) are promising approaches for elimination of artifacts and noise from EEG/MEG data [1] - [4], [15], [17], [19]. However, most of the methods require manual detection and classification of interference components and/or estimation of cross-correlation between each independent components and reference signals corresponding to specific artifacts [14], [19].

Main objective of this paper is to propose some relatively simple techniques to automatically detect and eliminate noise and some artifacts and classify independent 'brain sources'.

Evoked potentials (EPs) of the brain are meaningful for clinical diagnosis and they are important factors to understand higher order mechanism in the brain. The EPs are usually embedded in the ongoing EEG/MEG with signal to noise ratio (SNR) less than 0 dB, making them very difficult to extract using single trial. The traditional method of EPs extraction is by using ensemble averaging to improve the SNR. This often requires hundreds or thousands of trails to obtain a usable noiseless waveform. Therefore, it is important to develop novel techniques that can rapidly improve the SNR and reduce to minimum the number of ensembles (trials). Traditional signal processing techniques, such as Wiener filtering, adaptive noise canceler, latency-corrected averaging [9] and invertible wavelets transform filtering [18] have been recently proposed for SNR improvements and ensemble reduction. However, these methods require a priori knowledge pertaining to the nature of the signal. Since EPs signals are known to be non-stationary, sparse and changing their characteristic from trial to trail, it will be essential in the future to develop novel algorithms for enhancement of single trail EEG/MEG noisy data.

The formulation of the problem could be given in the following form. Denote by $\mathbf{x}(t) = (x_1(t), x_2(t), ..., x_n(t))^\top$ the observed n-dimensional vector of noisy signals that must be cleaned from the noise and interferences. Here we have two types of noise. The first is so called "inner" noise generated by some primary sources which cannot be observed directly but contained in the vector of observations that is mixture of useful signals and random noise signals or other undesirable sources, and second type of noise is the sensor additive noise (observation errors) at the output of measurement system. This noise is not measurable directly also. Formally we can write that observed n-dimensional vector of sensor signals $\mathbf{x}(t) = (x_1(t), x_2(t), ..., x_n(t))^\top$ is mixture of source signals plus observation errors:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{v}(t), \tag{1}$$

where $t = 0, 1, 2, ...$ is discrete time; $\mathbf{A}$ is a full rank ($n \times m$) mixing matrix; $\mathbf{s}(t) = (s_1(t), s_2(t), ..., s_m(t))^\top$ is m-dimensional vector of sources containing useful signals and

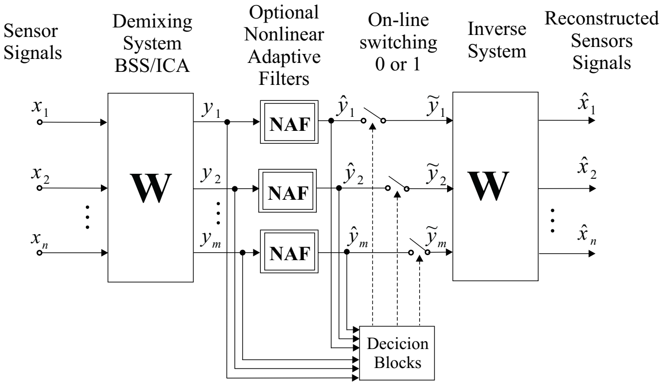

Figure 1: Conceptual model for cleaning multisensory (e.g. EEG/MEG) data.

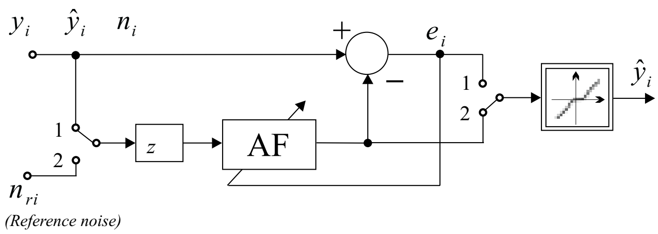

Figure 2: Adaptive filter configured for deconvolution (switches in position 1) and for standard noise cancellation (switches in position 2).

noise and $\mathbf{v}(t)$ is n-dimensional vector of additive white noise. We assume also here that some useful sources could be not necessarily statistically independent. It means that we can not achieve perfect separation of primary sources using any ICA procedure. However, our purpose is not separation of the sources but removing of independent noisy sources.

Let us emphasize that the problem consists in cancellation of the noise sources and reduce observation errors based on only information about observed vector $\mathbf{x}(t)$.

Conceptual model for elimination of noise and other undesirable components from multi-sensory data is depicted in Figure 1. Firstly, ICA is performed using any robust algorithms (in respect to Gaussian noise) [1], [2], [4], [6], [8], [12], [15] by linear transformation of sensory data as $\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t)$, where vector $\mathbf{y}(t)$ represents independent components. However, robust ICA methods allow us only to obtain unbiased estimation of demixing matrix $\mathbf{W}$, but due to memoryless they by definition can not remove the additive noise. This can be done using optional nonlinear adaptive filtering and nonlinear noise shaping (see Figure 2). In the next stage, we classify independent signals $\hat{y}_i(t)$ and remove noise and undesirable components by switching corresponding switches "off".

Projection of interesting or useful independent components (e.g. independent activation maps) $\tilde{y}_i(t)$ back onto the sensors (electrodes) can be done by: $\hat{\mathbf{x}}(t) = \mathbf{W}^{+}\tilde{\mathbf{y}}(t)$, where $\mathbf{W}^{+}$ is pseudo inverse of demixing matrix $\mathbf{W}$. In the typical case, when the number of independent components is equal to the number of sensor $\mathbf{W}^{+} = \mathbf{W}^{-1}$.

The standard adaptive noise and interference cancellation systems may be subdivided into the following classes [7], [9]:

1. Noise cancellation (see Figure 2). This term is normally referred to the case, when we have both a primary signal $y_i(t) = \hat{y}_i(t) + n_i(t)$ contaminated with noise and a reference noise $n_{ri}(t)$, which is correlated with the noise $n_i(t)$ but independent of the primary signal $\hat{y}_i(t)$. By feeding the reference signal to the linear adaptive filter we able to estimate or reconstruct the noise and subtract it from the primary signal thus enhancing the signal to noise ratio.

2. Deconvolution - Reverberation and echo canceling. This kind of interference canceling is often referred to as echo canceling because it enables the removal of reverberations and echo from a single observed signal. A delayed version of the primary input signal is fed to the linear adaptive filter thus enabling the filter to reconstruct and remove reverberation from the undelayed primary signal. The deconvolver may also be used to cancel periodic interference components in the primary input such as power line interference etc. The adaptive filter is able to extrapolate the periodic interference and subtract this component from the undelayed primary input (see Figure 2). The adaptability normally provide superior performance compared to e.g. standard notch or comb filtering.

3. Line Enhancement. In this case the objective is to find a periodic or quasi periodic signal which is buried in noise. The adaptive filter is receiving the same input as the deconvolver, however, instead of subtracting the extrapolated periodic signal from the input it is output directly.

4. Adaptive bandpass filtering. Often we may take advantage of some a priori knowledge regarding the bandwidth of the signal we wish to cleanse. By bandpass filtering the signal we eliminate parts of the frequency range where the useful signal is weak and the noise is comparatively strong thus enhancing the overall signal to noise ratio.

In traditional linear Finite Impulse Response (FIR) adaptive noise cancellation filter, the noise is estimated as a weighted sum of delayed samples of reference interference. However, linear adaptive noise cancellation systems mentioned above may not achieve acceptable level of cancellation of noise for many real world problems when interference signals are related the measured reference signals in a complex dynamic and nonlinear way. Optimum interference and noise cancellation usually requires nonlinear adaptive processing of recorded and measured on-line signals [7], [11].

A common technique for noise reduction is to split the signal in two or more bands. The high-pass bands are subjected to a threshold nonlinearity that suppresses low amplitude values while retaining high amplitude values [11].

2. STANDARD PCA AND ICA APPROACHES FOR PRELIMINARY NOISE REDUCTION

PCA is a standard technique for the computation of eigenvectors and eigenvalues of an estimated autocorrelation matrix $\widehat{\mathbf{R}}_{xx} = \left\langle \mathbf{x}\mathbf{x}^T \right\rangle = \frac{1}{M}\sum_{t=1}^{M} \mathbf{x}(t)\mathbf{x}^\top(t) = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\top \in R^{m\times m}$.

PCA enables to decompose mixed signals into two subspaces: the signal subspace corresponding to principal components associated with the largest eigenvalues $\lambda_1, \lambda_2, ...,$ λ_n, $(m > n)$ and the noise subspace corresponding to the minor components associated with the smallest eigenvalues $\lambda_{n+1}, ..., \lambda_m$, where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m \geq 0$ is assumed. The standard numerical procedure is as follows. First, estimate the $m \times m$ autocorrelation matrix $\mathbf{R}_{xx}$ of the zero-mean mixed signal vector $\mathbf{x}(t)$. Then compute the m eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_m \geq 0$ and corresponding eigenvectors $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m$. These eigenvalues and eigenvectors are obtained by the standard PCA. The subspace spanned by the n eigenvectors $\mathbf{v}_i$ corresponding to n largest eigenvalues be considered as an approximation of the noiseless signal subspace. Having constructed the signal and noise subspaces, we can project the original data onto the signal subspace by:

$$\hat{\mathbf{x}}(t) = \mathbf{Q}\mathbf{x}(t) = \mathbf{\Lambda}^{-1/2}\mathbf{V}^T\mathbf{x}(t), \qquad (2)$$

where $\mathbf{\Lambda} = \text{diag}\,\{\lambda_1, \lambda_2, ..., \lambda_n\}$ is the diagonal matrix containing the n largest eigenvalues and $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n]$ is the matrix of the associated eigenvectors. It is not difficult to show that the filtering problem is related to PCA method. Indeed, the filtering problem can be formulated as finding an estimate $\hat{\mathbf{x}}(t) = \mathbf{Q}\mathbf{x}(t)$ such that the mean square error $E\left\{\|\tilde{\mathbf{x}}(t) - \hat{\mathbf{x}}(t)\|^2\right\}$ is minimized, where $\tilde{\mathbf{x}}(t) = \mathbf{x}(t) - \mathbf{v}(t)$. Obviously that the filter based on minimization of the mean square error is nothing but inverse transformation of $\mathbf{x}(t)$ to the PCA basis of $\tilde{\mathbf{x}}(t)$ for the assumption that the additive noise is uncorrelated with vector $\tilde{\mathbf{x}}(t)$ and is white. One important advantage of this approach is that it enables not only reduction in the noise level, but also allows us to estimate the number of sources [15]. This approach have been applied to noise reduction in electroencephalographic signals [5], [17]. A problem arising from this approach, however, is how to correctly set or estimate the threshold which divides eigenvalues into the two subsets, especially when the noise is large (i.e., the SNR is low) [5], [15]. The uncorrelated principal components are ordered by decreasing values of their values.

Recently, it has been realized that ICA or at least combining of both techniques: PCA and ICA, is more appropriate for noise reduction and moreover such approach reveal underlying structure of signals better than PCA alone [11], [14], [19]. Moreover, using ICA we can achieve better results in the sense that PCA use only second-order statistics, but ICA can estimate a better basis by taking into account higher-order statistics inherent in the data and allow to build nonlinear estimator instead of linear one. ICA algorithms can be also robust, what is very important for noise cancellation applications. ICA allows to separate sources $\mathbf{s}(t)$ based on observations $\mathbf{x}(t)$ using maximum *a posteriori* method that is dispose of *a priory* information problem and allows to realize blind scenario. Using ICA we can find independent components, which are undesirable and can be thought as noisy sources and eliminated.

In this paper we propose to apply first ICA and next ordering the independent components (ICs) according to decreasing absolute value of their normalized kurtosis rather than their variances; since the normalized kurtosis $\kappa_4(y_i) = \frac{E\{y_i^4\}}{E^2\{y_i^2\}} - 3$ is natural measure of Gaussianity of signals. Using $\kappa_4(y_i)$ we can easily detect and remove white (colored) Gaussian noises form raw sensory data. Optionally we can use more robust measures to detect and classify specific ICs [10]. This is also the subject of next 2 sections.

2.1. Detection of i.i.d. and temporally structured components using linear predictor (e.g. enhancement of EPs)

In many applications only temporally structured sources are subject of interest where all i.i.d. (independent identically distributed) components should be removed. Lets us assume that primary source signals are modeled by stable autoregressive process as

$$\begin{aligned} s_i(t) &= \widetilde{s}_i(t) - \sum_{p=1}^{L} a_{ip} s_i(t-p) \\ &= \widetilde{s}_i(t) - A_i(z) s_i(t), \end{aligned} \qquad (3)$$

where $A_i(z) = \sum_{p=1}^{L} a_{ip} z^{-p}$ and $\widetilde{s}_i(t)$ are i.i.d. unknown innovation processes. In order to estimate primarily innovative source signal $e_i(t) \approx c_i \widetilde{s}_i(t - d_i)$ (here d_i is some possible delay and c_i is some possible scaling coefficient) we consider a linear predictor [1], [9]

$$e_i(t) = y_i(t) - B_i(z) y_i(t) = y_i(t) - \mathbf{b}_i^\top \mathbf{y}_i(t), \qquad (4)$$

where $B_i(z) = \sum_{p=1}^{M} b_{ip} z^{-p}$ with $M \geq L$, $\mathbf{b}_i = [b_{i1} \cdots b_{iM}]^T$ and $\mathbf{y}_i(t) = [y_i(t-1), y_i(t-2), \ldots, y_i(t-M)]^T$.

Applying the standard gradient descent technique for minimization of cost function $J(\mathbf{b}_i) = \frac{1}{2}E\{e_i^2(k)\}$, we obtain simple LMS on-line learning rule:

$$\mathbf{b}_i(t+1) = \mathbf{b}_i(t) + \eta_i e_i(t) \mathbf{y}_i(t), \qquad (5)$$

where $\eta_i > 0$ is the learning rate. Instead of on-line LMS algorithm we can use well known Wiener filter batch estimation as [9]

$$\mathbf{b}_i = \mathbf{R}_{y_i y_i}^{-1} \mathbf{p}_i, \qquad (6)$$

where $\mathbf{R}_{y_i y_i} = E\{\mathbf{y}_i \mathbf{y}_i^T\}$ and $\mathbf{p}_i = E\{\mathbf{y}_i y_i\}$. It should be noted that for a white (i.i.d.) signal y_i the cross-correlation vector $\mathbf{p}_i$ equal zero, so vector $\mathbf{b}_i$ also will be zero. This fact enable us very easily to detect and eliminate the white sources. In more general case, the vector $\mathbf{b}_i$ represents the temporal structure of the corresponding signal y_i, so some classification of temporal correlated sources is possible on basis of vector $\mathbf{b}_i$.

Temporal structure of sources can be described by more general means, e.g. using ARMA (autoregressive moving-average) process or HMM (Hidden Markov Model) which is able to represent high-order temporal statistics and facilitates EM learning rules [1], [3].

2.2. Detection and classification of independent components on basis of Hurst exponent

Studying of living organisms as complex nonlinear dynamic systems generating time series is of increasing interest to biology and neuroscience [13], [16], [20], [22]. The Hurst exponent H and associated fractal dimension $D = 2 - H$ is one possible parameter that characterize time series [13], [20]. Hurst in 1965 [13] develop the rescaled range (R/S) analysis for time series $y(t)$, $(t = 0, 1, 2, ...)$. Firstly, the range R defined as a difference between maximum and minimum "accumulated" values:

$$R(T) = \max_{1 \le t \le T} \{Y(t,T)\} - \min_{1 \le t \le T} \{Y(t,T)\}, \tag{7}$$

where

$$Y(t,T) = \sum_{t=1}^{T} \{y(t) - < y(t) >\},$$

and secondly, standard deviation S estimated from the observed value $y(t)$

$$S = \left(\frac{1}{T} \sum_{t=1}^{T} [y(t) - < y(t) >]^2 \right)^{\frac{1}{2}}. \tag{8}$$

Hurst found that the ration R/S is very well described for large number of phenomena by the following nonlinear empirical relation:

$$\frac{R}{S} = (cT)^H, \tag{9}$$

where T is the number of samples, c is some constant (typically $c = \frac{1}{2}$) and H is the Hurst exponent in the range from 0 to 1.

With this definition a Hurst exponent of value 0.5 correspond to a time series that is a truly random (e.g. Brown noise or Brownian motion). A Hurst exponent of $0 < H < 0.5$ shows so called antipersistent behavior, e.g. white uniform distributed noise has $H \cong 0.15$. At limit of $H = 0$ the time series must change direction every sample. On the other hand, a Hurst exponent of $0.5 < H < 1$ describes a temporally persistent or trend reinforcing time series. At limit a straight line with non zero slope will have a Hurst exponent of 1.

It was found by many researchers that a Hurst exponent H has value $0.70 - 0.76$ for many natural, economic and human phenomena.

In this paper we propose to apply the Hurst exponent H and/or fractal dimension D to classify and detect of independent components $\hat{y}_i(t)$ of EEG/MEG signals. Usually independent components $\hat{y}_i(t)$ can be considered as random or temporally independent processes if $H \le 0.6$. These components can be easily eliminated by open of switches in corresponding channels (see Figure 1). From the other hand, the most interesting or desirable components will have a Hurst exponent in the range $H = 0.70 - 0.76$. These components can be projected by pseudo inverse matrix $\mathbf{W}^+$ so corrected sensor signals enable us to localize corresponding "interesting" brain sources. Furthermore we have found by extensive computer experiments that some artifacts like eye blinking or heart beat artifacts have characteristic value of H, so they could be automatically identified and removed from sensor signals on basis of value of a Hurst exponent.

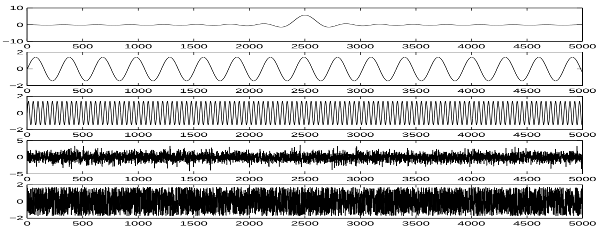

Figure 3: Set of sources for artificial data simulation.

For calculation of Hurst exponent we use recurrent method proposed in [20] in order to reduce the computation complexity. The recurrent method can be written in the following form

$$\begin{aligned}
H(t+1) &= \frac{\lg\left(\frac{R(t+1)}{S(t+1)}\right)}{\lg\left(\frac{t+1}{2}\right)}, \\
R(t+1) &= Y_{\max}(t+1) - Y_{\min}(t+1), \\
Y_{\max}(t+1) &= \begin{cases} Y(t+1), & Y(t+1) > Y_{\max}(t), \\ Y_{\max}(t+1), & Y(t+1) \le Y_{\max}(t), \end{cases} \\
Y_{\min}(t+1) &= \begin{cases} Y(t+1), & Y(t+1) < Y_{\min}(t), \\ Y_{\min}(t+1), & Y(t+1) \ge Y_{\min}(t), \end{cases} \\
Y(t+1) &= Y(t) + \frac{t}{t+1}(y(t+1) - < y(t) >), \\
< y(t) > &= \frac{t-1}{t} < y(t-1) > + \frac{1}{t} y(t), \\
S(t+1) &= D^{\frac{1}{2}}(t+1), \\
D(t+1) &= \frac{t}{t+1} D(t) + \frac{t^2}{(t+1)^3} \\
&\times \quad (y(t+1) - < y(t) >)^2, \\
Y_{\max}(0) &= Y_{\min}(0) = y(0), \\
H(0) &= D(0) = Y(0) = 0.
\end{aligned} \tag{10}$$

3. SIMULATION RESULTS

In this section we present the exemplary results of a computer simulations. We performed simulations both for artificially generated noisy signals as well as real-world single trial EEG/MEG data with 21/149 channels.

3.1. Artificial data simulation

In order to get an idea about the effectiveness and validity of proposed techniques for noise and interference cancellation for multi-sensory observations we performed experiments with set of known signals.

The set of useful signals was generated as follows: $s_1(t) = \text{sinc}\left(\frac{t}{15}\right)$, $s_2(t) = \sin\left(\frac{\omega_0 \pi t}{p}\right)$, $s_3(t) = \sin\left(\frac{7\omega_0 \pi t}{p}\right)$, where $\omega_0 = 50$ and $p = 755$ in our simulation example. It is easy to see that sources are independent expect of $s_2(t)$ and $s_3(t)$. The set of noise sources consists of two sources, where one of them was generated according to normal distribution low $N(0,1)$ and another was uniformly distributed in range $[-1, 1]$. The full set of sources is shown on the Figure 3. All signals were mixed using randomly chosen mixing matrix.

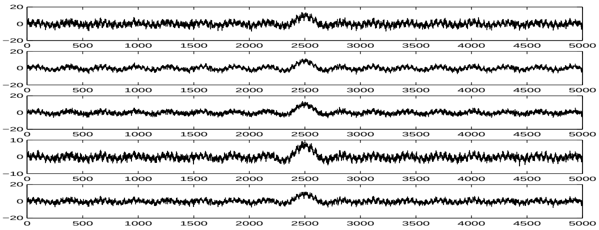

Figure 4: Sensor signals for artificial data simulation.

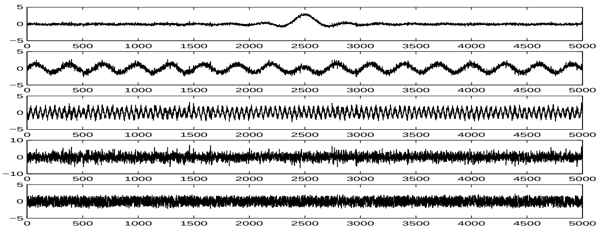

Figure 5: Separated signals (independent components) for artificial data simulation.

The Gaussian noise $N(0,1)$ with SNR 20dB was added to the outputs $\mathbf{x}(t)$ of mixture system. We assume in this simulation that only these noisy outputs of mixture system can be observed. The noisy outputs are shown in the Figure 4.

The purpose is to cancel all type of noise and interferences from observed signals. The scenario of numerical simulation is follows. At the first stage we separate signals from the noisy mixture using robust ICA algorithm. For our simulation we use High-Order Statistics (HOS) based algorithm proposed in [8]. The separated signals are shown in the Figure 5.

At the second stage Hurst exponents H_i and norms of FIR filter parameter vector $\|\mathbf{b}_i(t)\|$ are calculated and compared with some thresholds for observed and separated signals. For calculation of Hurst exponent we use recurrent method (10) and for FIR filter parameter vector estimation we use method proposed in [21].

Based on comparison of Hurst exponents and norms of linear predictor parameter vector with correspondent thresholds the decision to switch off or switch on the corresponding independent component is taken automatically.

The results for calculation of Hurst exponents and norms of linear predictor parameter vector are shown in the Table 1. It is obvious that we can say nothing about which signals are noise and which signals are useful for sensor observations but it is relatively easy to find noisy signals for separated sources. The thresholds for Hurst exponent and norm of linear predictor parameter vector are 0.6 and 0.05.

The separated signals 4 and 5 can be removed. Then reconstructed sensor signals are as shown on the Figure 6. Inversion and projection of separated signals that were passed through on-line switching system (see Figure 1) is the last third stage of our procedure.

Table 1: Hurst exponent and norm of vector $\mathbf{b}_i$ of linear predictor for each signal shown in Figure 4 and Figure 5.

Signal number i	H_i		$\max\|\mathbf{b}_i(t)\|$	
	$x_i(t)$	$y_i(t)$	$x_i(t)$	$y_i(t)$
1	0.7297	0.7997	0.8797	2.4555
2	0.7353	0.7220	1.1638	1.7780
3	0.7523	0.6255	1.3596	0.3541
4	0.7559	0.5182	1.2208	0.0221
5	0.7429	0.4949	1.0621	0.0151

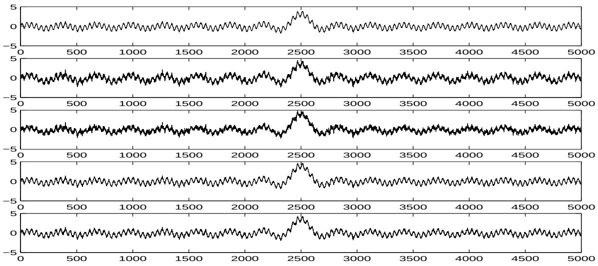

Figure 6: Reconstructed sensor signals after removing ICs 4 and 5 for artificial data simulation.

3.2. Real data analysis application

Similar experiments were performed for real-world data - EEG and MEG signals with 21 and 149 channels respectively. The exemplary observed selected EEG signals are shown in Figure 7. Here we plotted only 5 from 21 sensor signals. After application of ICA procedure for separation of the sensor 21 EEG signals we have obtained the results shown in the Figure 8. Analysis of the values H_i and $\|\mathbf{b}_i\|$ for observed sensor signals and separated signals (independent components) have completely different distributions and identification or detection of random signals is only possible after applying ICA. For this case ICs number 3 should be removed.

4. CONCLUSIONS

In this paper we have proposed several methods for signal detection (identifying the presence in independent components random signals or deterministic signals with specific features or temporal structure) and classification (assignment of a independent component to a particular class) par-

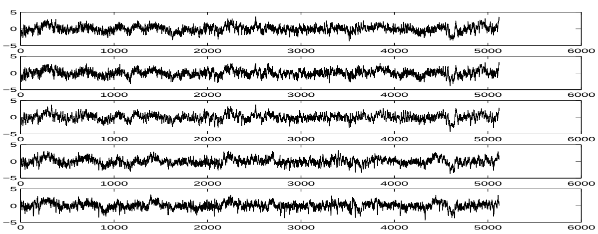

Figure 7: Exemplary observed noisy EEG data.

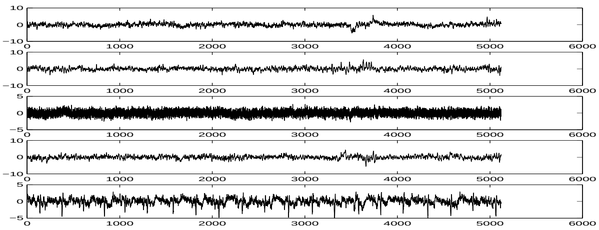

Figure 8: Plots of the first 5 ICs for EEG data.

Table 2: Hurst exponent and norm of vector $\mathbf{b}_i$ of linear predictor for each signal shown in Figure 7 and Figure 8.

Signal number i	H_i		$\max \|\mathbf{b}_i(t)\|$	
	$x_i(t)$	$y_i(t)$	$x_i(t)$	$y_i(t)$
1	0.7290	0.7513	0.3657	1.0431
2	0.7117	0.6989	0.3348	0.4245
3	0.6970	0.5418	0.4647	0.0173
4	0.7120	0.7361	0.5257	0.9348
5	0.7232	0.6452	0.5333	0.2197

ticularly in the high noise regime. Especially, application of Hurst exponent together with ICA has been investigated to our knowledge for the first time for automatic elimination of noise and undesirable interference. Of course, these methods exploit only small part of possible techniques which can be used together with ICA for noise and interference cancellation. Especially, higher order and nonlinear correlation methods are very promising. However, the proposed and investigated methods appear to be simple and efficient for specific applications, especially for enhancement of single trial EEG/MEG data, what has been confirmed by extensive computer simulation experiments.

5. REFERENCES

[1] Amari, S. (1999) ICA of temporally correlated signals - learning algorithm, *Proc. ICA'99*, Aussois, France, pp. 13-18.

[2] Amari, S.I. and Cardoso, J.F. (1997) Blind source separation – semiparametric statistical approach, *IEEE Transactions on Signal Processing* **45**(11):2692–2697.

[3] Attias, H. (2000) Independent factor analysis with temporally structured factors. Advances in Neural Information Processing Systems 12 (Ed. by T. Leen et al.), MIT Press, Cambridge, MA. *(in print)*.

[4] Bell, A.J. and Sejnowski, T.J. (1996) An information maximization approach to blind separation and blind deconvolution, *Neural Computation* **7**:1129–1159.

[5] Callaerts, D. et al. (1998) On-line algorithm for signal separation based on SVD, in E.F. Deprettere, ed., *SVD and Signal Processing: Algorithms, Applications and Architectures*, Elsevier, Amsterdam, pp. 269-276.

[6] Cichocki, A., Thawonmas, R. and Amari, S. (1997) Sequential blind signal extraction in order specified by stochastic properties, *Electronics Letters*, **33**(1):64–65.

[7] Cichocki, A., Vorobyov, S.A. and Rutkowski, T. (1999) Nonlinear interference cancellation using neural networks, *Proceed. of Int. Symposium on Nonlinear Theory and its Applications*, vol.2, pp.875-878.

[8] Cruces, S., Castedo, L. and Cichocki, A. (2000) Novel blind source separation algorithms using cumulants, *Proceedings of the ICASSP*, Istanbul, Turkey.

[9] Haykin, Y. (1996) *"Adaptive filter theory"*, Prentice-Hall, Inc., Upper Saddle River, New Jersey, Third Edition.

[10] Hyvärinen. A. (1999) Gaussian moments for noisy independent component analysis, *IEEE Signal Processing Letters* **6**(6):145–147.

[11] Hyvärinen, A., Hoyer, P. and Oja, E. (1999) Sparse code shrinkage: denoising by nonlinear maximum likelihood estimation, *Advances in Neural Information Processing Systems 11 (NIPS*98)*.

[12] Hyvärinen, A. and Oja, E. (1997) A fast fixed-point algorithm for independent component analysis, *Neural Computation*, **9**:1483–1492.

[13] Hurst, H.E., Black, R.P. and Simaika, Y.M. (1965) *"Long term storage"*, Constable, London, U.K.

[14] Jung, T.P., Humphries, C., Lee, T-W., McKeown, M., Iragui, V., Makeig, S., Sejnowski, T.J. (1998) Removing electroencephalographic artifacts: Comparison between ICA and PCA, *IEEE International Workshop on Neural Networks for Signal Processing*, pp. 63-72.

[15] Karhunen, J., Cichocki, A., Kasprzak, W., and Pajunen, P. (1997) On neural blind separation with noise suppression and redundancy reduction, *Int. Journal of Neural Systems* **8**(2):219-237.

[16] Katz, M.J. (1988) Fractals and the analysis of waveforms, *Comput. Biol. Med.* **18**:145-155.

[17] Sadasivan, P.K. and Dutt, D.N. (1996) SVD based technique for noise reduction in electroencephalographic signals, *Signal Processing* **55**:179-189.

[18] Vetterli, M. and Kovacević, J. (1995) *"Wavelets and subband coding"*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

[19] Vigario, R., Jousmaki, V., Hamalainen, M., Hari, R., and Oja, E. (1997) Independent component analysis for identification of artifacts in magnetoencephalographic recordings, *Advances in Neural Information Processing System 10 (Proc. NIPS*97)*, MIT Press, pp. 229 - 235.

[20] Vorobyov, S.A. and Shilo, A.V. (1998) The recurrent normalized range method for revelation of fractal structure and its application to analysis of EEG, *Radioelectronics and Informatics* (3):162-165. (*in Russian*)

[21] Vorobyov, S.A. and Bodyanskiy, Ye.V. (1999) On a non-parametric algorithm for smoothing parameter control in adaptive filtering, *Engineering Simulation* **16**:341-350.

[22] Turner, K.M. (1993) Reliability of storage scheme, tree rings, and Hurst phenomena", *Proc. of the American Geophysical Union Twelfth Annual Hydrology Days*, pp.230-242.