BLIND SIGNAL EXTRACTION OF SIGNALS WITH SPECIFIED FREQUENCY BAND

Andrzej Cichocki, Tomasz Rutkowski and Krzysztof Siwek
Laboratory for Advanced Brain Signal Processing,
Brain Science Institute, Riken,
2-1 Hirosawa, Wako-shi, Saitama 351-0198, JAPAN
Phone: (+81) 48 467 9668
Fax: (+81) 48 467 9686
E-mail: cia@bsp.brain.riken.go.jp
http://www.bsp.brain.riken.go.jp

Abstract. Blind Sources Separation, Independent Component Analysis (ICA) and related methods are promising approaches for analysis of biomedical signals, especially for EEG/MEG and fMRI data. However, most of the methods extract all sources simultaneously, so it is time consuming and not reliable especially, when the number of sensors is large (more than 100 sensors) and signals are contaminated by huge noise. Main objective of this paper is to present a new method for extraction of specific source signals using bandpass filters approach. Such a method allows us to extract source signals with specific stochastic properties, e.g., extraction narrow band sources with specific frequency bandwidth.

INTRODUCTION AND PROBLEM DETAILED ELABORATION

The problem of independent Component Analysis (ICA) and related problems like blind sources separation (BSS) and blind source extraction (BSE) have received wide attention in various fields such as biomedical signal analysis and processing (EEG, MEG, fMRI), speech enhancement, geophysical data processing, data mining, wireless communications and image processing [1-17].

There are two principal approaches to solve such problems. The first approach separates all sources simultaneously [2, 9, 12, 13, 14]. In the second approach, we extract sources one by one sequentially rather than separating them all together. In many applications, a large number of sensors (electrodes, microphones or transducers) are available but only a very few source signals are the subject of interest. For example, in high density array EEG or MEG systems, we observe typically more than 100 sensor signals and only a few source signals are considered interesting, the rest are considered to be interfering noise. Another example is the cocktail party problem; it is usually applied to extract voices of specific persons rather than separate all the available source signals from an array of microphones. For such applications, it is essential to develop reliable, robust and effective learning algorithms which enable us to extract only the small number of source signals that are potentially interesting and contain useful information

The blind signal extraction approach may have the following advantages over simultaneous blind separation [2, 4, 7]:

1. Signals can be extracted in a specified *order* according to the stochastic features of the source signals, (e.g., in the order determined by absolute values of generalized normalized kurtosis, some measures of sparseness, non-Gaussianity, smoothness or linear predictability.) The blind extraction can be considered as a generalization of PCA (principal components analysis), where decorrelated output signals (principal components) are extracted according to the decreasing order of their variances. Analogously, independent components can be ordered according for example to the decreasing absolute value of normalized kurtosis which is a measure of non-Gaussianity or according to any higher order normalized moment or cumulant.

2. The approach is flexible, because many different criteria based on HOS and SOS can be applied for extraction of wide spectrum of sources, like i.i.d. sources, colored Gaussian, sparse sources, non-stationary sources, smooth sources with relative high measure of predictability, etc. In fact in each stage of extraction, we can use various criteria and corresponding algorithms depending on requirement to extract sources with specific features.

3. Only "interesting" signals need to be extracted. For example, if the source signals are mixed with a large number of noise sources or interferences, we may extract only signals with some desired stochastic properties. For example, in EEG/MEG biomedical signal processing is often desired to extract brain waves (e.g., alpha, beta, gamma, theta waves) with specific frequency bandwidth. Moreover, it important to extract so called evoked potentials with non-symmetric distributions from symmetric distributed noises and interferences of on going brain activity [2, 9, 10, 17].

The main objective of this paper is presentation a family of simple robust algorithms for blind source extraction suitable for analysis of noisy biomedical signals.

BLIND SOURCE EXTRACTION USING LINEAR FIR FILTER

Let us assume for simplicity, that we want to extract only one source signal with temporal structure , say $s_j(k)$, from the available sensor vector $\mathbf{x}(k) =$

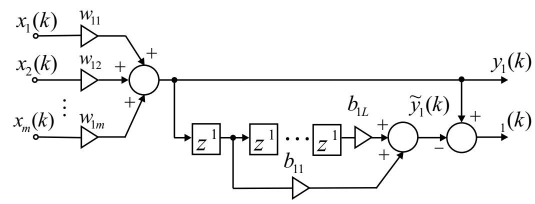

Figure 1: The neural network structure for one-unit extraction using adaptive FIR filter.

$\mathbf{H}\,\mathbf{s}(k)+\boldsymbol{\nu}(k)$, where $\mathbf{H}$ is a full column rank mixing matrix. For this purpose, we employ the single processing unit described as (see Fig.1):

$$y_1(k) = \mathbf{w}_1^T \mathbf{x}(k) = \sum_{i=1}^{m} w_{1i} x_i(k), \tag{1}$$

$$\begin{aligned} \varepsilon_1(k) &= y_1(k) - \sum_{p=1}^{L} b_{1p} y_1(k-p) \\ &= \mathbf{w}_1^T \mathbf{x}(k) - \mathbf{b}_1^T \bar{\mathbf{y}}_1(k), \end{aligned} \tag{2}$$

where $\mathbf{w}_1 = [w_{11}, w_{12}, \ldots, w_{1m}]^T$, $\bar{\mathbf{y}}_1(k) = [y_1(k-1), \ldots, y_1(k-L)]^T$, $\mathbf{b}_1 = [b_{11}, b_{12}, \ldots, b_{1L}]^T$ and $B_1(z) = \sum\limits_{p=1}^{L} b_{1p} z^{-p}$ is a transfer function of the corresponding FIR filter. It should be noted that the FIR filter can have a sparse representation, in particular, only one single processing unit, say with delay p and $b_{1p} \neq 0$ can be used instead of L parameters [2, 3].

The processing unit has two outputs: $y_1(k)$ which estimates the extracted source signals, and $\varepsilon_1(k)$, which represents a prediction error or innovation, after passing the output signal $y_1(k)$ through FIR filter.

In general, our objective is to estimate the optimal values of vectors $\mathbf{w}_1$ and $\mathbf{b}_1$, in such a way that the processing unit successfully extracts one of the sources. This is achieved, if the global vector defined as $\mathbf{g}_1 = \mathbf{H}^T \mathbf{w}_1 = \left(\mathbf{w}_1^T \mathbf{H}\right)^T = c_j \mathbf{e}_j$ contains only one nonzero element, say in the j-th row, such that $y_1(k) = c_j s_j$, where c_j is an arbitrary nonzero scaling factor. For this purpose, we reformulate the problem as a minimization of the cost function

$$\mathcal{J}(\mathbf{w}_1, \mathbf{b}_1) = E\left\{\varepsilon_1^2\right\}. \tag{3}$$

subject to the constraint $||\mathbf{w}_1|| = 1$,
where $\varepsilon_1(k) = y_1(k) - \widetilde{y}_1(k)$, $y_1 = \mathbf{w}_1^T \mathbf{x}_1$, $\mathbf{x}_1 = \mathbf{Q}\mathbf{x}$, $\mathbf{Q}$ is optional prewhitening matrix, $\widetilde{y}_1(k) = \mathbf{b}_1^T \bar{\mathbf{y}}_1(k) = \sum_{p=1}^{L} b_{1p} y_1(k-p)$ and $\bar{\mathbf{y}}_1 = [y_1(k-1), \ldots, y_1(k-L)]^T$.

The main motivation of applying such a cost function is the assumption that primary sources signals (signals of interest) have temporal structures and can be modeled, e.g., by the autoregressive model [1, 3, 9].
The cost function can be evaluated as follows:

$$E\left\{\varepsilon_1^2\right\} = \mathbf{w}_1^T \widehat{\mathbf{R}}_{\mathbf{x}_1\mathbf{x}_1} \mathbf{w}_1 - 2\mathbf{w}_1^T \widehat{\mathbf{R}}_{\mathbf{x}_1\bar{\mathbf{y}}_1} \mathbf{b}_1 + \mathbf{b}_1^T \widehat{\mathbf{R}}_{\bar{\mathbf{y}}_1\bar{\mathbf{y}}_1} \mathbf{b}_1, \tag{4}$$

where $\widehat{\mathbf{R}}_{\mathbf{x}_1\mathbf{x}_1} \simeq E\{\mathbf{x}_1\mathbf{x}_1^T\}$, $\widehat{\mathbf{R}}_{\mathbf{x}_1\bar{\mathbf{y}}_1} \approx E\{\mathbf{x}_1\bar{\mathbf{y}}_1^T\}$ and $\widehat{\mathbf{R}}_{\bar{\mathbf{y}}_1\bar{\mathbf{y}}_1} \approx E\{\bar{\mathbf{y}}_1\bar{\mathbf{y}}_1^T\}$, are the estimators of the true values of correlation and cross-correlation matrices: $\mathbf{R}_{\mathbf{x}_1\mathbf{x}_1}, \mathbf{R}_{\mathbf{x}_1\bar{\mathbf{y}}_1}, \mathbf{R}_{\bar{\mathbf{y}}_1\bar{\mathbf{y}}_1}$, respectively. In order to estimate vectors $\mathbf{w}_1$ and $\mathbf{b}_1$, we evaluate the gradients of the cost function and equalize them to zero as follows:

$$\frac{\partial \mathcal{J}_1(\mathbf{w}_1, \mathbf{b}_1)}{\partial \mathbf{w}_1} = 2\widehat{\mathbf{R}}_{\mathbf{x}_1\mathbf{x}_1}\mathbf{w}_1 - 2\widehat{\mathbf{R}}_{\mathbf{x}_1\bar{\mathbf{y}}_1}\mathbf{b}_1 = \mathbf{0}, \tag{5}$$

$$\frac{\partial \mathcal{J}_1(\mathbf{w}_1, \mathbf{b}_1)}{\partial \mathbf{b}_1} = 2\widehat{\mathbf{R}}_{\bar{\mathbf{y}}_1\bar{\mathbf{y}}_1}\mathbf{b}_1 - 2\widehat{\mathbf{R}}_{\bar{\mathbf{y}}_1\mathbf{x}_1}\mathbf{w}_1 = \mathbf{0}. \tag{6}$$

Solving the above matrix equations, we obtain an iterative algorithm

$$\mathbf{w}_1^+ = \widehat{\mathbf{R}}_{\mathbf{x}_1\mathbf{x}_1}^{-1}\widehat{\mathbf{R}}_{\mathbf{x}_1\bar{\mathbf{y}}_1}\mathbf{b}_1, \quad \mathbf{w}_1 = \frac{\mathbf{w}_1^+}{||\mathbf{w}_1^+||}, \tag{7}$$

$$\mathbf{b}_1 = \widehat{\mathbf{R}}_{\bar{\mathbf{y}}_1\bar{\mathbf{y}}_1}^{-1}\widehat{\mathbf{R}}_{\bar{\mathbf{y}}_1\mathbf{x}_1}\mathbf{w}_1 = \widehat{\mathbf{R}}_{\bar{\mathbf{y}}_1\bar{\mathbf{y}}_1}^{-1}\widehat{\mathbf{R}}_{\bar{\mathbf{y}}_1 y_1}, \tag{8}$$

where the matrices $\widehat{\mathbf{R}}_{\bar{\mathbf{y}}_1\bar{\mathbf{y}}_1}$ and $\widehat{\mathbf{R}}_{\bar{\mathbf{y}}_1 y_1}$ are estimated on the basis of the parameters $\mathbf{w}_1$ obtained in the previous iteration step.
In order to avoid the trivial solution $\mathbf{w}_1 = \mathbf{0}$, we normalize the vector $\mathbf{w}_1$ to unit length in each iteration step as $\mathbf{w}_1(l+1) = \mathbf{w}_1^+(l+1)/\left\|\mathbf{w}_1^+(l+1)\right\|$ (which ensures that $E\{y_1^2\} = 1$).

It should be emphasized here that in the derivation $\widehat{\mathbf{R}}_{\bar{\mathbf{y}}_1\bar{\mathbf{y}}_1}$ and $\widehat{\mathbf{R}}_{\bar{\mathbf{y}}_1 y_1}$ are assumed to be independent of the actually evaluated vector $\mathbf{w}_1(l+1)$, i.e., they are estimated based on $\mathbf{w}_1(l)$ in the previous iteration step. This two-phase procedure is similar to the expectation maximization (EM) scheme: (i) freeze the correlation and cross-correlation matrices and learn the parameters of the processing unit $(\mathbf{w}_1, \mathbf{b}_1)$; (ii) freeze $\mathbf{w}_1$ and $\mathbf{b}_1$ and learn new statistics (i.e., matrices $\widehat{\mathbf{R}}_{\bar{\mathbf{y}}_1 y_1}$ and $\mathbf{R}_{\bar{\mathbf{y}}_1\bar{\mathbf{y}}_1}$) of the estimated source signal, then go back to (i) and repeat. Hence, in phase (i) our algorithm extracts a source signal, whereas in phase (ii) it learns the statistics of the source.

The above algorithm can be considerably simplified. It should be noted, that in order to avoid the inversion of the autocorrelation matrix $\mathbf{R}_{\mathbf{x}_1\mathbf{x}_1}$ in each iteration step, we can apply as a preprocessing the robust pre-whitening or standard PCA and next normalize the sensor signals to unit variance [2]. In such cases, $\widehat{\mathbf{R}}_{\mathbf{x}_1\mathbf{x}_1} = \mathbf{I}_n$ and the algorithm is simplified to [2, 3]

$$\mathbf{w}_1^+ = \widehat{\mathbf{R}}_{\mathbf{x}_1\bar{\mathbf{y}}_1}\mathbf{b}_1 = \widehat{\mathbf{R}}_{\mathbf{x}_1\widetilde{y}_1}, \quad \mathbf{w}_1 = \frac{\mathbf{w}_1^+}{||\mathbf{w}_1^+||} \tag{9}$$

where $\widehat{\mathbf{R}}_{\mathbf{x}_1\widetilde{y}_1} = \frac{1}{N}\sum_{k=1}^{N}\mathbf{x}_1(k)\,\widetilde{y}_1(k)$.

It is interesting to note that the algorithm can be formulated in an equivalent form as

$$\mathbf{w}_1(l+1) = \frac{\langle \mathbf{x}_1(k)\widetilde{y}_1(k)\rangle}{\langle y_1^2(k)\rangle}. \tag{10}$$

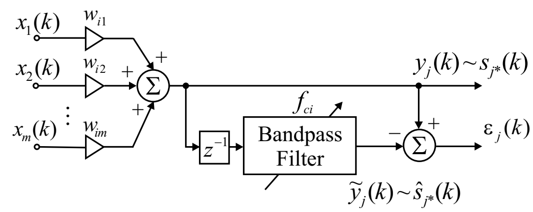

Figure 2: The conceptual model of single processing unit for extraction of sources using an adaptive IIR bandpass filter.

In order to reduce bias caused by white additive noise, we can modify the formula (10) as

$$\mathbf{w}_1(l+1) = \frac{\langle \mathbf{x}_1(k)\widetilde{y}_1(k)\rangle}{\langle y_1(k)\widetilde{y}_1(k)\rangle}. \tag{11}$$

From analysis (9)-(10) it follows that our algorithm is similar to the power method for finding the eigenvector $\mathbf{w}_1$ associated with the maximal eigenvalue of the matrix $\mathbf{R}_{\mathbf{x}_1}(\mathbf{b}_1) = E\{\sum_{p=1}^{L} b_{1p}\mathbf{x}_1(k)\mathbf{x}_1^T(k-p)\}$. This observation suggests that it is not needed to minimize the cost function with respect to parameters $\{b_{1p}\}$ but it is enough to choose the arbitrary set of them for which the largest eigenvalue is unique. More generally, if all eigenvalues of the generalized covariance matrix $\mathbf{R}_{\mathbf{x}_1}(\mathbf{b}_1)$ are distinct, then we can extract all sources simultaneously by estimating principal eigenvectors of $\mathbf{R}_{\mathbf{x}_1}(\mathbf{b}_1)$.

BLIND EXTRACTION OF SOURCES USING A BANK BAND-PASS FILTERS

For noisy data instead of linear predictor (FIR filter), we can use a bandpass IIR filter (or in a parallel way several processing units with a bank of bandpass filters) with fixed or adjustable center frequency and a bandpass bandwidth. The approach is illustrated in Fig. 2. By minimizing the cost function

$$\mathcal{J}(\mathbf{w}_j) = E\{\varepsilon_j^2\} \tag{12}$$

subject to constraint $\|\mathbf{w}_j\|_2 = 1$, we obtain the on-line learning rule

$$\Delta\mathbf{w}_j(k) = -\eta_1(k)\ [\langle \varepsilon_j(k)\widetilde{\mathbf{x}}_1(k)\rangle - \gamma_j(k)\mathbf{w}_1(k)] \tag{13}$$

where $\gamma_j(k) = -\beta_j\,[1 - \hat{\sigma}_{y_j}^2(k)]$ is a forgetting factor, $\mathbf{x}_1(k) = \mathbf{Q}\,\mathbf{x}(k)$, $\widetilde{\mathbf{x}}_1(k) = \mathbf{x}_1(k) - B_j(z)\mathbf{x}_1(k)$ and $B_j(z)$ means transfer function of the bandpass filters.

For pre-whitened sensor signals, we can easily derive simple batch algorithm as

$$\mathbf{w}_j(l+1) = \frac{\langle \mathbf{x}_1(k)\widetilde{y}_j(k)\rangle}{\langle y_j^2(k)\rangle} \tag{14}$$

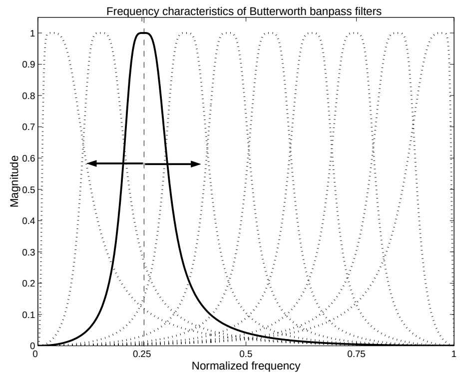

Figure 3: Exemplary frequency characteristics of 4-th order Buttherworth bandpass filter with adjustable center frequency and fixed bandwidth.

or alternatively

$$\begin{aligned}\mathbf{w}_j^{+}(l+1) &= \langle \mathbf{x}_1(k)\tilde{y}_j(k)\rangle = \frac{1}{N}\sum_{k=1}^{N}\mathbf{x}_1(k)\tilde{y}_j(k), \\ \mathbf{w}_j(l+1) &= \frac{\mathbf{w}_j^{+}(l+1)}{||\mathbf{w}_j^{+}(l+1)||}, \end{aligned} \tag{15}$$

where $y_j(k) = \mathbf{w}_j^T(l)\,\mathbf{x}_1(k)$, $\tilde{y}_j(k) = [B_j(z)]y_j(k) = \tilde{\mathbf{x}}_1^T(k)\mathbf{w}_j(l)$.

The above algorithms extract successfully sources if the covariance matrix $\mathbf{R}_{\mathbf{x}_1\tilde{\mathbf{x}}_1} = E\{\mathbf{x}_1\tilde{\mathbf{x}}_1\}$ has unique maximum eigenvalue. It should be noted that the proposed algorithm (15) is insensitive to white noise and even colored noise which is out off the bandwidth of the bandpass filter. Moreover, the processing unit is able to extract the filtered version of a source signal if it is narrow band signal. In practical implementation we can use the 4-th order Buttherworth filter with easy adjustable central frequency and bandwidth [2, 8]. In contrast to the second order BPF the forth-order Buttherworth filter has a flat characteristic around central frequency and enable enhance arbitrary narrow-band source signal with low distortion (see Fig.3). By changing or adjusting the center frequency and bandwidth of the band pass filter, we can extract different narrow band sources using the some processing unit. We can also extract sources simultaneously by employing several processing units with bandpass filters with different bandwidths and center frequencies.

In fact, the above algorithm algorithm can be also considered as modified version of power method for finding the eigenvector corresponding to the maximal eigenvalue of the time-delayed covariance matrix $\mathbf{R}_{\mathbf{x}_1\,\tilde{\mathbf{x}}_1}$. This means that the problem is equivalent to the eigenvalue problem of finding an eigenvector $\mathbf{w}$ corresponding to the largest eigenvalue of the covariance matrix $\mathbf{R}_{\mathbf{x}_1\,\tilde{\mathbf{x}}_1}$, thus any efficient algorithm for estimation extremal eigenvalue and associated eigenvector can be employed. The problem has a solution if this largest eigenvalue is distinct from the other eigenvalues. If the largest eigenvalue is multiple, we must choose the FIR filter with more delays.

It should be noted that the convergence rate of the algorithm depends on a ratio λ_2/λ_{max}, where λ_2 is the second largest eigenvalue of $\mathbf{R}_{\mathbf{x}_1\tilde{\mathbf{x}}_1}$. This ratio is generally smaller than one, allowing adequate convergence of the algorithm. However, if the eigenvalue $\lambda_1 = \lambda_{max}$ has one or more other eigenvalues of $\mathbf{R}_{\mathbf{x}_1\tilde{\mathbf{x}}_1}$ close by, in other words, when λ_1 belongs to a cluster of eigenvalues then the ratio can be very close to one, causing very slow convergence and in consequence the estimated eigenvector $\mathbf{w}$ may be inaccurate. For multiple eigenvalues the power method fails to converge.

SEQUENTIAL EXTRACTION OF CONVOLVED AND MIXED SOURCES

The criteria and algorithms discussed in the previous sections for blind signal extraction from instantaneous mixture can be extended or generalized to the problem of extraction of convolved (filtered) and mixed independent sources. In this section, we illustrate this by a simple extension of the linear predictor and BPF approach.

In multichannel blind deconvolution, an m dimensional vector of received signals $\mathbf{x}(k) \in \mathbb{R}^m$ is assumed to be generated from an n dimensional vector of spatially independent, temporally unknown source signals $\mathbf{s}(k) = [s_1(k), s_2(k), \ldots, s_n(k)]^T$ using the multi-variate linear time invariant (LTI) filters, i.e.,

$$\mathbf{x}(k) = \sum_{p=-\infty}^{\infty} \mathbf{H}_p \, \mathbf{s}(k-p) = [\mathbf{H}(z)] \, \mathbf{s}(k) \tag{16}$$

or equivalently in the scalar form

$$x_i(k) = \sum_{p=-\infty}^{\infty} \sum_{j=1}^{n} h_{ij,p} \, s_j(k-p), \quad (i = 1, \ldots, m) \tag{17}$$

where $\mathbf{H}(z) = \sum_{p=-\infty}^{\infty} \mathbf{H}_p z^{-p}$ is an unknown $(m \times n)$ polynomial matrix with $m \geq n$, and z^{-p} is the delay operator such that $z^{-p}\mathbf{s}(k) = \mathbf{s}(k-p)$. A task of multichannel deconvolution is to recover the source signals $\mathbf{s}(k)$ from the received signals $\mathbf{x}(k)$, up to a scaled, permuted, and delayed version of the source signals, i.e., to estimate sources, $\mathbf{y}(k) = \hat{\mathbf{s}}(k) = \mathbf{P} \, \mathbf{\Lambda} \, \mathbf{D}(z) \, \mathbf{s}(k)$, where $\mathbf{P} \in \mathbb{R}^{n \times n}$ is a permutation matrix, $\mathbf{\Lambda} \in \mathbb{R}^{n \times n}$ is a nonsingular scaling diagonal matrix, and $\mathbf{D}(z)$ is a diagonal matrix whose ith diagonal element is given by z^{-d_i}. In some application the condition are further relaxed in the sense that extracted (output) signals is filtered version of single source [2].

Let us consider the model similar to that shown in Fig.1 and Fig.2 in which synaptic weights w_{ji} represent convolutive FIR filters. Such processing unit with output $y_j(k)$ is described by

$$y_j(k) = \sum_{i=1}^{m} \sum_{p} w_{ji\,p} \, x_i(k-p), \tag{18}$$

where $\{w_{jip}\}$ are the coefficients of ji-th FIR filter and $\{x_i(k)\}$ is the ith sensor output.

In this model, we are exploiting the temporal structure of signals rather than the statistical independence [2, 16]. Intuitively speaking, the source signals s_j have less complexity than mixed sensor signals x_j. In other words, the degree of temporal predictability of any source signal is higher than (or equal to) that of any of the convolutive mixture under assumption that the convolutive filters are high-pass filters. For example, the waveforms of convolutive mixture of several sine waves with different frequencies are usually more complex or less predictable than either of the original sine waves. This means that applying the standard linear predictor and by minimizing the least mean error $E\{\varepsilon^2\}$, which is measure of predictability, we can separate or extract signals with different temporal structures. More precisely, by minimizing the error, we maximize a measure of temporal predictability for each recovered signal [2, 16].

It is interesting to note that there is some analogy between the measure of temporal predictability and the measure of non-Gaussianity. The central limit theorem ensures that the probability density function (pdf) of any mixture is closer to the Gaussian distribution than (or equal to) any of its component source signals. As some measure of non Gaussianity or statistical independence, we can use absolute value of the kurtosis and the generalized kurtosis. However, it should be noted, that these two measures: temporal linear predictability and non-Gaussianity based on kurtosis may lead to different results. Temporal predictability forces the extracted signal to be smooth and possibly to be of low complexity while the non-Gaussianity measure forces the extracted signals to be independent as possible and have sparse representation for sources with positive kurtosis.

Using the concept of linear predictor and adaptive bandpass filters (cf. Eqs. (7)-(11) and (14)- (15)) we can derive the following iterative algorithm for deconvolutive FIR filters:

$$\mathbf{w}_{ji}^{+}(l+1) = \widehat{\mathbf{R}}_{\mathbf{x}_i\mathbf{x}_i}^{-1} \left\langle \widetilde{y}_j(k)\bar{\mathbf{x}}_i(k) \right\rangle, \qquad (19)$$

$$\mathbf{w}_{ji}(l+1) = \frac{\mathbf{w}_{ji}^{+}(l+1)}{\|\mathbf{w}_{ji}^{+}(l+1)\|}, \qquad (20)$$

where

$$\widetilde{y}_j(k) = [B_j(z)]y_j(k), \qquad (21)$$

$\mathbf{w}_{ji} = [w_{ji0}, w_{ji1}, \ldots, w_{jiM}]^T$ and $\bar{\mathbf{x}}_i(k) = [x_i(k), x_i(k-1), \ldots, x_i(k-M)]^T$.

The advantage of the proposed approach is that, we can avoid the deflation procedure by using the bandpass filters with different frequency characteristics. Furthermore, the algorithm allows us to extract colored sources with different shape spectra.

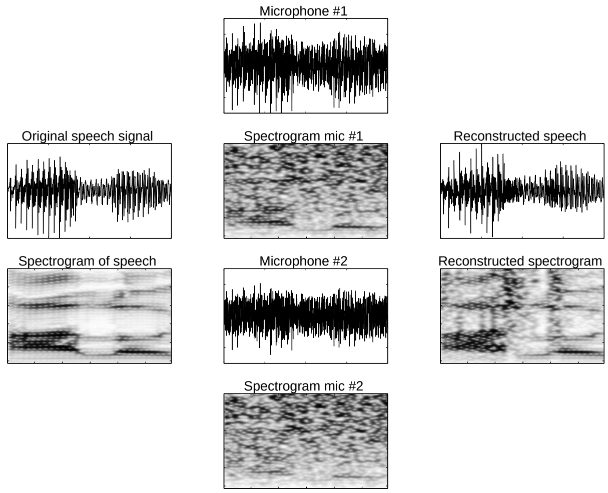

Figure 4: The example of sequential reconstruction of speech signal recorded using 2 microphones in noisy environment applying the model from the Fig.2 with 4th order IIR adjustable bandpass filter.

CONCLUSIONS

In summary, blind signal extraction is a useful approach when it is desired to extract several source signals with specific stochastic properties from a large number of mixtures. The proposed method does not need to apply deflation procedure, since processing unit can extract all desired narrow-band sources sequentially one-by-one by adjusting the center frequency and bandwidth of the bandpass filter. Parallel extraction of arbitrary group of sources is also possible by employing several bandpass filters with different suitably chosen frequency characteristics. The proposed algorithms are computationally very simple, efficient and robust with respect to an additive noise, both white and out of band colored noise. In contrast to other methods the covariance matrix of the noise does not need to be estimated or modeled. The approach can be extended to the problem of extraction of colored narrow band sources from convolutive mixture. On the other hand, the sequential blind extraction approach may give poorer performance in comparison to the simultaneous blind separation/deconvolution approach for ill-conditioned problems.

REFERENCES

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

[2] A. Cichocki and S. Amari. *Adaptive Blind Signal and Image Processing.* John Wiley, Chichester, 2002.

[3] A. K. Barros and A. Cichocki, "Extraction of specific signals with temporal structure", *Neural Computation*, 13(9):1995–2000, September 2001.

[4] A.K. Barros, H. Kawahara, A. Cichocki, S. Kojita, T. Rutkowski, M. Kawamoto, and N. Ohnishi, "Enhancement of a speech signal embedded in noisy environment using two microphones", in *Proceedings of the Second International Workshop on ICA and BSS, ICA'2000*, pages 423–428, Helsinki, Finland, 19-22 June 2000.

[5] T. Rutkowski, A. Cichocki and A. K. Barros, "Speech enhancement from interfering sounds using CASA techniques and blind source separation", 3rd International Conference on Independent Component Analysis and Signal Separation, San Diego, USA, December 9-13, 2001, pp. 728-733.

[6] 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)*.

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

[8] A. Cichocki and A. Beluochrani, "Sources separation of temporally correlated sources from noisy data using bank of band-pass filters", in *Third International Conference on Independent Component Analysis and Signal Separation (ICA-2001)*, pages 173–178, San Diego, USA, Dec. 9-13 2001.

[9] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis.* John Wiley, New York, 2001.

[10] S. Makeig, T.-P. Jung, A.J. Bell, D. Ghahremani, and T.J. Sejnowski, "Blind separation of auditory event-related brain responses into independent components", in *Proceedings of the National Academy of Sciences*, volume 94, pages 10979–10984, 1997.

[11] V. Calhoun, T. Adali, and G. Pearlson, "Independent component analysis applied to fMRI Data: a natural model and order selection", in Proc. IEEE Workshop on Nonlinear Signal and Image Processing, Baltimore, MD, June 2001.

[12] B. Karlsen, J. Larsen, H.B.D. Sorensen and K.B. Jokobsen, "Comparison of PCA and ICA based clutter reduction in GPR systems for anti-personal landmine detection", Proc. of 11th IEEE Workshop on Statistical Signal Processing, Singapore, Aug. 6-8, 2001.

[13] S.C. Douglas and S.Y. Kung, "Kuicnet algorithms for blind deconvolution", In *Proc. IEEE Workshop on Neural Networks for Signal Processing*, pages 3–12, Cambridge, UK, August 1998.

[14] S. Choi, H. Hong, H. Glotin, and F. Berthommier (2002), "Multichannel signal separation for cocktail party speech recognition: A dynamic recurrent network," accepted for publication in Neurocomputing, 2002 (to appear).

[15] D. P. Mandic and J.A. Chambers, *Recurrent Neural Networks for Prediction*, John Wiley, Chichester, 2001.

[16] J.V. Stone. Blind source separation using temporal predictability. *Neural Computation*, 13(7):1559–1574, 2001.

[17] 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.