Le CNRS | Accueil INIST

Titre du document / Document title

Blind source separation using time-delayed signals

Auteur(s) / Author(s)

TOME A. M. (1) ; TEIXEIRA Ana R. (1) ; LANG E. W. (2) ; STADLTHANNER K. (2) ; ROCHA A. P. (3) ; ALMEIDA R. (3)

Affiliation(s) du ou des auteurs / Author(s) Affiliation(s)

(1) DETI/IEETA Universidade de Aveiro, 3810 Aveiro, PORTUGAL

(2) Institute of Biophysics University of Regensburg, 93040 Regensburg, ALLEMAGNE

(3) DMA/FCUP Universidade do Porto, Porto, PORTUGAL

Résumé / Abstract

In this work a modified version of AMUSE, called dAMUSE, is proposed. The main modification consists in increasing the dimension of the data vectors by joining delayed versions of the observed mixed signals. With the new data a matrix pencil is computed and its generalized eigendecomposition is performed as in AMUSE. We will show that in this case the output (or independent) signals are filtered versions of the source signals. Some numerical simulations using artificially mixed signals as well as biological data (RR and QT intervals of Electrocardiogram) are presented.

Revue / Source

Congrès

2004 International Joint Conference on Neural Networks : (proceedings) (Budapest, Hungary, 25-29 July, 2004)

International Joint Conference on Neural Networks, Budapest , HONGRIE (25/07/2004)

2004, vol. 4 XLVII-3302 p.,] (16 ref.), [Notes: "IEEE Catalog Number: 04CH37541"--T.p. verso] ISBN 0-7803-8359-1 ; Illustration : Illustration ;

Langue / Language

Anglais

Editeur / Publisher

IEEE, Piscataway NJ, ETATS-UNIS (2004) (Monographie)

Mots-clés anglais / English Keywords

Filter ; Matrix pencil ; Delay time ; Electrocardiography ; Signal sources ; Source separation ; Blind separation ; Delay ; Neural network ;

Mots-clés français / French Keywords

Filtre ; Pinceau matriciel ; Temps retard ; Electrocardiographie ; Source signal ; Séparation source ; Séparation aveugle ; Retard ; Réseau neuronal ;

Mots-clés espagnols / Spanish Keywords

Filtro ; Pincel matricial ; Tiempo retardo ; Electrocardiografía ; Separación señal ; Separación ciega ; Retraso ; Red neuronal ;

Localisation / Location

2

Blind Source Separation Using Time-delayed Signals

A.M.Tomé, Ana R. Teixeira
DETUA/IEETA
Universidade de Aveiro
P-3810 Aveiro, Portugal
email: ana@ieeta.pt

E.W.Lang, K. Stadlthanner
Institute of Biophysics
University of Regensburg
D-93040 Regensburg, Germany
email:elmar.lang@biologie.uni-regensburg.de

A.P. Rocha, R.Almeida
DMA/FCUP
Universidade do Porto
Porto, Portugal
email: aprocha@fc.up.pt

Abstract—**In this work a modified version of AMUSE is proposed. The main modification consist in increasing the vector data dimension by joining delayed versions of the observed mixed signals. With the new data a matrix pencil is computed and its generalized eigendecomposition is performed as in AMUSE. We will show that in this case the output (or independent) signals are filtered versions of the source signals. Some numerical simulations using artificially mixed signals are presented as well as some preliminary results of a Heart Rate Variability (HRV) study showing that the output signals are related with LF (low frequency) and HF (high frequency) fluctuations.**

I. Introduction

Blind Source Separation methods based on the generalized eigendecomposition (GED) of a matrix pencil estimated using second order statistics are sensitive to noise [1][2]. There are several proposals to improve the efficiency of those algorithm when noise is present[1][3]. Besides there are denoising techniques whose first step consist in increasing the dimension of the data set by joining delayed versions of the signals [4][5]. Then, the de-noising is achieved by the projection of the data into a lower dimension subspace. In this work we will present a modified version of AMUSE [6] which joins both strategies. AMUSE is a second-order statistics whose solution is based on the simultaneous diagonalization of matrix pencil (a correlation matrix and a time-delayed matrix). The proposed algorithm also comprises the simultaneous diagonalization but the pencil is computed after increasing the data set dimension by joining delayed versions of each mixed signal.

In the following section we will show, considering a noise free model of linearly mixed signals, that the computed independent signals are filtered versions of source (unmixed) signals. We will also present an algorithm to compute the eigenvector matrix of the matrix pencil which involves a two step procedure based on the standard eigendecomposition approach. The advantage of this procedure is concerned with a dimensional reduction between the two steps as well as a reduction in the number of independent signals. Then, some simulations with artificially mixed signals are discussed to illustrate the formulated method. The proposed algorithm is also applied to the study of Heart Variability Rate using (RR,QT) sequences of the ECG signal. These preliminary results shows that the independent signals are related to LF (low frequency) and HF (High Frequency) fluctuations[7].

II. Generalized Eigendecomposition Using Time-delayed Signals

Considering a group of N sensor signals $x^{(n)} = \begin{bmatrix} x_1, & x_2, & \cdots & x_N \end{bmatrix}^T$ where each component, $i = 1 \ldots N$, includes M delayed versions of the sensor signal

$$x_i = \begin{bmatrix} x_i[n], & x_i[n-1], & \cdots & x_i[n-M-1] \end{bmatrix}$$

The data vector $x^{(n)}$ is used to compute a pair of time-delayed correlation matrices

$$R_x(k_i) = E\left\{x^{(n)} x^{(n-k_i)^T}\right\} \quad (1)$$

The generalized eigendecomposition of the pair $(R_x(k_1), R_x(k_2))$, with size $NM \times NM$, in particular its eigenvector matrix E_x, will be used to transform the NM dimensional sensor signals $x^{(n)}$, $i.e$, $y = E_x^T x^{(n)}$. The signals y are filtered versions of the the source signals, s, as we will prove next.

Assuming that each mixed signal is a linear combination of N source signals, the vector of the source signals $s^{(n)} = \begin{bmatrix} s_1, & s_2, & \ldots & s_N \end{bmatrix}^T$ has components s_i, $i = 1 \ldots N$,with M delayed versions of each source signal,

$$s_i = \begin{bmatrix} s_i[n], & s_i[n-1], & \cdots & s_i[n-M-1] \end{bmatrix}.$$

Writing the vector of mixed signals $x^{(n)} = Hs^{(n)}$, the mixing matrix H will be a block matrix with a diagonal matrix in each block

$$H = \begin{bmatrix} h_{11}I_{M\times M} & h_{12}I_{M\times M} & .. & h_{1N}I_{M\times M} \\ h_{21}I_{M\times M} & h_{22}I_{M\times M} & & \\ & & & \\ h_{N1}I_{M\times M} & & & h_{NN}I_{M\times M} \end{bmatrix} \quad (2)$$

where I_{MxM} is the identity matrix and h_{ij} is the mixing coefficient that relates the sensor signal i with the source signal j. As we are dealing with instantaneous mixing model all delayed versions of a signal are related by the same coefficient.

Using the mixing matrix equation $x^{(n)} = Hs^{(n)}$, we obtain

$$R_x(k_i) = HE\left\{s^{(n)} s^{(n-k_i)^T}\right\} H^T = HR_s(k_i)H^T \quad (3)$$

The equation (3) relates the time delayed correlation matrices of the mixed, $R_x(k_i)$, with the identical matrices computed using source signals $R_s(k_i)$. The two pair of matrices

$(R_x(k_1), R_x(k_2)$ and $(R_s(k_1), R_s(k_2)$ are congruent pencils [8] and so:

- the eigenvalues are the same,i.e., the eigenvalue matrices of both pencils are $D_x = D_s$
- And consequently having unique eigenvalues (distinct values in the diagonal of the matrix $D_x = D_s$), the eigenvectors are also related by $E_s = H^T E_x$

Assuming that all sources are uncorrelated the matrices $R_s(k_i)$ are block diagonal having a matrix $R_{mm}(k_i) = E\left\{s_m^{(n)} s_m^{(n-k_i)}\right\}$ in diagonal

$$R_s(k_i) = \begin{bmatrix} R_{11}(k_i) & 0 & \cdots & 0 \\ 0 & R_{22}(k_i) & \cdots & \cdots \\ 0 & 0 & \ddots & 0 \\ 0 & \cdots & 0 & R_{NN}(k_i) \end{bmatrix}$$

The eigenvector matrix of the generalized eigendecomposition of the pencil $(R_s(k_1), R_s(k_2))$ can be written as

$$E_s = \begin{bmatrix} E_{11} & 0 & \cdots & 0 \\ 0 & E_{22} & \cdots & \cdots \\ 0 & 0 & \ddots & 0 \\ 0 & \cdots & 0 & E_{NN} \end{bmatrix} \tag{4}$$

where E_{mm} is the $M \times M$ eigenvector matrix of the generalized eigendecomposition of the pencil $(R_{mm}(k_1), R_{mm}(k_2))$.

The mixed signals are linearly transformed by E_x^T,i.e.,

$$y = E_x^T H s^{(n)} = E_s^T s^{(n)}$$

The eigenvector matrix E_s (eqn. 4) is a block diagonal matrix, then there are M signals in the vector y which are a linear combination of one of the source signal and its delayed versions. For instance, the block m depends on the source signal m

$$\sum_{k=1}^{M} E_{mm}(k,j) s_m(n-k+1) \tag{5}$$

The equation (5) defines the convolution operation between column j of E_{mm} and source signal s_m.Then the columns of the matrix E_{mm} are the impulse response of finite impulse response (FIR) filters. Considering that all the columns of E_{mm} are different its frequency response might provide different shaping of the source signal spectra. Then, in a total of NM output signals y there are M filtered versions of each source/unmixed signal.

III. Algorithm Implementation

There are several ways to compute the eigenvalues or the eigenvectors of a matrix pencil if one of the matrices is symmetric positive[9]. One of those methods is based on the standard eigendecomposition applied in two consecutive steps. If one of the matrices of the pair is computed for delay zero we might achieve the symmetric positive definite condition. So, considering the pencil $(R_x(k_2), R_x(0))$ the following steps are proposed:

- Compute standard eigendecomposition of $R_x(0), i.e$, the eigenvectors (ν_i) and eigenvalues (λ_i). As the matrix is symmetric positive the eigenvalues can be organized descending order $(\lambda_1 > \lambda_2 > ... > \lambda_{NM})$. In AMUSE (and other algorithms) this procedure is used to estimate the number of sources and it is also considered a strategy to reduce the noise. In this work, we consider a variance criterium so that the most significant eigenvalues can be chosen according to
$$\frac{\lambda_1 + \lambda_2 + ... + \lambda_l}{\lambda_1 + \lambda_2 + ... \lambda_{NM}} \geq TH \tag{6}$$
If we are interested in the eigenvectors corresponding to the directions of higher variance of the signals, the threshold (TH) should be chosen so that the maximum energy of the signals is preserved.
- The transformation matrix can be computed using the l most significant eigenvalues or all the eigenvalues and respective eigenvectors. If the Cholesky decomposition is chosen then,
$$Q = \wedge^{-\frac{1}{2}} \vee^T \tag{7}$$
Then Q is a $l \times NM$ matrix if a reduction in dimension is considered.
- Compute the matrix $C = QR_x(k_1)Q^T$ and its the standard eigendecomposition: the eigenvectors (U) and eigenvalues (D_x).

The eigenvectors (no normalized to unit length) of the pencil $(R_x(k_2), R_x(0))$ are

$$E_x = V \wedge^{-\frac{1}{2}} U \tag{8}$$

Note that if $TH = 1$, the same result is obtained with $eig(R_x(k_2), R_x(0))$ in the recent version of MATLAB.
In what concerns the independent components, the mixed signals and its delayed versions should be transformed and l (or NM) signals are obtained,

$$y^{(n)} = E_x^T x^{(n)} = U^T \wedge^{-\frac{1}{2}} \vee^T x^{(n)} \tag{9}$$

IV. Results

The proposed algorithm was applied to artificially mixed signals and biological signals (RR and QT sequences of the Electrocardiogram-ECG). The artificial mixtures allows the illustration of the method once every parameter of model is known.

The heart rate variability study (HRV), based on RR and QT sequences, illustrate a possible application to real data. The Heart Rate Variability (HRV) spectrum is currently separated into three frequency bands: Very Low Frequency band (VLF range below 0.04Hz), Low Frequency band (LF range 0.04-0.15HZ) and high-frequency band (HF range 0.15- 0.4HZ). Hence it is interesting to known if the extracted signals can be related with LF and HF fluctuations of cardiovascular signals.

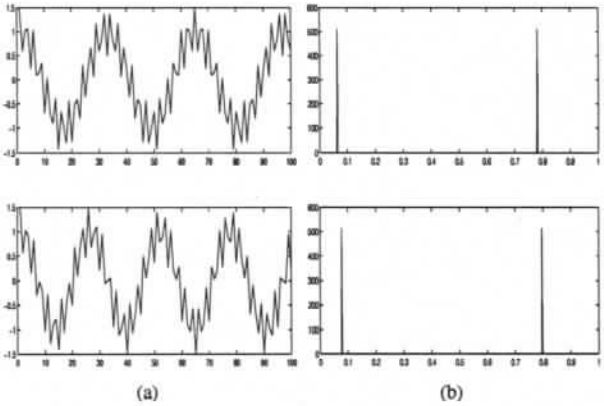

Fig. 1. Artificial Signals: (a)-source signals; (b)-Spectral contents of the source signals

A. Artificially mixed signals

The simulations were designed to illustrate the method and to study the influence of its parameters (specially M and TH) on the respective performance. As the separation process involves also linear filtering operation, each computed independent component have its maximum correlation with one of the source signal for a non-zero delay, besides the usual indetermination in order and amplitude.

In the first experience, two source signals with very similar frequency contents were considered (see figure 1). The signals are randomly mixed and the algorithm was applied for different values of $M, TH = 1, k_1 = 0$ and $k_2 = 1$. The figure (2) shows the results achieved when $M = 3$. In that case six independent components are computed: three (figure 2-a) are correlated with the first source signal while the other three (figure 2-b) were correlated with second source signal. We can see that those components have different frequency contents of each of the source signals, for instance figure 2-a) shows that: the first component has the high frequency, the last has the low frequency and the middle has the frequency contents of the source. The figure 2-(c) and (d) represents the frequency response magnitude of the filters considering the columns of the $E^T H$ as coefficients of a filter. The filters are then paired with the respective independent source signals and we can verify that the frequency response confirms the frequency contents of each of the independent signals as we have a low-pass, a band-pass and high-pass filter. Increasing the number of delays (M) the independent components show either the low or high frequency contents of the source signals. In this case, the frequency response of the columns of $E^T H$ will not have a pass-band bandwidth to include all frequency contents of each source signal. Another problem is that the eigenvalues of the first standard eigendecomposition start to have very low values. In that case it is better to consider $TH < 1$ and reduce (from NM to l) the number of signals at the output of the first step.

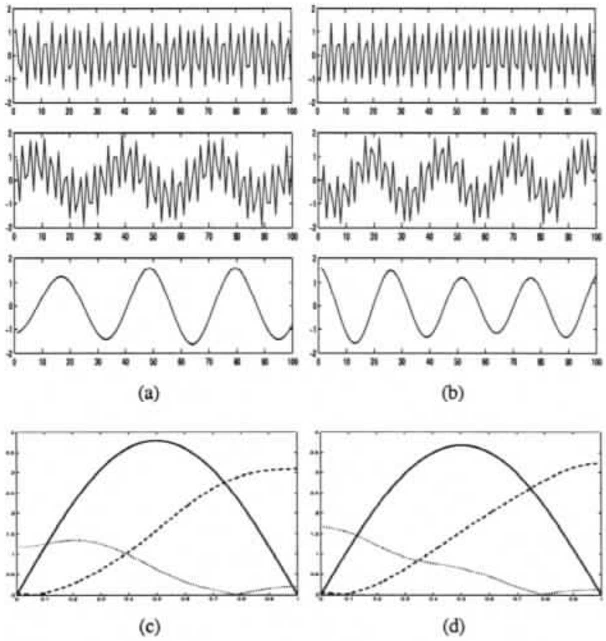

Fig. 2. (a)-Independent Components correlated with source signal 1 , (b)-Independent Components correlated with source signal 2 ,(c)-Frequency response magnitude of filters related with components represented in (a),(d)-Frequency response magnitude of filters related with components represented in (b)

The second simulation considers the influence of the noise on the algorithm performance. Then considering three narrow-band source signals (sinusoids) linearly mixed a random noise was added such as the signal-to-noise ratio in the range of $[0, 20]dB$. The algorithm parameters were: $M = 8, TH = 0.95, k_1 = 0$ and $k_2 = 1$. After choosing the three computed independent components that have the highest correlation with each of the source signals the mean square error (MSE) is computed. Figure (3) shows the results as well as a comparison with AMUSE (with the pencil computed with the same delays) for identical noise level. It can be verified that the proposed algorithm is always better than AMUSE and that the difference increases with the increase in noise energy.

Another important aspect of this experience is related with the influence of the a parameter TH when noise is considered. The parameter has always the same value, and as the noise energy increases the number of signals (l) after the application of the first step also increases (second column of table I). Consequently the number of independent signals computed at the output is also l (last column of table I). Nevertheless, computing the correlation coefficients (using the frequency information of the signals) between output signals and noise or source signals we confirm that some components are related with the sources and others with noise. The maximum values of the correlation coefficients are distributed between the

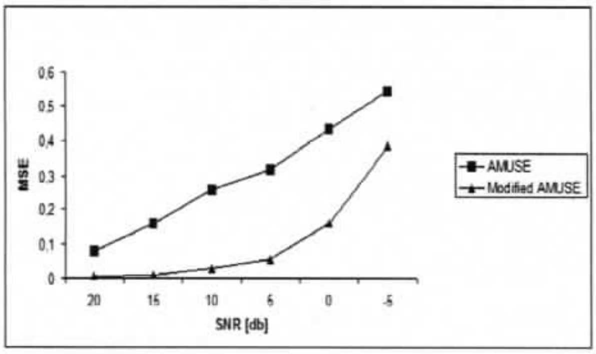

Fig. 3. Comparing AMUSE and modified AMUSE using MSE error

groups, in table I the third column shows the number of independent components that are related with sources while fourth column shows the number that are correlated with noise.

		Output correlated with		
Noise	**1st step**	**Sources**	**Noise**	**Total**
20db	5	5	0	5
15db	7	6	1	7
10db	10	6	4	10
5db	18	14	4	18
0db	19	10	9	19

TABLE I

NUMBER OF SIGNALS AT THE OUTPUTS OF THE 2 MAIN STEPS OF THE ALGORITHM FOR DIFFERENT LEVELS OF NOISE

Those results show that for a low noise level the first step (which is mainly a principal component analysis in a space of dimension NM) achieves good solutions. But when noise level increases the second step improves those results. This was verified for one of the high energy noise by taking only the five most significant eigenvalues, and corresponding eigenvectors, at the output of the fist step. In these cases the source signals were not recovered.

B. Heart Rate Variability study

We use ECG signals of two young normal subjects from POLY/MEDLAV database [10]. One lead of the stored ECG signal was processed by a wavelet transform based automatic delineation system [11]and the RR and QT sequences measured over the obtained marks.The ECG signals of two subjects were chosen because the corresponding respiration signals have spectral contents in distinct bands: (1)-ECG1 - in the low frequency band $(0.05 - 0.1Hz)$; (2)-ECG2 - in high frequency band $(0.2 - 0.4Hz)$

The proposed algorithm was applied $[RR(k), QT(k)]^T$ sequences. In all the experiments the free parameters of the algorithm are: $M = 8$, $k_1 = 0$, $k_2 = 1$ and $TH = 0.95$. The computed independent component signals are not correlated in time domain and suffer from an indetermination on its amplitude as it is expected on these methods. Then in order to achieve a direct comparison they are normalized to unitary variance and amplitude in range [-1,1]. However, in HRV studies, the frequency contents in distinct bands are the relevant information. An analysis in frequency is performed for each independent signal using the Welch method. The

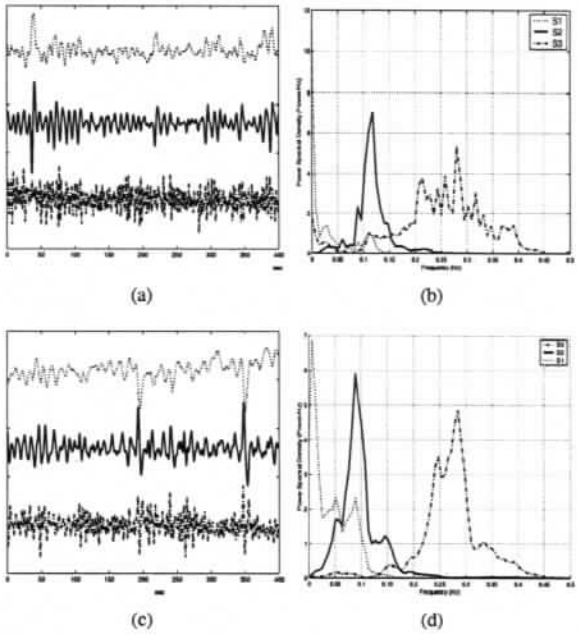

Fig. 4. ECG1: (a)-Independent Components, (b)-Spectral contents of independent components; ECG2: (a)-Independent Components,(b)-Spectral contents of independent components

results in frequency domain are compared graphically and the correlation coefficients are also computed. Three different independent components $(S_i, i = 1...3)$ are computed having frequency contents in distinct frequency ranges as shown in figure 4. In table II we can also see that the correlation

	ECG1			ECG2		
	S_1	S_2	S_3	S_1	S_2	S_3
S_1	1	0.10	0.09	1	0.38	0.14
S_2	0.10	1	0.03	0.38	1	0.13
S_3	0.09	0.03	1	0.14	0.13	1

TABLE II

CORRELATION COEFFICIENTS BETWEEN THE SPECTRA OF THE INDEPENDENT COMPONENTS

between the frequency contents of the computed independent signals are very low. And we can also verify that two of those signals have a frequency contents in the range of the LF and HF fluctuations considered in the HRV studies.

V. CONCLUSIONS

In this work we propose a modified version of the algorithm AMUSE. The algorithm is based on the generalized eigendecomposition of a matrix pencil computed with time-delayed mixed data. It was shown that the independent components achieved on the output are filtered versions of the source signals. An implementation was also proposed which differs from the algorithm AMUSE (and other blind source separation algorithms) because uses a different criteria for choosing the number of signals at the the output of the first step (which is very similar to a principal component decomposition). The algorithm has a set of parameters whose choice must be further studied, particularly the number of delays(M) used to build the the data set. It's choice naturally constraints the linear filtering operation that characterizes method as it was shown in the simulations. The simulation also reveal that the noise reduction was not completely achieved by the PCA. Then having at the output an high number of signals it is also important to find an automatic procedure to choose the relevant signals some might only be related with noise. The choice of time-delayed matrices to compute the pencil was done according to the algorithm AMUSE as well as the values for time-delays (k_1 and k_2)[1]. Nevertheless other second-order statistics [12][13] might be considered because were achieved better results when compared with AMUSE[12].

In what concerns the HRV study, blind source separation techniques were used to study identical time-series signals (RR and QT) and to discuss the relation of the independent components with the LF and HF fluctuations[14]. This preliminary study achieves similar results in spite of the use of different processing steps namely the pre-processing of the temporal sequences and the method to compute the independent signals. It seems a promising tool to HRV studies. Nevertheless a validation against a conventional method as well as a large data set should be used to verify the reliability and the performance of the method.

REFERENCES

[1] A. Chichocki and S.-I. Amari, *Adaptive Blind Signal and Image Processing*. Wiley, 2002.

[2] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using secon-order statistics," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 434–444, 1997.

[3] R. Gharieb and A. Chichocki, "Second-order statistics based blind source separation using a bank of subband filters," *Digital Signal Processing*, vol. 13, pp. 252–274, 2003.

[4] T. Schereiber and H. Kantz, "Nonlinear projective filtering ii: Application to real time series," in *Proceedings of Nolta*, 1998.

[5] R. Vetter, J. Vesin, P. Celka, P. Renevey, and J. Krauss, "Automatic nonlinear noise reduction using local principal component analysis and mdl parameter selection," in *IASTED Internation Conference on Signal Processing, Pattern RecognitionApplications*, (Crete, Greece), pp. 290–294, 2002.

[6] L. Tong, R.-w. Liu, V. C. Soon, and Y.-F. Huang, "Indeterminacy and identifiability of blind identification," *IEEE Transactions on Circuits and Systems*, vol. 38, no. 5, pp. 499–509, 1991.

[7] "Heart rate variability - standards of measurement, physiological interpretation, and clinical use," tech. rep., Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, 1996.

[8] A. M. Tomé and N. Ferreira, "On-line source separation of temporally correlated signals," in *European Signal Processing Conference, EUSIPCO2002*, (Toulouse, France), 2002.

[9] Parlett, *The Symmetric Eigenvalue Problem*. SIAM's Classics in Applied Mathematics, SIAM, 1998.

[10] F. Pinciroli, G. Pozzi, R. Rossi, M. Piovosi, A. Capo, R. Olivieri, and M. D. Torre, "A respiration-related ekg database," in *Computers in Cardiology* (I. C. Society, ed.), pp. 477–480, 1988.

[11] J. P. Martínez, S. Olmos, and L. Laguna, "Evaluation of wavelet-based ecg waveform detector on the qt database," in *Computers in Cardiology* (I. C. Society, ed.), pp. 81–84, 2000.

[12] A. M. Tomé, "Separation of a mixture of signals using linear filtering and second order statistics," in *10th European Symposium on Artificial Neural Networks*, (Brugges), pp. 307–312, 2002.

[13] J. V. Stone, "Blind source separation using temporal predictibility," *Neural Computation*, vol. 13, no. 7, pp. 1557–1574, 2001.

[14] R. Vetter, N. Virag, J. M. Vesin, P. Celka, and U. Scherrer, "Observer of autonomic outflow based on blind source separation of the ecg parameters," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 5, pp. 578–582, 2000.