Extraction of temporally correlated sources with its application to non-invasive fetal electrocardiogram extraction

Zhi-Lin Zhang, Zhang Yi

Computational Intelligence Laboratory, School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 610054, People's Republic of China.

Abstract

This letter addresses the problem of fetal electrocardiogram (FECG) extraction. A class of source extraction algorithms is proposed, which uses a combination of two different approaches: Barros's source extraction approach and Hyvärinen's fixed-point approach. The proposed algorithms can rapidly extract the desired FECG with little noise, and can work well in some bad situations. The validity and performance of the algorithms are confirmed by extensive computer simulations and experiments on real-world data.

Key words: Source extraction; Independent component analysis (ICA); Blind source separation(BSS); Fetal electrocardiogram (FECG);

1 Introduction

The extraction of fetal electrocardiogram (FECG) [1–3] is of vital importance from the clinical point of view, because it provides information about the health and the possible diseases of a fetus. Before delivery, non-invasive techniques to acquire the FECG are preferred. However, the desired fetal heartbeat signal appearing at the electrode output is always corrupted by a lot of noise, such as the maternal electrocardiogram (MECG) contributions with extremely high amplitude, the mother's respiration, the power line interference, and the thermal noise due to electronic equipment. Therefore obtaining FECG is a difficult task.

Email address: zlzhang@uestc.edu.cn, zhangyi@uestc.edu.cn (Zhi-Lin Zhang, Zhang Yi).

Many methods have been proposed to address this problem. A promising approach is blind source separation (BSS) or independent component analysis (ICA) [1–5]. However, separating all of the source signals from a large number of observed sensor signals takes a long time and is often not necessary. Thus a better choice may be the source extraction methods [4,6,7,9], which are closely related to the BSS.

However, many source extraction algorithms [4,7,9,10] are not suitable to extract the FECG, since the objective is to obtain the FECG as the first output signal. It seems that only the algorithm proposed by Barros and Cichocki [6] is good at extracting FECGs. The simple and fast algorithm requires *a priori* information about the desired FECG, namely, a suitable time delay at which the autocorrelation of the FECG has a large absolute value, while those of other source signals have very small value. However, in practice the extracted FECG by this algorithm is often mixed with some noise, such as the mother's breathing artifact. What is more, it is sensitive to the estimation error of the time delay. To overcome these drawbacks, we propose a class of algorithms, which combines the Barros's algorithm and Hyvärinen's fixed-point algorithm [10].

The rest of the letter is organized as follows. In section 2 we derive the Barros's algorithm [6] from a different perspective. Then we introduce the ICA model and the fast fixed-point algorithm [10] in Section 3. In Section 4 we propose a class of algorithms. Extensive simulations and experiments on real-world data are carried out in Section 5, and conclusions are drawn in Section 6.

2 Maximization of autocorrelation at a specific time delay

Assume that the desired source signal s_i (in the case of the FECG extraction, s_i is the desired FECG) is temporally correlated, satisfying the following relations for a specific time delay τ^*:

$$\begin{cases} E\left\{s_i(k)s_i(k-\tau^*)\right\} \neq 0 \\ E\left\{s_i(k)s_j(k-\tau^*)\right\} = 0 \\ E\left\{s_j(k)s_j(k-\tau^*)\right\} = 0 \quad \forall j \neq i, \end{cases} \tag{1}$$

where s_j are other source signals, k is the time index and τ^* is an integer delay. Ideally, under the condition $\|\mathbf{w}\| = 1$, maximizing the objective function

$$J(\mathbf{w}) = E\{y(k)y(k-\tau^*)\} \tag{2}$$

leads to the desired source signal. Here, $y(k) = \mathbf{w}^T\mathbf{x}(k)$, and $\mathbf{x}(k)$ are the prewhitened sensor signals. When $J(\mathbf{w})$ reaches its maximum, $y(k)$ is the desired signal. The reason for this proposal is that for the desired source signal, this autocorrelation should have a large value, while for other source signals this value should be very small.

By the standard gradient approach, we can derive the following algorithm: [1]

$$\begin{cases} \mathbf{w}^+ = (R_x(\tau^*) + R_x(\tau^*)^T)\mathbf{w} \\ \mathbf{w} = \mathbf{w}^+/\|\mathbf{w}^+\| \end{cases} \tag{3}$$

where $R_x(\tau^*) = E\{\mathbf{x}(k)\mathbf{x}(k-\tau^*)^T\}$. Note that the algorithm can be considered an extension of the Barros's algorithm [6], preserving all of the merits of the latter. Most importantly, it is relatively insensitive to the estimation error of the time delay τ^*, in contrast to the Barros's algorithm. However, in most cases the assumption (1) is not strictly satisfied. In other words, although the desired source signal s_i is strongly autocorrelated at the time delay τ^*, it also weakly correlates with some of the other source signals. What is more, some other source signals may be autocorrelated at the time delay τ^*, too. Therefore the extracted signal is often mixed with some other source signals. Later we will find ways to overcome these drawbacks.

3 The one-unit fixed-point algorithm

The basic ICA model can be summarized as follows: assume that there exist mutually independent unknown sources s_j $(j = 1, \cdots, N)$, and that each has zero mean and unit variance. The sources are linearly mixed with an unknown $M \times N (M \geq N)$ matrix $\mathbf{A}$:

$$\mathbf{x} = \mathbf{A}\mathbf{s}, \tag{4}$$

where $\mathbf{s} = [s_1, s_2, \cdots, s_N]^T$ and $\mathbf{x} = [x_1, x_2, \cdots, x_M]^T$ are respectively N-dimensional sources and M-dimensional mixed signals. The basic goal is to find an $M \times N$ separating matrix $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_N]$ without knowing the mixing matrix $\mathbf{A}$, such that

$$\mathbf{y} = [y_1, y_2, \cdots, y_N]^T = \mathbf{W}^T\mathbf{x} \tag{5}$$

[1] we will drop the index k from now on, and use it only when necessary to avoid confusion.

is the estimate of $\mathbf{s}$ to within the well-known permutation and scaling ambiguities [4,9].

To find one independent source signal $y = \mathbf{w}^T\mathbf{x}$ ($\mathbf{w}$ is any column of $\mathbf{W}$, and $\mathbf{x}$ are the prewhitened mixed signals), Hyvärinen and Oja [10] propose to maximize the standard kurtosis

$$J(\mathbf{w}, y) = \kappa_4(y(\mathbf{w})) = E\{(\mathbf{w}^T\mathbf{x})^4\} - 3[E\{(\mathbf{w}^T\mathbf{x})^2\}]^2, \tag{6}$$

under the constraint $\|\mathbf{w}\| = 1$. Based on (6), they obtain the following one-unit fixed-point algorithm:

$$\begin{cases} \mathbf{w}^+ = E\{\mathbf{x}(\mathbf{w}^T\mathbf{x})^3\} - 3\mathbf{w} \\ \mathbf{w} = \mathbf{w}^+/\|\mathbf{w}^+\|. \end{cases} \tag{7}$$

Applying a deflation technique [10,11] the above algorithm can extract the subsequent source signals. It has many desirable properties. For example, its convergence is cubic.

In a similar way, Cichocki and Amari [4] derive a modified fixed-point algorithm for the generalized normalized kurtosis

$$\kappa_{p,q}(y(\mathbf{w})) = \frac{1}{p}\left(\frac{E\{|y|^p\}}{E^q\{|y|^{p/q}\}} - c_{pq}\right), \tag{8}$$

where c_{pq} is a positive constant such that for the Gaussian distribution $\kappa_{p,q} = 0$. In the special case, for $p = 4, q = 2, c_{pq} = 3$, applying the standard gradient method to (8), one can obtain the following learning rule

$$\begin{cases} \mathbf{w}^+ = \frac{E\{y^3\mathbf{X}\}}{E\{y^4\}} - \mathbf{w} \\ \mathbf{w} = \mathbf{w}^+/\|\mathbf{w}^+\|. \end{cases} \tag{9}$$

The above algorithm is more robust to outliers and spiky noise than the original fixed-point algorithm (7). And it can also be extended to the muli-unit algorithm using the deflation method or other approaches [4,11].

By the algorithm (7) or (9), however, any source signal could be extracted as the first one. So one cannot ensure that the first extracted signal is the desired one. To solve this problem, in the next section we will propose a class of algorithms, which combines the extension of the Barros's algorithm (3) and the fixed-point algorithm (7) or (9).

4 The proposed algorithms

In order to extract the desired source signal, we propose to roughly divide the extraction procedure into two stages. The first stage is a capture stage, in which the desired source signal is coarsely extracted by the algorithm (3). But the extracted signal is often mixed with noise (i.e., some other source signals), just as we have mentioned. In the second stage, the noisy extracted signal is processed as clearly as possible by the one-unit fixed-point algorithm (7) or (9), based on the assumption that the desired source signal is independent of the other source signals. Thus this stage, to some extent, can be regarded as a fine extraction stage.

Following the idea above we propose to maximize the following objective

$$J_1(\mathbf{w}) = \frac{1}{4}\bigg(E\{y^4\} - 3[E\{y^2\}]^2\bigg) + \lambda E\{y(k)y(k-\tau^*)\} + F(\|\mathbf{w}\|^2), \quad (10)$$

which combines the objective function (2) and (6), or, to maximize

$$J_2(\mathbf{w}) = \frac{1}{4}\bigg(\frac{E\{|y|^4\}}{E^2\{|y|^2\}} - 3\bigg) + \lambda E\{y(k)y(k-\tau^*)\} + F(\|\mathbf{w}\|^2), \quad (11)$$

which combines the objective (2) and (8). λ is a nonnegative parameter whose value is sufficiently large at the capture stage and sufficiently small at the fine extraction stage. $E\{y(k)y(k-\tau^*)\}$ is the autocorrelation (lies in the range $[-1, 1]$) at the time delay τ^*. $F(\cdot)$ is a penalty term due to the constraint $\|\mathbf{w}\| = 1$.

From (10) and (11), we have

$$\frac{\partial J_1(\mathbf{w})}{\partial \mathbf{w}} = E\{\mathbf{x}(\mathbf{w}^T\mathbf{x})^3\} - 3\mathbf{w} + f(\|\mathbf{w}\|^2)\mathbf{w} + \lambda(R_x(\tau^*) + R_x(\tau^*)^T)\mathbf{w}, \quad (12)$$

$$\frac{\partial J_2(\mathbf{w})}{\partial \mathbf{w}} = \frac{E\{y^3\mathbf{x}\}}{E\{y^4\}} - \mathbf{w} + f(\|\mathbf{w}\|^2)\mathbf{w} + \lambda(R_x(\tau^*) + R_x(\tau^*)^T)\mathbf{w}, \quad (13)$$

where $f(\|\mathbf{w}\|^2)\mathbf{w}$ is the gradient of $F(\|\mathbf{w}\|^2)$. Note that as long as $F(\|\mathbf{w}\|^2)$ is a function of $\|\mathbf{w}\|^2$ only, its gradient has the form $scalar \times \mathbf{w}$. Similar to [10], the fixed point $\mathbf{w}$ of the learning rule (10) is obtained by equating the change in $\mathbf{w}$ to 0:

$$\begin{cases} \mathbf{w}(k+1)^+ = E\{\mathbf{x}(\mathbf{w}(k)^T\mathbf{x})^3\} - 3\mathbf{w}(k) + \lambda(R_x(\tau^*) + R_x(\tau^*)^T)\mathbf{w}(k) \\ \mathbf{w}(k+1) = \mathbf{w}(k+1)^+/\|\mathbf{w}(k+1)^+\|. \end{cases} \quad (14)$$

Similarly, from (13) we obtain another learning rule as follows

$$\begin{cases} \mathbf{w}(k+1)^{+} = \frac{E\{y^3\mathbf{X}\}}{E\{y^4\}} - \mathbf{w}(k) + \lambda(R_x(\tau^*) + R_x(\tau^*)^T)\mathbf{w}(k) \\ \mathbf{w}(k+1) = \mathbf{w}(k+1)^{+}/\|\mathbf{w}(k+1)^{+}\|, \end{cases} \tag{15}$$

which is expected to be robust to outliers. With regard to the value of λ, we adopt the following form (but other strategies that follow the two-stage idea are also feasible):

$$\lambda(k) = \frac{\lambda_0}{k+1}, \tag{16}$$

where λ_0 is a positive constant with a large value. In practice, we find $\lambda_0 = 50 \sim 200$ is good for successful extraction of the FECG.

On the other hand, for estimating the time delay τ^*, there are several methods, such as computing the autocorrelation of a sensor signal [6], and the heart instantaneous frequency estimation technique [8]. In many applications the task of estimating τ^* is not difficult, but the estimation error is often inevitable. For many extraction algorithms based on time delays, such as the Barros's algorithm [6], their performance is greatly affected by the estimation error. We will see, however, that the new algorithms (14) and (15) are robust to the error as long as it is not too large.

5 Simulations and experiments on real-world data

To check the validity and good performance of the proposed algorithms, we have performed extensive computer simulations and experiments on real-world data. But due to limit of space, we only present several typical results.

In the first simulation, we generated five zero-mean and unit-variance source signals, shown in Fig.1. Each signal had 2500 samples. They were, respectively, a muscle artifact, a breathing artifact, a Gaussian signal, an MECG, and an FECG whose period was 112 sampling period (i.e., $\tau^* = 112$). These source signals were randomly mixed. After prewhitening the mixed signals, we ran the algorithms (14), (15) and the Barros's extraction algorithm [6] (we called it, for simplicity, autocorrelation maximization algorithm (AM algorithm)). To compare the extraction performance, we adopted the following measure:

$$PI = -10E\{lg(s(k) - \tilde{s}(k))^2\}, \quad (dB) \tag{17}$$

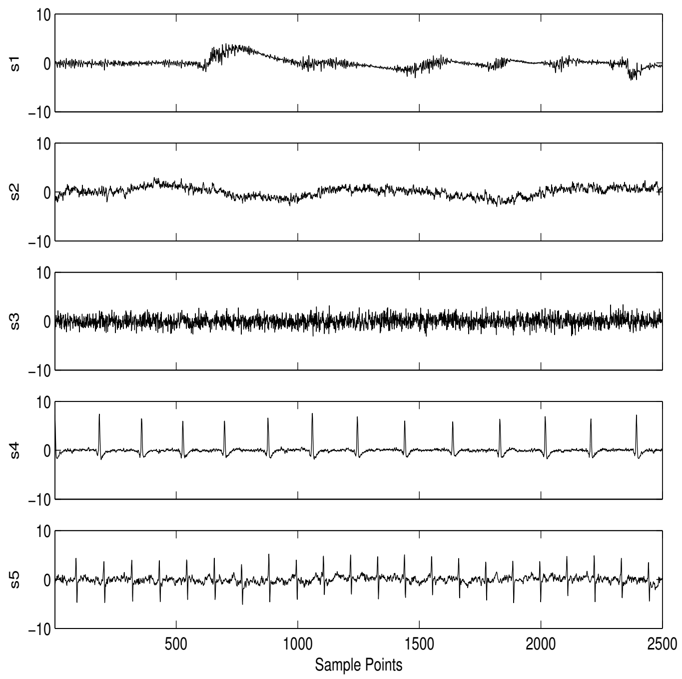

Fig. 1. Five source signals. From the top down, they were respectively a muscle artifact, a breathing artifact, a Gaussian signal, an MECG, and an FECG.

Table 1

The extraction performance indices averaged over 1000 independent trials, measured by the equation (17).

	Simulation One	**Simulation Two**
Algorithm (14)	22.2 (dB)	17.4 (dB)
Algorithm (15)	20.1 (dB)	23.2 (dB)
AM algorithm [6]	7.0 (dB)	7.3 (dB)

where $s(k)$ was the desired source signal (i.e., the FECG), and $\tilde{s}(k)$ was the extracted signal (both of them were normalized to be zero-mean and unit-variance). The higher PI was, the better the performance was. The simulation was independently repeated 1000 times, and the averaged performance indices are shown in the second column of Table 1. It is clear to see that the AM algorithm had poorest performance, due to the fact that the FECG extracted by it was always mixed with some other source signals.

To investigate the robustness of the algorithm (15), we randomly added 25 outliers whose values were 10 in each source signal. Then the source signals were randomly mixed, followed by prewhitening. Since the robust estimation of the covariance matrix was a classic problem independent of the robustness of the algorithms, we used in this simulation a hypothetical robust estimator of covariance, which was simulated by estimating the covariance matrix from the original source signals without outliers. Averaged extraction performance

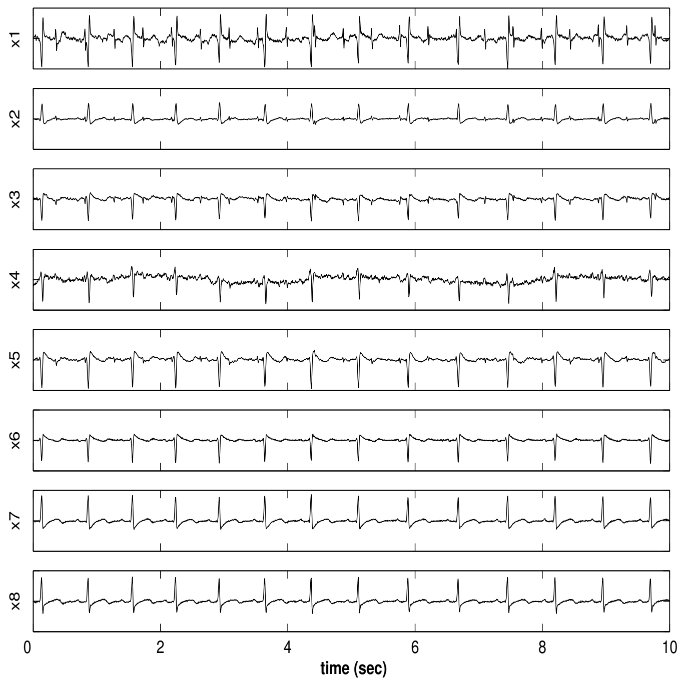

Fig. 2. ECG data measured from a pregnant woman.

indices over 1000 independent trials are given in the third column of Table 1. Obviously, the algorithm (15) outperformed the other two algorithms, as expected, and again the AM algorithm performed poorly.

Next we used real-world ECG data to test our algorithms. The ECG data set used in this experiment was distributed by De Moor [12], which was measured from a pregnant woman (Fig.2). Sampling frequency was 250 Hz (Although in De Moor's homepage he claimed the sampling frequency was 500 Hz, Barros et al. assured it was 250 Hz [6]). By carefully examining the autocorrelation of the sensor signal x_1, and using *a priori* knowledge that the fetal heart should strike every 0.5 second or so, we found that a peak lay at $\tau^* = 112$ sampling period (i.e., 0.448 second) and believed it was just the time delay for extracting the FECG. After prewhitening the sensor signals, we ran our algorithms (14), (15) and the AM algorithm. Results are shown in Fig.3, which illustrates that our algorithms greatly reduced the mother's breathing artifact and obtained clearer FECGs, compared with the AM algorithm.

To check that our algorithms were not sensitive to the estimation error of the time delay τ^*, we carried out another experiment. We intentionally set the time delay $\tau^* = 114$ (i.e., 0.456 second) for the three algorithms (Remember the true time delay was $\tau^* = 112$). Results are shown in Fig.4. One can deduce that the estimation error of the time delay has little effect on the performance of our algorithms, but badly affects the AM algorithm.

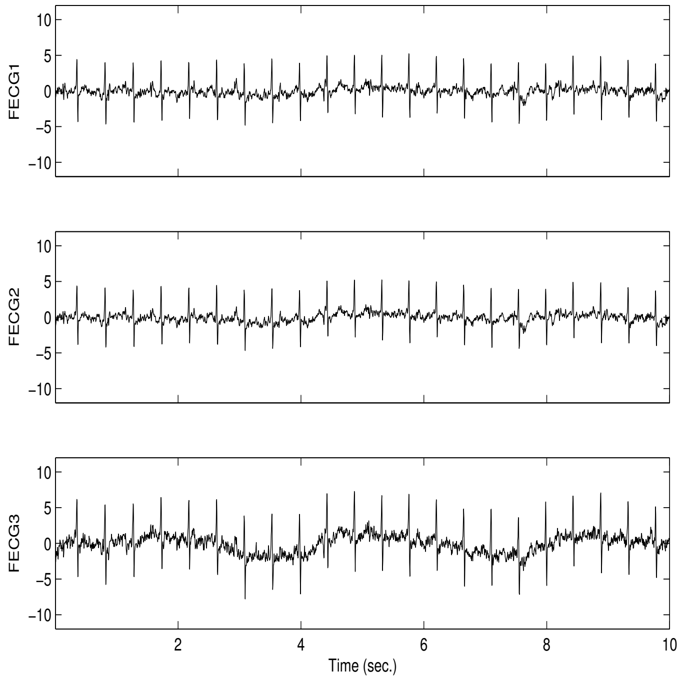

Fig. 3. Comparison of the extracted FECGs when there was no estimation error of the time delay. The top two FECGs were extracted by our algorithm (14) and (15), respectively. The bottom one was extracted by the AM algorithm.

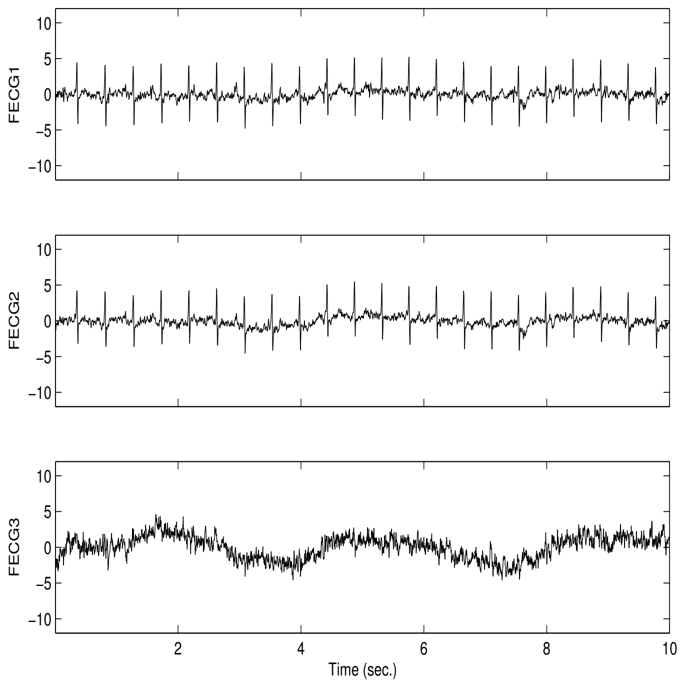

Fig. 4. Comparison of the extracted FECGs when the estimation error of the time delay was introduced. The top two FECGs were extracted by our algorithm (14) and (15), respectively. The bottom one was extracted by the AM algorithm.

6 Conclusions

This letter presents two algorithms to deal with the FECG extraction problem. They are based on the joint maximization of the (generalized) kurtosis and autocorrelation (at a specific time delay) of the output signal. The extraction procedure can be roughly divided into two stages. At the first stage, the autocorrelation property of source signals is used to extract the desired source signal. But in this stage the extracted one is often mixed with some noise. At the second stage, the signal is processed as clearly as possible, based on the fact that the desired source signal is independent of the noise.

The proposed algorithms have many advantages. First, their convergence is fast (only 5-10 iterations are needed to achieve convergence). Second, the two algorithms have high extraction performance. Furthermore, one of the proposed algorithms is robust to the outliers and spiky noise. In addition, the two algorithms are insensitive to the estimation error of the time delay as long as the error is not too large. The effectiveness and good performance of the proposed algorithms are confirmed by extensive simulations and experiments on real-world ECG data.

Acknowledgments

We are indebted to Prof. Fusheng Yang of Tsinghua University for his help and discussions. We also thank Prof. Robert W. Newcomb for his help in revising the manuscript. And we are grateful to all the anonymous reviewers who provided helpful comments. This work was supported by the National Science Foundation of China under Grant No. 60471055.

References

[1] V. Zarzoso, A.K. Nandi, Noninvasive fetal electrocardiogram extraction blind separation versus adaptive noise cancellation, IEEE Trans. On Biomedical Engineering 48 (1) (January, 2001) 12-18

[2] M.G. Jafari, J.A. Chambers, Fetal electrocardiogram extraction by sequential source separation in the wavelet domain, IEEE Trans. On Biomedical Engineering 52 (3) (Mar, 2005) 390-400

[3] A.K. Barros, et al., Removing artifacts from electrocardiographic signals using independent component analysis, Neurocomputing 22 (1-3) (November, 1998) 173-186

[4] A. Cichocki, S. Amari, Adaptive blind signal and image processing: learning algorithms and applications, New York: John Wiley & Sons, 2002

[5] M.G. Jafari, J.A. Chambers, Adaptive noise cancellation and blind source separation, 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003), Nara, Japan, April, 2003, pp.627-632

[6] A.K. Barros, A. Cichocki, Extraction of specific signals with temporal structure, Neural Computation 13 (9) (September, 2001) 1995-2003

[7] A. Cichocki, R. Thawonmas, S. Amari, Sequential blind signal extraction in order specified by stochastic properties, Electronics Letters 33 (1) (January, 1997) 64-65

[8] A.K. Barros, N. Ohnishi, Heart instantaneous frequency (HIF): an alternative approach to extract heart rate variability, IEEE Trans. On Biomedical Engineering 48 (8) (August, 2001): 850-855

[9] S.A. Cruces-Alvarez, A. Cichocki, S. Amari, From blind signal extraction to blind instantaneous signal separation: criteria, algorithm, and stability, IEEE Trans. On Neural Networks 15 (4) (July, 2004) 859-873

[10] A. Hyvärinen, E. Oja, A fast fixed-point algorithm for independent component analysis, Neural Computation 9 (7) (October, 1997) 1483-1492

[11] N. Delfosse, P. Loubaton, Adaptive blind separation of independent sources: A deflation approach, Signal Processing 45 (1) (July, 1995) 59-83

[12] D. De Moor (Ed.): Daisy: Database for the identification of systems. Available online at: http://www.esat.kuleuven.ac.be/sista/daisy