

Blind noise reduction for multisensory signals using ICA and subspace filtering, with application to EEG analysis

Sergiy Vorobyov*, Andrzej Cichocki**

Laboratory for Advanced Brain Signal Processing, Brain Science Institute, RIKEN, 2-1 Hirosawa, Wako-shi, Saitama, 351-0198, Japan

Received: 6 November 2000 / Accepted in revised form: 12 November 2001

Abstract. In many applications of signal processing, especially in communications and biomedicine, preprocessing is necessary to remove noise from data recorded by multiple sensors. Typically, each sensor or electrode measures the noisy mixture of original source signals. In this paper a noise reduction technique using independent component analysis (ICA) and subspace filtering is presented. In this approach we apply subspace filtering not to the observed raw data but to a demixed version of these data obtained by ICA. Finite impulse response filters are employed whose vectors are parameters estimated based on signal subspace extraction. ICA allows us to filter independent components. After the noise is removed we reconstruct the enhanced independent components to obtain clean original signals; i.e., we project the data to sensor level. Simulations as well as real application results for EEG-signal noise elimination are included to show the validity and effectiveness of the proposed approach.

1 Introduction and problem formulation

In many real-world applications of signal processing, especially in communications and biomedicine, the problem of noise cancellation is important (Widrow and Walach 1996). Noise cancellation is a subject of wide interest in physical and communication systems. Several methods have been suggested in the literature for noise reduction. Signal processing techniques using for noise elimination include band-pass filtering, the fast Fourier transform, autocorrelation, autoregressive modeling, adaptive filtering, Kalman filtering, and singular value decomposition (SVD) (Akay 1996; Arnold et al. 1998; Sadasivan and Dutt 1996; Thakon 1987; Walter

1969; de Weerd and Martens 1978; Widrow and Walach 1996). Recently, the principal component analysis (PCA) (Callaerts et al. 1988; Laguna et al. 1999; Sadasivan and Dutt 1996) and independent component analysis (ICA) (Cichocki and Vorobyov 2000; Lee 1998) approaches have become very popular for the analysis of biomedical data e.g., EEG and MEG). One of the main advantages of these approaches relates to their applicability to multisensory observations of mixed signals.

In this paper we consider the following linear mixture model for measured signals

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

where $t = 0, 1, 2, \dots$ is discrete time; $\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_n(t)]^T$ is an n -dimensional vector of observed noisy sensor signals; \mathbf{A} is an $n \times m$ unknown full-rank mixing matrix; $\mathbf{s}(t) = [s_1(t), s_2(t), \dots, s_m(t)]^T$ is an m -dimensional unknown vector of primary sources; and $\mathbf{v}(t)$ is n -dimensional also unknown vector of additive white (generally, could be colored) Gaussian noise represented measurement and environmental noise. Furthermore, we assume that the vector $\mathbf{s}(t)$ contains a subset of useful or “interesting” sources with temporal structure, and “uninteresting” interferences or “inner” noises.

Our objective is to reduce the influence of additive noise $\mathbf{v}(t)$ and eliminate “inner” noise. In other words, our task is to obtain corrected or “cleaned” sensor signals which contain only useful or “interesting” sources with temporal structure. By useful or “interesting” signals we mean short-duration (sparse) signals with temporal structure such as the evoked potential/event-related potential (EP/ERP) in biomedical signal analysis applications (Goldstein and Alrich 1999; Niedermeyer and de Silva 1999). If these signals are statistically independent, our second objective is to estimate the corresponding sources.

We apply model (1) to EEG signal analysis, by describing all variables in model (1) as follows. Sources are original sparse signals generated by the brain. Some of sources can be noise sources. A noisy instantaneous mixture of original sources is available for measurement.

Correspondence to: S. Vorobyov
(Fax: +1-905-5212922, e-mail: svor@mail.ece.mcmaster.ca)

*Department of Electrical and Computer Engineering, McMaster University, 1280 Main St.W., Hamilton, Ontario L8S 4K1, Canada

**Warsaw University of Technology Poland

Let us emphasize that the problem consists of reduction of both the “inner” noise and an additive noise based on only information about observed vector $\mathbf{x}(t)$ (i.e., a blind scenario).

Such a formulation of the problem lies outside traditional noise-cancellation approaches (Haykin 1996; Widrow and Walach 1996), because the model (1) contains an “inner” noise as well as additive noise. However, traditional filtering techniques are based on the assumption that the noise is only additive. PCA and SVD are standard methods used to find and remove a noise subspace of a signal (Callaerts et al. 1998; Laguna et al. 1999; Sadasivan and Dutt 1996). However, both PCA and SVD perform well only if: (a) the noise level is small enough and (b) a signal subspace and a noise subspace are orthogonal to each other. For practical applications the orthogonality requirement for signal and noise subspaces is usually not valid. Moreover, “inner” noise sources in the model influence strongly the real signal-to-noise ratio (SNR). Moreover, PCA and SVD approaches are not able to distinguish “inner” and additive noises. Thus, good estimation of signal subspace is not possible.

Recently, it was found that one of the most promising approaches to the problem of blind noise reduction is application of ICA (Cichocki and Vorobyov 2000). Recent ICA algorithms are robust with respect to additive noise, which makes it possible to use them successfully for model (1). However, ICA cannot guarantee that some individual independent components (ICs) contain only noise and do not contain information about useful sources, especially in biomedical applications. Hence, the problem of detection and filtering of “useful” part of each IC is still open, and additional tools are needed to solve it. We will see that for biological signal processing applications these problems can be solved relatively easy, because usually we are interested in extraction of sparse signals, such as the EP’s. Hence, we have some a priori knowledge about signals which we are looking for at the stage of IC filtering.

In this paper, we propose a noise reduction technique for multisensory signals described by model (1). This formulation is closed to the description of real biomedical data measurements such as EEG/MEG signals (Dogandzic and Nehorai 2000). Section 2 concentrates on the description of our approach to blind noise reduction using ICA and subspace filtering. In Sect. 3 we give some details about an algorithm realizing the proposed approach, such as choosing useful ICs and subspace filtering. Section 4 presents some simulation results for artificially generated data to show the validity and performance of the proposed method. Results for real EEGs are also given. Section 5 provides the conclusions.

2 Schema of blind noise reduction based on ICA and subspace filtering

The block diagram illustrating the ICA technique application for noise and interference reduction is shown in Fig. 1 (see also Cichocki and Vorobyov 2000). Sensor

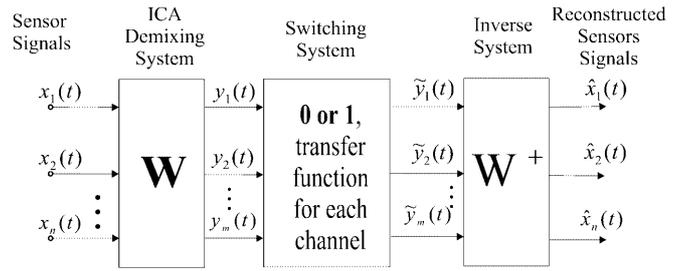


Fig. 1. Block diagram illustrating the independent component analysis (ICA) method for noise reduction. See text for details

signals are processed by a “ICA demixing system” that is described by the model

$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t) , \quad (2)$$

where $\mathbf{y}(t) = [y_1(t), y_2(t), \dots, y_m(t)]^T$ is an m -dimensional vector of independent components with $m \leq n$, and \mathbf{W} is a separation matrix.

The outputs of the “ICA demixing system” pass through the “switching system” which takes the binary decision $\{0, 1\}$. Thus, the signals that corresponds to the ICs obtained after separation can be passed through when the corresponding switch is on, or perfectly eliminated when the switch is off. The results go through the “inverse system” represented by the inverse of an estimated separation matrix. At the output of the system we have reconstructed sensor signals $\hat{\mathbf{x}}(t) = [\hat{x}_1(t), \hat{x}_2(t), \dots, \hat{x}_n(t)]^T$.

It should be noted that such a simple approach is valid under the assumptions that: (a) interferences and noises are extracted as ICs and can be easily detected and recognized, and (b) there is no additive noise in the model (1). The analysis of biomedical signals shows that such a simple schema is usually not valid (Dogandzic and Nehorai 2000). There are several reasons for this, and the most important ones are the following. First, besides “inner” noise that is biological in nature (e.g., from cardiac, smooth, and skeletal muscles, the EEG recorded using surface electrodes is always buried in additive noise that is electrical in nature which emanates from a variety of sources such as instrumentation, recording electrodes, and surrounding power lines. Second, a spatially correlated noise between sensors with unknown covariance should be taken into account.

Thus, separation on \mathbf{W} can be biased by noise. A dependance between noise components also leads to biased estimation of separation matrix. Note that even if noise components are independent, after separation we will not obtain clean signals but estimated signals of the sources plus the multiplication of the estimated separation matrix by the vector of additive noise:

$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t) + \mathbf{W}\mathbf{v}(t) = \hat{\mathbf{s}}(t) + \mathbf{W}\mathbf{v}(t) . \quad (3)$$

If some separated sources are just “inner” noise components, we have at the output of the separation system simply the accumulation of two types of noises. Such ICs could be detected and completely rejected. However, for ICs which contain a “useful” signal, a

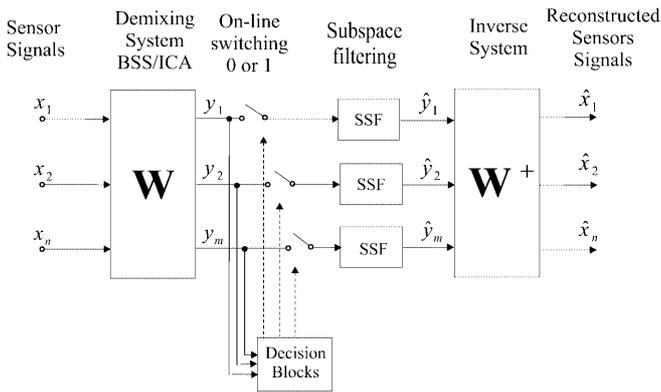


Fig. 2. Block diagram illustrating the ICA method combined with higher-order correlation-based filtration for noise reduction

filtering procedure must be performed to reduce the term $\mathbf{W}\mathbf{v}(t)$ in (3). Hence, the problem consists now of recognition of ICs that contain useful (“interesting”) signals buried in noise, and filtering them.

Note that this reduction of the initial problem, that is actually filtering of measured signals (1), has a strong motivation because “inner” noise signals have equivalent or often higher power than “interesting” signals. This leads to dramatic decreases in the SNR for the measured signals. For example, in biomedical applications the measured signals typically suffer from SNRs of -10 dB to -20 dB. Traditional filtering techniques do not perform well for signals with such a low SNR.

The schema illustrating the method for which the framework has been given above is shown in Fig. 2. Instead of “switching system” in the schema in Fig. 1, we incorporate a filter in each channel in Fig. 2 (see Cichocki and Amari 2002 for alternative models).

3 Algorithm for blind noise reduction based on ICA and subspace filtering

3.1 ICs extraction

First, we apply ICA to obtain the ICs. Any algorithm that ensures robust unbiased estimation of the separation matrix can be employed. Recently, several robust-to-additive-noise algorithms for unbiased estimation of separation matrix have been investigated (Choi and Cichocki 2000; Cruces et al. 2000; Cichocki and Amari 2002). The modification of the algorithm described by Cruces et al. (2000) (see Appendix A) is used in the simulations for this paper.

We will not dwell on the details of the algorithm, which can be found in the original paper. Here we are more interested to answer the question: is ICA really helpful and give some reasonable benefits to finding ICs describing noise or artifacts in comparison to direct analysis of the originally measured EEG signals? To answer this question we need to define the measure for estimation of “usefulness” of signals in application to EEG analysis. At the same time, the answer to this question gives us the solution to a switching stage (see Fig. 2).

3.2 Detection and classification of ICs

Obviously, there are several possibilities for classification of signals based on a “usefulness” criterion. For possible comparison of results we need to consider at least two different criteria. However, first we have to specify the sense of “usefulness” in EEG-analysis application.

This is not a trivial problem. However, we always can be certain that if a signal has no temporal structure or is independent identically distributed (i.i.d.) it can give us no information for analysis, except the information that the signal is unpredictable. Such a signal can be rejected from the analysis. Further, electrical noise and artifacts can still have a temporal structure. Hence, we need also to define a more general measure of randomness. Here we discuss the Hurst exponent as a measure of “usefulness”. In Sect. 3.2.3 we give a simulation example for comparison of original EEG signals and ICs from the viewpoint of detecting “useless” components.

3.2.1 Detection of i.i.d. and temporally structured components using a linear predictor. In many applications only temporally structured sources are of interest, where all i.i.d. components should be removed. Let us assume that the primary source signals are modeled by a stable autoregressive process as

$$s_i(t) = \tilde{s}_i(t) - \sum_{p=1}^N a_{ip}s_i(t-p) = \tilde{s}_i(t) - A_i(z)s_i(t), \quad (4)$$

where $A_i(z) = \sum_{p=1}^N a_{ip}z^{-p}$ and $\tilde{s}_i(t)$ are i.i.d. unknown innovation processes. In order to estimate the primarily innovative source signal $e_i(t) \approx c_i\tilde{s}_i(t-d_i)$ (here d_i is some possible delay and c_i is some possible scaling coefficient), we consider a linear predictor (Haykin 1996):

$$e_i(t) = y_i(t) - B_i(z)y_i(t) = y_i(t) - \mathbf{b}_i^T \mathbf{y}_i(t), \quad (5)$$

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

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

$$\mathbf{b}_i(t+1) = \mathbf{b}_i(t) + \eta_i e_i(t) \mathbf{y}_i(t), \quad (6)$$

where $\eta_i > 0$ is the learning rate. Instead of the on-line Least-Mean Square (LMS) algorithm, we can use the Wiener filter batch estimation (Haykin 1996)

$$\mathbf{b}_i = \mathbf{R}_{y_i y_i}^{-1} \mathbf{p}_i, \quad (7)$$

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

The temporal structure of sources can be described by more general means, for example using an autoregressive moving-average process or a hidden Markov model, which is able to represent high-order temporal statistics and facilitates expectation maximization learning rules (Amari 1999; Attias 2000; Cichocki and Amari 2002). However, we can avoid these details because the set of signals which we are looking for is not limited only to temporally uncorrelated signals. The investigation of a more general measure is more important than increasing the performance of the method described above.

3.2.2 Detection and classification of ICs on the basis of the Hurst exponent. Studying living organisms as complex nonlinear dynamic systems generating time series is of increasing interest to biology, and neuroscience in particular (Hurst et al. 1965; Katz 1988; Turner 1993; Vorobyov and Shilo 1998). The Hurst exponent H (and associated fractal dimension $D = 2 - H$) is one possible parameter for characterizing a time series (Hurst et al. 1965; Vorobyov and Shilo 1998). Hurst et al. (1965) developed the rescaled range (R/S) analysis for a time series $y(t)$, ($t = 0, 1, 2, \dots$). Firstly, the range R was defined as a difference between maximum and minimum “accumulated” values:

$$R(T) = \max_{1 \leq t \leq T} \{Y(t, T)\} - \min_{1 \leq t \leq T} \{Y(t, T)\} , \quad (8)$$

where

$$Y(t, T) = \sum_{i=1}^T [y(t) - \langle y(t) \rangle] ; \quad (9)$$

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

$$S = \left(\frac{1}{T} \sum_{i=1}^T [y(t) - \langle y(t) \rangle]^2 \right)^{\frac{1}{2}} . \quad (10)$$

Hurst et al. found that the ratio R/S is very well described for a large number of phenomena by the following nonlinear empirical relation:

$$\frac{R}{S} = (cT)^H , \quad (11)$$

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 the Hurst exponent of value 0.5 corresponds to a time series that is truly random (e.g., Brown noise or Brownian motion). The Hurst exponent of $0 < H < 0.5$ exhibits the so-called antipersistent behavior; e.g., white uniformly distributed noise has $H \cong 0.15$. At the limit of $H = 0$, the time series must change direction every sample. On the other hand, the Hurst exponent of $0.5 < H < 1$ describes a temporally persistent or trend-reinforcing time series. At the limit a straight line with nonzero slope will have the Hurst exponent of 1. It was found by many

researchers that the Hurst exponent H has a value equal to 0.70–0.76 for many natural, economic, and human phenomena.

In this paper we propose to apply the Hurst exponent H to classify and detect the ICs $\hat{y}_i(t)$, $i = 1, \dots, m$ of EEG/MEG signals. IC $\hat{y}_i(t)$ can be considered as a random or temporally independent process if $H \leq 0.6$. Such an IC can be easily eliminated by closing the closing switch in the corresponding channel (see Fig. 2). The most interesting or desirable components have a Hurst exponent in the range $H = 0.70 - 0.76$. These components can be projected by a pseudo inverse matrix \mathbf{W}^+ . Thus, corrected sensor signals enable us to localize corresponding “interesting” brain sources. Furthermore, we have found by extensive computer experiments that some artifacts, such as those from eye blinking or heart beats, have a specific value of H . Thus, they could be automatically identified and removed from sensor signals on the basis of the value of the Hurst exponent. For example, heart-beat artifacts are usually characterized by $H = 0.64 - 0.69$, and eye blinking by $H = 0.58 - 0.64$.

For calculating the Hurst exponent we use the recurrent method proposed in Vorobyov and Shilo (1998) in order to reduce the computation complexity. A summary of the procedure for calculating the Hurst exponent is given in Appendix B.

3.2.3 Comparison of original EEG signals and ICs from the viewpoint of detecting “useless” components. In this section we summarize some experiment results for real-world data (i.e., EEG signals) that demonstrate possible advantages of the application of ICA to the processing of biomedical signals. The purpose is to compare original EEG signals and ICs after ICA application from the viewpoint of detecting “useless” components. Clearly, the performance of this analysis generally depends on the performance of the ICA procedure. A modified version of the procedure developed in Cruces et al. (2000) (see Appendix A) is used for the extraction of ICs. The observed EEG signals are shown in Fig. 3. After ICA separates the sensor EEG signals, we obtain the results shown in the Fig. 4. Analysis of the values H_i and $\|\mathbf{b}_i\|$ for the observed sensor signals and separated signals (ICs) reveal completely different distributions, which can be seen from the Table 1. It follows that the identification or detection of random signals is only possible for ICs obtained after applying ICA. It is obvious that the third IC should be removed, and the fifth IC probably corresponds to heart-beat artifact and should also be removed. Hence, *ICA is helpful for EEG denoising and analysis.*

Assume now that we could decompose the signals from the electrodes, such that we know ICs corresponding to original sources, and reject ICs corresponding to noise sources and artifacts. Then for each unrejected IC we need to search for event-related signal corresponding to the original source; i.e. we need to reduce the influence of additive noise ($\mathbf{W}\mathbf{v}(t)$ in Eq. 3). We discuss a useful filtering technique to solve this problem in Sect. 3.3.

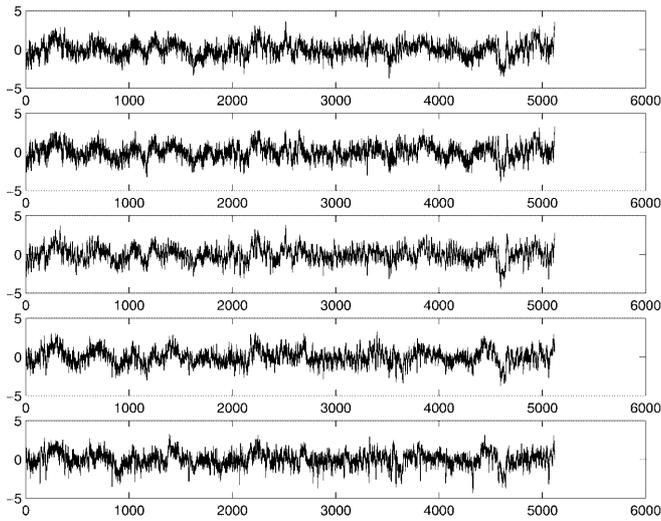


Fig. 3. Observed noisy EEG data

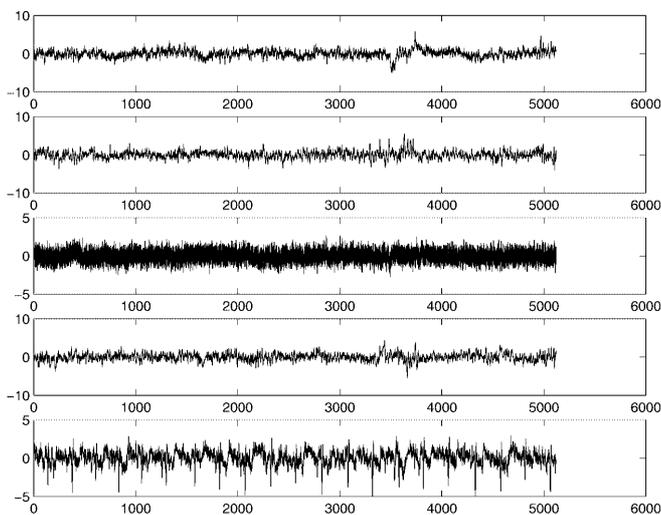


Fig. 4. Plots of the ICs for the EEG data

3.3 Subspace-based filtering

For processing the ICs for noise elimination a least-squares filter can be applied. However, event-related signals have a special structure that allow us to look for a low-rank filter instead of “full-rank” methods (Akay 1996; van der Veen et al. 1993; Haykin 1996; Pisarenko 1973; Strobach 1996). We introduce here a filter based on an approximate dominant eigendecomposition of the data covariance matrix.

The signals generated by the brain, such as EPs and ERPs, have a very specific sparse structure. The ICs, which are indeed estimations of brain-generated signals, are defined by (3), where $\mathbf{W}\mathbf{v}(t)$ is the noise term. To obtain an accurate estimation of a sparse signal—such as a EP or a ERP—contaminated by noise, which is no more unstructured because of multiplication by matrix \mathbf{W} , we first need to estimate the signal subspace and then filter the signal with respect to the filter parameters estimated only based on that subspace. Otherwise, the

Table 1. The Hurst exponent and norm of vector \mathbf{b}_i of linear predictor for each signal shown in Figs. 3 and 4

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

parameters estimated based on a whole signal and noise subspace describe a noise structure as well as a signal structure, and cannot provide an accurate estimation of parameters of the signal only. However, sparse signals are characterized only by few principal components. Thus, the signal subspace for the ICs (3) is easily separated from noise subspace by eigenvalue decomposition of the covariance matrices of the ICs.

Let us define the i th unrejected IC as $\tilde{y}_i(t)$ and introduce an autoregressive (AR) model

$$\tilde{y}_i(t) = \mathbf{c}_i^T \boldsymbol{\varphi}_i(t) + e_i(t) , \quad (12)$$

where $\mathbf{c}_i = (c_{i1}, c_{i2}, \dots, c_{iL})^T$ is an L -dimensional AR coefficient vector; $\boldsymbol{\varphi}_i(t) = [\tilde{y}_i(t-1), \tilde{y}_i(t-2), \dots, \tilde{y}_i(t-L)]^T$ is an L -dimensional vector of prehistory; and $e_i(t)$ is white noise with zero mean and $\sigma_{e_i}^2 < \infty$. Here L is the window length over which the covariance matrix is computed. The choice of the order of L for each IC depends on the shape of a corresponding event-related signal which we are looking for. However, such information often is not available a priori. Practically, L must be large enough, typically, $L \geq 20$.

The AR coefficient vector \mathbf{c}_i for i th IC can then be estimated using Wiener filter

$$\hat{\mathbf{c}}_i = -\hat{\mathbf{R}}_i^{-1} \hat{\mathbf{r}}_i . \quad (13)$$

However, to obtain an accurate estimation of the filter coefficients $\hat{\mathbf{c}}_i$ we need to find the signal subspace and estimate $\hat{\mathbf{c}}_i$ only on this subspace. The estimate of the filtered IC then can be calculated as

$$\hat{y}_i(t) = (\hat{\mathbf{c}}_i^{\text{pc}})^T \boldsymbol{\varphi}_i(t) , \quad (14)$$

where $\hat{\mathbf{c}}_i^{\text{pc}}$ is a Wiener filter (Haykin 1996) parameter vector estimated based only on the signal subspace. The details of the $\hat{\mathbf{c}}_i^{\text{pc}}$ calculation are given in Appendix C.

3.4 Summary

Based on the previous sections, we can introduce a blind noise-reduction algorithm for multisensory signals using ICA and subspace filtering. The algorithm can be defined by the following implementation:

1. Estimate the ICs of the set of signals using an ICA algorithm that is robust to additive noise (see Appendix A).

2. Test the hypothesis:

$$H_0 : \tilde{y}_i(t) = \tilde{s}_i(t) + n_i(t) , \quad (15)$$

$$H_1 : \tilde{y}_i(t) = \xi_i(t) + n_i(t) , \quad (16)$$

where $\tilde{s}_i(t)$ is the target signal of interest in the i th IC, $n_i(t) = \mathbf{w}_i^T \mathbf{v}(t)$, $\mathbf{w}_i(t)$ is the i th row vector of separation matrix \mathbf{W} , and $\xi_i(t)$ is the source corresponding to “inner” noise.

From the assumption that additive noise $v_i(t)$ in the model (1) is Gaussian, it can be inferred that the noise $n_i(t)$ is also Gaussian. If the “inner” noise is Gaussian, then the testing of hypothesis (15) is reduced to checking the Gaussianity of a tested signal. For this simplest case we can estimate the normalized kurtosis

$$K^4\{y_i(t)\} = \frac{E\{y_i^4\}}{E^2\{y_i^2\}} - 3 \quad (17)$$

and compare it with some small-enough threshold c , $|K^4\{y_i(t)\}| \leq c$ (Cichocki and Vorobyov 2000). If the “inner” noise is not Gaussian, the procedure based on the Hurst exponent calculation from Sect. 3.2.2 is applied (see Appendix B).

3. Reject the ICs which satisfy hypothesis H_1 , and filter ICs which satisfy hypothesis H_0 using the subspace-filtering procedure described in Sect. 3.3 and Appendix C.

4. Perform inverse projection of “interesting” filtered ICs $\hat{y}_i(t)$ back onto the sensors level by a linear transformation:

$$\hat{\mathbf{x}}(t) = \mathbf{W}^+(t)\hat{\mathbf{y}}(t) , \quad (18)$$

where \mathbf{W}^+ is a pseudo inverse of separation matrix \mathbf{W} ($\mathbf{W}^+ = \mathbf{W}^{-1}$ when the number of independent components is equal to the number of sensor).

Note that the above algorithm presents the performance of only the most basic schema. Further improvements may be achieved by “optimizing” the algorithm using, for example, low-rank adaptive tracking filtration (van der Veen et al. 1993; Nijima and Veno 2002; Pisarenco 1973; Raghothanan et al. 2000; Strobach 1996), or a robust prewhitening procedure for the case of correlated noise (Cichocki and Amari 2002).

4 Results

We divide this section into two subsections. Section 4.1 discusses the results obtained from the studies performed with known signals, and Sect. 4.2 discusses the results obtained from the studies on recorded EEG signals.

4.1 Simulation

In order to get an idea about the effectiveness of the proposed algorithm, we performed studies with known signals. The set of observed signals is generated using a mixing model (1), where the set of sources consists of

three useful signals and two noise signals. The set of useful signals contains $s_1(t)$ – sparse rectangular pulse; $s_2(t)$ – one period of a damped sine wave; and $s_3(t)$ – a long-duration rectangular pulse. One of the noise sources is generated according to a normal distribution with variance $\sigma_{s_4}^2 = 2.65$, and the other is uniformly distributed with variance $\sigma_{s_5}^2 = 5.33$. The vector of additive noise $\mathbf{v}(t)$ is generated according to a normal distribution. The SNRs of the mixed signals to additive noise are 7–10 dB. The set of noiseless signals (mixture of signals s_1, s_2, s_3 , and $s_4(t) = s_5(t) = 0$, and $\mathbf{v}(t) = \mathbf{0}$) is shown in Fig. 5, and the set of observed signals is shown in the Fig. 6.

Direct application of subspace-based filters for each observed signal gives the results shown in Fig. 7. The characteristics of the filters applied for each individual channel separately were as follows. The length of the FIR filters (see Eq. 12) was $L = 25$. The number of principal components used for estimation of the filter coefficients (see Eq. C6) was the same for each individual channel, and equal $q = 5$.

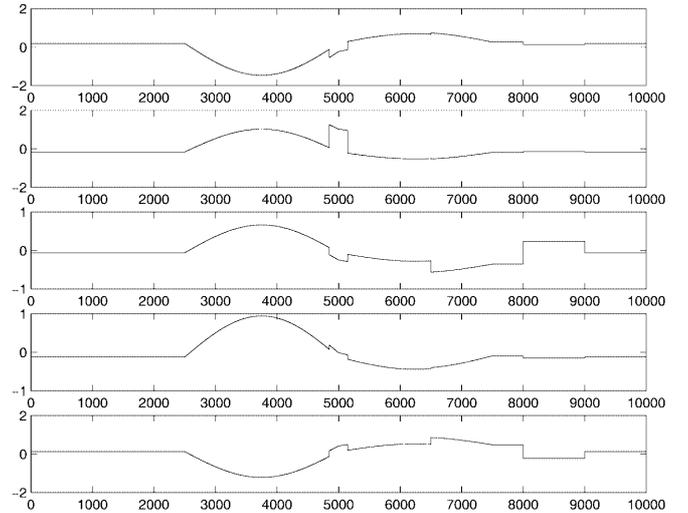


Fig. 5. The set of noiseless mixed signals (assumed to be unknown)

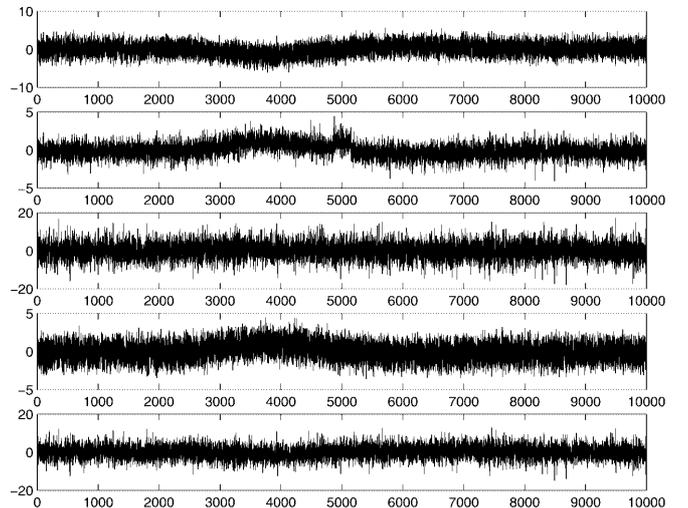


Fig. 6. The set of observed noisy signals

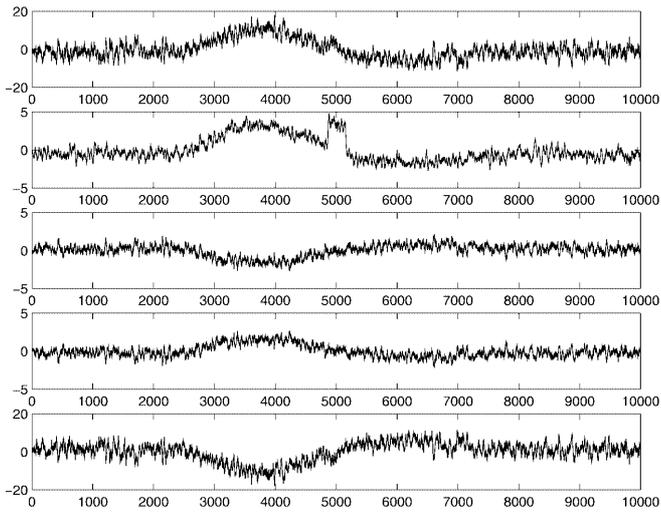


Fig. 7. Filtered signals without separation

Direct filtering of signals does not allow us to separate the “inner” noise, since by direct filtration we filter the measured signals without taking into account the “inner” mixing model (1) of the signals. We can see that the results are rather poor. Generally speaking, the performance of direct application of subspace-based filtering may possibly be slightly improved by choosing the parameters L and q appropriately. However, any filtering techniques used in practice exploit the assumptions that the noise in the model is additive and the SNR is high enough. These assumptions are obviously not valid for the model (1). The first assumption is not valid because we have “inner” noise in the model (1) as well as additive one. However, even if “inner” noises can be thought of as several additive noise components of measured signals, the second assumption is still important for traditional filtering techniques. In fact, “inner” noise signals have similar or often higher power than “interesting” signals in biological applications, which leads to low SNRs for the measured signals. Traditional filtering techniques do not perform well for signals with low SNR, because of subspace swap effect. The poor results of this simulation is caused also because the full SNR (SNR calculated with respect to additive noise as well as “inner” noise) is very low. The SNR equals -15 dB to -25 dB, depending on the channel. The correct estimation of a signal subspace is not possible for such low SNRs.

Note that the conditions of this simulation are well motivated from a practical viewpoint. EEGs that are recorded using surface electrodes are buried in noise which is both electrical and biological in nature, and are characterized by very poor SNRs of less than -20 dB in the presence of larger electrical interference.

The results of our blind noise-reduction algorithm are shown in Figs. 8–10. The extracted ICs (signals separated from the observed mixed signals) are shown in Fig. 8. For the third and fourth signals in Fig. 8 the values of the Hurst exponents are 0.538 and 0.551, respectively. It means that these separated signals are noise components and can be removed. For the first and the second signals we use the subspace-filtering proce-

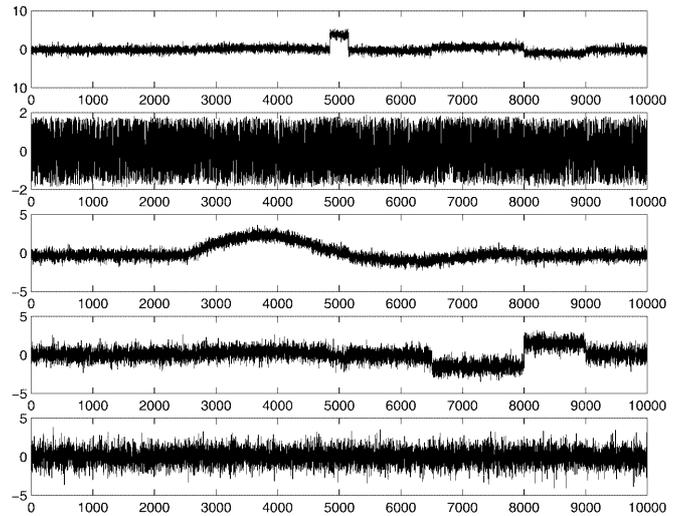


Fig. 8. ICs after applying ICA

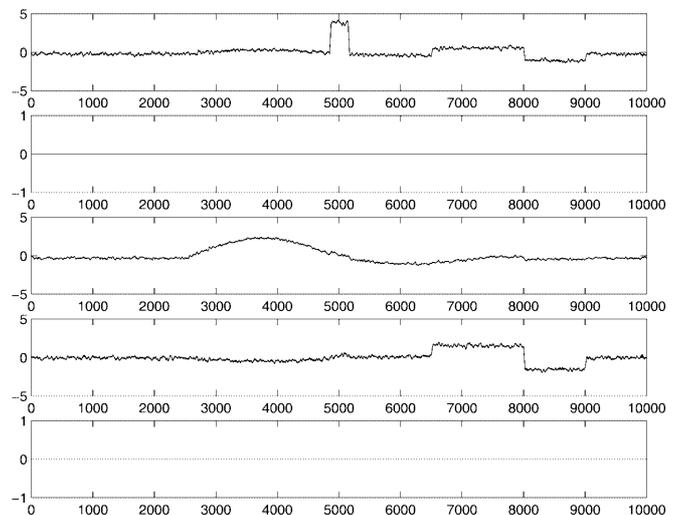


Fig. 9. Filtered ICs

cedure with the same FIR filter characteristics that have been used in previous examples of direct filtering of measured signals. These results are shown in Fig. 9. Finally, we show the reconstructed signals in Fig. 10, which can be compared with the original noiseless signals in Fig. 5. It is not difficult to see even visually (it is, in fact, the purpose for real medical applications) that performance increased dramatically.

4.2 EEG analysis – noise elimination results

Recently, the model (1) has been discussed as a possible model for measured EEG signals (Dogandzic and Nehorai 2000). The “inner” signals are sparse for many mental states. However, we can measure only the noisy mixture of these source signals.

Studies similar to the above were performed for recorded EEG signals as well. We consider the EEG signals recorded from seven positions on the scalp. Thus,

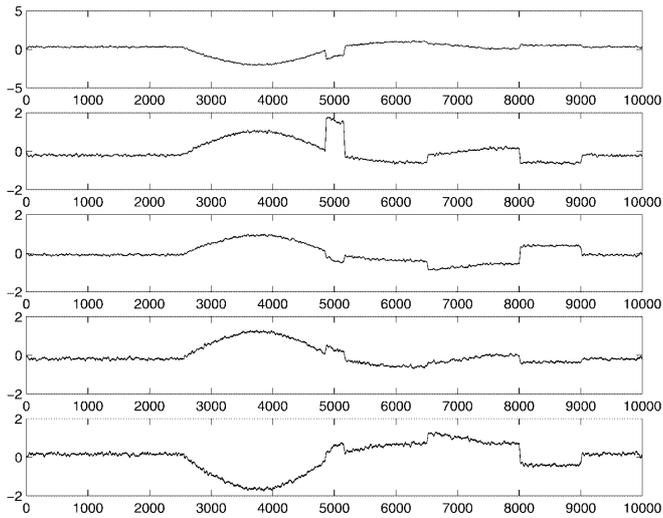


Fig. 10. Reconstructed sensor signals

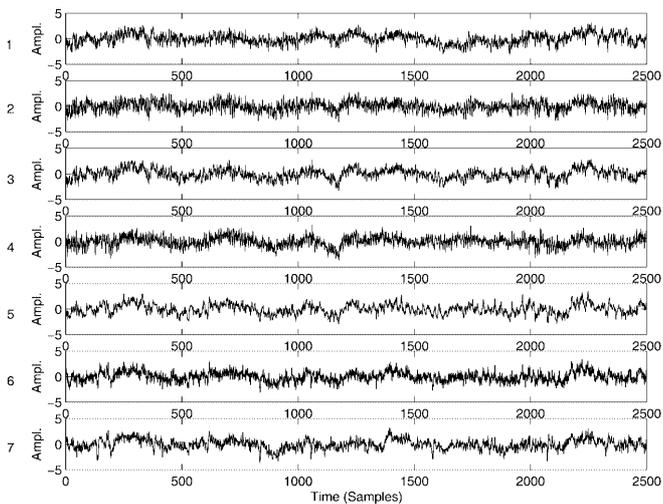


Fig. 11. Recorded row EEG signals

the data matrix is seven rows and 2500 columns, since across the columns are samples taken at 250 Hz for 10 s. Recordings were made with reference to electrically linked mastoids A1 and A2. Recording was performed with a bank of Grass 7P511 amplifiers whose band-pass analog filters were set at 0.1 Hz to 100 Hz. The EEG signals were recorded by Zak Keirn at Purdue University and are accessible at <http://www.cs.colostate.edu/~anderson/res/eeeg/>. The raw recorded EEG signals are shown in Fig. 11. The electrodes are numbered and ordered from 1 to 7.

The result of direct application of subspace filtering to the raw recorded EEG is shown in Fig. 12. For the first and the second EEG signals the filtration results are very poor. This may be explained by the corresponding subspace-based filters not being able to estimate the signal subspace correctly because of a low SNR. The ordering of the EEG signals is the same as in Fig. 11. The characteristics of the subspace filters are the same for each channel: $L = 15$ and $q = 7$. Note that the peak

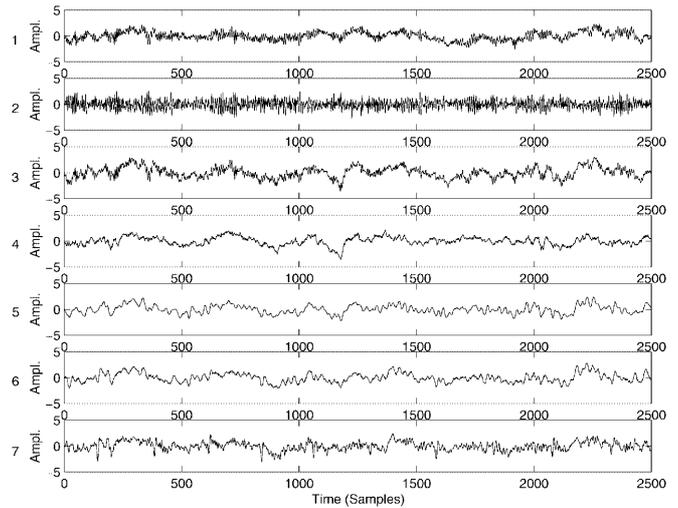


Fig. 12. Filtered EEG signals

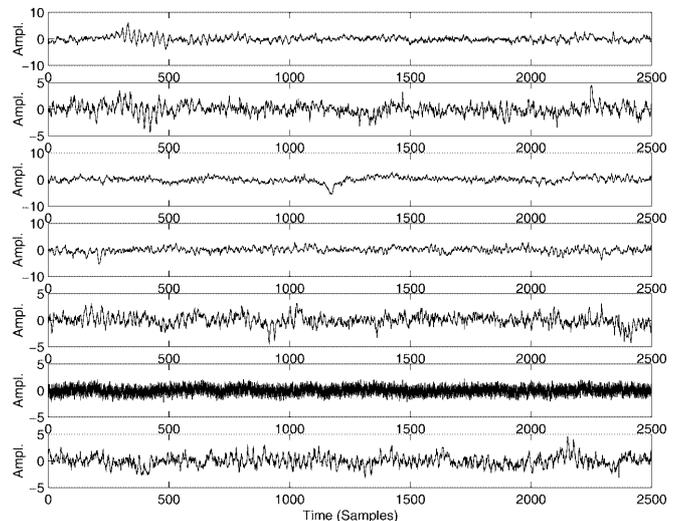


Fig. 13. ICs after applying ICA

between the 1080 and 1220 time samples in the third, fourth and fifth filtered EEG signals correspond to eye blinking. This peak was not removed by direct application of subspace filtering.

The ICA procedure (A1), (A2) is then applied to the set of seven recordings shown in Fig. 11. The ICs shown in Fig. 13 have the following values of the corresponding Hurst exponents: 1: 0.763; 2: 0.785; 3: 0.628; 4: 0.701; 5: 0.778; 6: 0.511; and 7: 0.753. The sixth IC has also the value of kurtosis equal to 0.01. Thus, this component should be rejected as one that has no temporal structure and does not include any “interesting” information. The third IC has a specific value of the Hurst exponent corresponding to the eyes blinking, and should also be removed. The numbering of ICs is arbitrary and does not related to the numbering of recorded EEG signals.

Note here that ICA actually does not solve the inverse problem. Scott Makeig at <http://www.cnl.salk.edu/~scott/tutorial/icafaq.html>, observes that from the

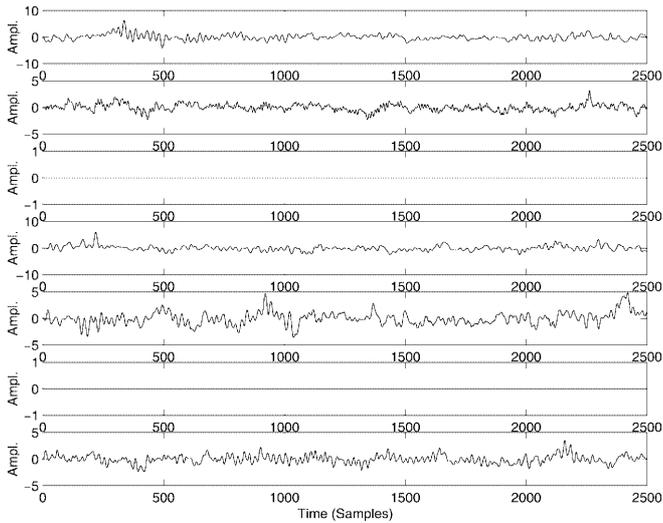


Fig. 14. Filtered ICs

viewpoint of ICA, “brain sources” are not necessarily “fMRI hot spots” or “single equivalent dipoles,” but rather “concurrent electromagnetic activity that is spatially fixed and temporally independent of activity arising in other spatially fixed sources and summed with it in the input data.” Networks producing such concurrent activity are defined not by compact spatial distributions in the brain, but by the covering field measurements they produce at the scalp sensors. In general, “sources” of ICA ICs may be (one or more) distributed brain networks rather than physically compact active brain regions. These networks may be functionally linked, forming a (possibly transient) larger network, or may simply be activated concurrently in the input data, by chance as it were. In our simulation the third and sixth ICs correspond to “biological noise sources,” and can be easily found by the associated values of the Hurst exponent.

The first, second, fourth, fifth and seventh ICs are filtered using the subspace-based filtering technique. The results are shown in Fig. 14. Here the numbering is also not related to the numbering of the recorded EEG signals.

After reconstruction of the filtered ICs back to the sensor level by the inverse operation (18) we obtain the final results presented in Fig. 15. The numbering here is the same as the numbering of recorded EEG signals. In fact, the signals in Fig. 15 are EEG signals with reduced “inner” and additive noise. This results can be compared with raw recorded EEG signals in Fig. 11 and directly filtered EEG signals in Fig. 12.

From the above we conclude that the proposed approach is an effective tool for both artifact classification and removing, and electrical noise elimination from contaminated EEG signals.

5 Conclusions

In this paper, we explored the method for blind noise reduction in EEG signals using ICA and subspace

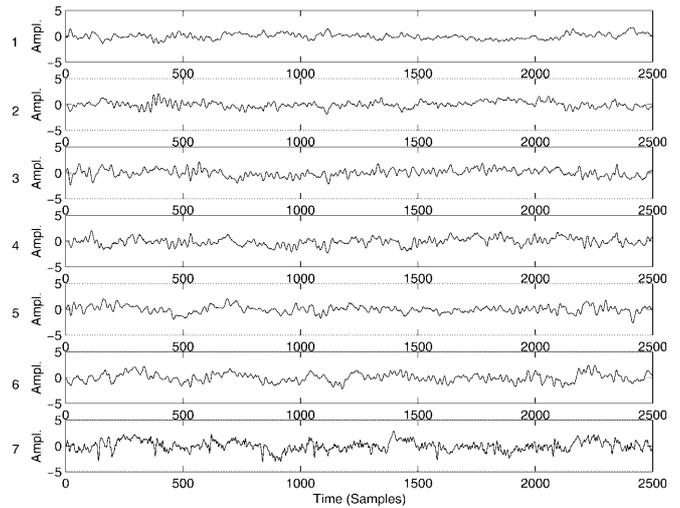


Fig. 15. Reconstructed EEG signals

filtering. The key point which makes the ICA technique important and promising for the blind noise-reduction problem is the “inner” structure of observed signals, which are nothing but a noisy mixture of some signals from real sources and the “inner” noise signals. For detection of ICs containing “interesting” signals (signals of interest), we apply the procedure that is based on the Hurst exponent calculation. For filtering of “interesting” ICs after separation of the mixture, we apply the subspace filtering method. The simulation as well as real applications of the proposed method demonstrates the effectiveness of the proposed approach. On the other hand, direct application of filters to measured signals does not take into account the special structure of measured signals and, therefore, does not allow us to obtain acceptable results of noise reduction. Observed EEG data can be described using the same mixing model that has been discussed in the paper. The real application shows the effectiveness of the proposed method for blind noise reduction in EEG data. The proposed algorithm presents the performance of only the most basic schema. Further improvements can be achieved by “optimizing” the algorithm using, for example, low-rank adaptive-tracking filtering or a robust prewhitening procedure for the case of coherent additive noise.

Appendix A: Summary of the procedure for blind source extraction in Gaussian noise

Let us explain in more detail what the block “ICA demixing system” in Fig. 1 makes. The blind source extraction procedure (Cruces et al. 2000) consists of two stages. In the first stage we identify the mixture matrix \mathbf{A} that is unknown using following iteration:

$$\hat{\mathbf{A}}(t+1) = \hat{\mathbf{A}}(t) + \mu(t) \left(\mathbf{C}_{x,y}^{1,3} \mathbf{S}_y^3 - \hat{\mathbf{A}}(t) \right), \quad (\text{A1})$$

where $\hat{\mathbf{A}}(t)$ is estimation of the unknown mixture matrix \mathbf{A} at step t ; $\mu(t)$ is a learning-rate parameter;

$\mathbf{C}_{x,y}^{1,3} = M_{x,y}^{1,3} - 3M_{x,y}^{0,2}M_{x,y}^{1,1}$ is the fourth-order cross-cumulant matrix; $M_{x,y}^{k,l} = E\{\underbrace{x \times \dots \times x}_k \underbrace{(y \times \dots \times y)}_l\}^T$; and \mathbf{S}_y^3 is the diagonal matrix of cumulant signs $[S_y^3]_{ii} = \text{sign}([C_{x,y}^{1,3}]_{ii})$.

At the second stage we calculate immediately a separation matrix $\mathbf{W}(t)$ based on estimation of mixture matrix $\hat{\mathbf{A}}(t)$:

$$\mathbf{W}(t) = (\hat{\mathbf{A}}^T(t)\mathbf{R}_{xx}^{-1}\hat{\mathbf{A}}(t))^{-1}\hat{\mathbf{A}}^T(t)\mathbf{R}_{xx}^{-1}, \quad (\text{A2})$$

where $\mathbf{R}_{xx} = \frac{1}{t} \sum_{i=1}^t \mathbf{x}(i)\mathbf{x}^T(i)$ is a correlation matrix of the observations.

As it was shown in Cruces et al. 2000, this algorithm is robust in the sense that for a additive Gaussian noise presence in observations, the obtained estimates are asymptotically unbiased. Let us note also that a stability condition for this algorithm is $\mu < \frac{1}{2}$.

Appendix B: Hurst exponent recurrent calculation

The recurrent method for calculation of the Hurst exponent (Vorobyov and Shilo 1998) can be written in the form that follows. It follows from (11) that

$$H(t+1) = \frac{\log_{10}\left(\frac{R(t+1)}{S(t+1)}\right)}{\log_{10}\left(\frac{t+1}{2}\right)}, \quad (\text{B1})$$

where the range

$$R(t+1) = Y_{\max}(t+1) - Y_{\min}(t+1) \quad (\text{B2})$$

is recursively defined as a difference between maximum and minimum ‘‘accumulated’’ values

$$Y_{\max}(t+1) = \begin{cases} Y(t+1), & Y(t+1)Y_{\max}(t), \\ Y_{\max}(t+1), & Y(t+1) \leq Y_{\max}(t), \end{cases} \quad (\text{B3})$$

$$Y_{\min}(t+1) = \begin{cases} Y(t+1), & Y(t+1) < Y_{\min}(t), \\ Y_{\min}(t+1), & Y(t+1) \geq Y_{\min}(t). \end{cases} \quad (\text{B4})$$

Here $Y(t+1)$ defined in (9) is recursively computed as

$$Y(t+1) = Y(t) + \frac{t}{t+1}(y(t+1) - \langle y(t) \rangle), \quad (\text{B5})$$

and $\langle y(t) \rangle$ is averaging of the process $y(t)$:

$$\langle y(t) \rangle = \frac{t-1}{t} \langle y(t-1) \rangle + \frac{1}{t} y(t). \quad (\text{B6})$$

The standard deviation $S(t)$ (10) is recursively computed according to the following obvious equations:

$$S(t+1) = D^{\frac{1}{2}}(t+1),$$

$$D(t+1) = \frac{t}{t+1}D(t) + \frac{t^2}{(t+1)^3}(y(t+1) - \langle y(t) \rangle)^2. \quad (\text{B7})$$

The initial conditions are $Y_{\max}(0) = Y_{\min}(0) = y(0)$ and $H(0) = D(0) = Y(0) = 0$.

Appendix C: Calculation of $\hat{\mathbf{a}}_i^{\text{pc}}$ in (14)

The autocorrelation matrix $\hat{\mathbf{R}}_i$ in the Wiener filter (13) is

$$\hat{\mathbf{R}}_i = \begin{pmatrix} \hat{r}_{\tilde{y}_i\tilde{y}_i}(0) & \hat{r}_{\tilde{y}_i\tilde{y}_i}(-1) & \dots & \hat{r}_{\tilde{y}_i\tilde{y}_i}(-L-1) \\ \hat{r}_{\tilde{y}_i\tilde{y}_i}(1) & \hat{r}_{\tilde{y}_i\tilde{y}_i}(0) & \dots & \hat{r}_{\tilde{y}_i\tilde{y}_i}(-L) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{r}_{\tilde{y}_i\tilde{y}_i}(L-1) & \hat{r}_{\tilde{y}_i\tilde{y}_i}(L-2) & \dots & \hat{r}_{\tilde{y}_i\tilde{y}_i}(0) \end{pmatrix}. \quad (\text{C1})$$

Hence, $\hat{\mathbf{R}}_i$ is $L \times L$ positive definite Hermitian matrix with components estimated based on IC \tilde{y}_i according to

$$\hat{r}_{\tilde{y}_i\tilde{y}_i}(k) = \begin{cases} \frac{1}{N} \sum_{t=1}^{N-1-k} \tilde{y}_i(t)\tilde{y}_i(t+k), & \text{for } k = 0, 1, \dots, L, \\ \hat{r}_{\tilde{y}_i\tilde{y}_i}^*(-k), & \text{for } k = -L+1, -L+2, \dots, -1. \end{cases} \quad (\text{C2})$$

The $L \times 1$ autocorrelation lag vector $\hat{\mathbf{r}}_i$ in the Wiener filter (13) is

$$\hat{\mathbf{r}}_i = \begin{pmatrix} \hat{r}_{\tilde{y}_i\tilde{y}_i}(1) \\ \hat{r}_{\tilde{y}_i\tilde{y}_i}(2) \\ \vdots \\ \hat{r}_{\tilde{y}_i\tilde{y}_i}(L) \end{pmatrix}. \quad (\text{C3})$$

The eigenvalue decomposition of autocorrelation matrix $\hat{\mathbf{R}}_i$ can be given as

$$\hat{\mathbf{R}}_i = \sum_{j=1}^L \lambda_j \mathbf{u}_j \mathbf{u}_j^T, \quad (\text{C4})$$

where λ_j are the eigenvalues of the matrix $\hat{\mathbf{R}}_i$ and \mathbf{u}_i are the eigenvectors correspondent to eigenvalues. Eigenvalues λ_j are real and positive, and eigenvectors are orthonormal because $\hat{\mathbf{R}}_i$ is positive definite. Hence, we can write the Wiener filter (13) in the form

$$\hat{\mathbf{c}}_i = - \sum_{j=1}^L \frac{1}{\lambda_j} \mathbf{u}_j \mathbf{u}_j^T \mathbf{r}_i. \quad (\text{C5})$$

However, depending on the eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_q \geq \lambda_{q+1} \geq \dots \geq \lambda_L$, we can use only the first q dominant principal components with the largest eigenvalues:

$$\hat{\mathbf{c}}_i^{\text{pc}} = - \sum_{j=1}^q \frac{1}{\lambda_j} \mathbf{u}_j \mathbf{u}_j^T \mathbf{r}_i. \quad (\text{C6})$$

Hence, (C6) uses only q eigenvectors that belong to the signal subspace. Often q can be easily defined from the distribution of eigenvalues, but some knowledge about shapes of event-related brain signals also can give

additional information for choosing q . For example, $q = 2$ is enough for the representation of sine wave signals.

References

- Akay M (1996) Detection and estimation methods for biomedical signals. Academic Press, Inc., San Diego
- Amari S (1999) ICA of temporally correlated signals—learning algorithm. In: Proceedings of ICA'99, Aussois, France, January 11–15, pp 13–18
- Arnold M, Miltner WH, Witte H, Bauer R, Braun C (1998) Adaptive AR modeling of nonstationary time series by means of Kalman filtering. *IEEE Trans Biomed Eng* 45: 553–562
- Attias H (2000) Independent factor analysis with temporally structured factors In: Leen T, et al. (eds) *Advances in neural information processing systems*, 12. MIT Press, Cambridge, Mass.
- Callaerts D et al (1988) On-line algorithm for signal separation based on SVD. In: Deprettere, EF (ed) *SVD and signal processing: algorithms, applications and architectures*, Elsevier, Amsterdam, pp 269–276
- Choi S, Cichocki A (2000) Blind separation of nonstationary sources in noisy mixtures. *Electron Lett* 36: 848–849
- Cichocki A, Amari S (2002) *Adaptive Blind Signal and Image Processing*, Wiley, Chichester
- Cichocki A, Vorobyov SA (2000) Application of ICA for automatic noise and interference cancellation in multisensory biomedical signals, In: Proceedings of the Second International Workshop on Independent Component Analysis and Blind Signal Separation, Helsinki, Finland, June 19–22, pp, 621–626
- Cruces S, Cichocki A, Castedo L (2000) An iterative inversion approach to blind source separation. *IEEE Trans Neural Netw* 11: 1423–1437
- Dogandzic A, Nehorai A (2000) Estimating evoked dipole responses in unknown spatially correlated noise with EEG/MEG arrays. *IEEE Trans Signal Process* 48: 13–25
- Goldstein R, Aldrich WM (1999) *Evoked potential audiometry. Fundamentals and applications*, Allyn and Bacon, Boston
- Haykin S (1996) *Adaptive filter theory*, 3rd edn. Prentice Hall, Upper Saddle River, N.J.
- Hurst HE, Black RP, Simaika YM (1965) *Long term storage*. Constable, London
- Katz MJ (1988) Fractals and the analysis of waveforms. *Comput Biol Med* 18: 145–155
- Laguna P, Moody GB, Garcia J, Goldberger AL, Mark RG (1999) Analysis of the ST-T complex of the electrocardiogram using the Karhunen–Loève transform: adaptive monitoring and alternans detection. *Med Biol Eng Comput* 37: 175–189
- Lee TW (1998) *Independent component analysis. Theory and applications*, Kluwer Academic, Boston
- Nijima S, Amari S (2002) MEG source estimation using fourth order MUSIC method, *IEICE Trans Information and Systems* E85-D: 167–174
- Niedermeyer E, da Silva FL (1999) *Electroencephalography. Basic principles, clinical applications, and related fields*, 4th edn. Williams and Wilkins, Philadelphia, Pa.
- Pisarenco VF (1973) The retrieval of harmonics from a covariance function, *Geophys J R Astron Soc* 33: 347–366
- Raghothaman B, Linebarger DA, Begusic D (2000) A new method for low rank transform domain adaptive filtering. *IEEE Trans Signal Process* 48:1097–1109
- Sadasivan PK, Dutt DN (1996) SVD based technique for noise reduction in electroencephalographic signals. *Signal Process* 55: 179–189
- Strobach P (1996) Low rank adaptive filters. *IEEE Trans Signal Process* 44: 2932–2947
- Thakor NV (1987) Adaptive filtering of evoked potentials. *IEEE Trans Biomed Eng* 34: 6–12
- Turner KM (1992) Reliability of storage scheme, tree rings, and Hurst phenomena, In: Proceedings of the American Geophysical Union Twelfth Annual Hydrology Days, Colorado State University, Fort Collins, Colorado, April, Hydrology Days Publications, Atherton, California, pp 230–242
- van der Veen A-J, Deprettere EDF, Swindlehurst AL (1993) Subspace based signal analysis using singular value decomposition. *IEEE Proceedings* 81: 1277–1308
- Vorobyov SA, Shilo AV (1998) The recurrent normalized range method for revelation of fractal structure and its application to analysis of EEG (in Russian). *Radioelectron Inform* 3: 162–165
- Walter DO (1969) A posteriori filtering of average evoked responses. *Electroencephalography Clin Neurophysiol* 27: 61–70
- Weerd JPC de, Martens WLJ (1978) Theory and practice of a posteriori Wiener filtering of averaged evoked potentials. *Biol Cybern* 30: 81–94
- Widrow B, Walach E (1996) *Adaptive inverse control*. Prentice Hall, Upper Saddle River, N.J.