FOETAL ECG EXTRACTION USING BLIND SOURCE SEPARATION METHODS

E. Bacharakis, A. K. Nandi and V. Zarzoso

Signal Processing Division, Department of Electronic and Electrical Engineering
University of Strathclyde, 204 George Street, Glasgow G1 1XW, U.K.
e-mail: asoke@eee.strath.ac.uk

ABSTRACT

Three methods to achieve Blind Source Separation are applied to the foetal electrocardiogram (ECG) extraction problem: Principal Component Analysis (PCA), Higher-Order Singular Value Decomposition (HOSVD) and Higher-Order EigenValue Decomposition (HOEVD). The first one gives uncorrelated source signals by means of second-order tools, while the last two resort to higher-order statistics of the data signals, so higher-order independence is attained. When tested on real ECG data, the last two produce better results than the former, with the HOEVD yielding the best performance, as expected from the theoretical unfolding.

1 INTRODUCTION

1.1 Foetal Electrocardiogram Extraction

One interesting problem on Biomedical Engineering is the foetal electrocardiogram (ECG) extraction. So-called *invasive techniques* have been shown to be very accurate, but require access to the foetal scalp. An earlier diagnosis using *non-invasive techniques* is desirable. These usually involve measurements form electrodes located on different points of the mother's skin. As a result, foetal ECG (FECG) and maternal ECG (MECG) components appear mixed together in the recordings. In addition, other random disturbances must be considered, which become important due to the low level of the FECG signals: mains interference, maternal electromyogram (EMG) interference, thermal noise from the electrodes and other electronic equipment, etc. Hence, it is necessary to *separate* the desired foetal contributions from the rest of the non-desired signals that disturb the cutaneous recordings. This task of obtaining the FECG from skin electrode signals is an example of the so-called *inverse problems*, which can be coped by using Blind Source Separation tools.

1.2 Blind Source Separation (BSS)

Let $\{x_i(t),\ i = 1, 2, \ldots, q\}$ be the set of q source signals, statistically independent and zero-mean, and $\{y_i(t),\ i = 1, 2, \ldots, p\}$ the p observed signals, formed by unknown linear combinations of the unknown sources, plus some additive noise $\{n_i(t),\ i = 1, 2, \ldots, p\}$, assumed to be Gaussian, mutually independent as well as independent from the source signals. Then, the following matrix equality can be established:

$$\begin{bmatrix} y_1(t) \\ \vdots \\ y_p(t) \end{bmatrix} = \begin{bmatrix} m_{11} & \ldots & m_{1q} \\ \vdots & \ddots & \vdots \\ m_{p1} & \ldots & m_{pq} \end{bmatrix} \cdot \begin{bmatrix} x_1(t) \\ \vdots \\ x_q(t) \end{bmatrix} + \begin{bmatrix} n_1(t) \\ \vdots \\ n_p(t) \end{bmatrix} \qquad (1)$$

or $Y = M \cdot X + N$. The aim of BSS is to retrieve X, the source signals, and M, the linear transformation from the sources to the observations, from the knowledge of the data signals contained in Y.

As far as the FECG extraction problem is concerned, we assume that the data consists of p ECG signals which are obtained by p electrodes located on the mother's body. After digitisation of the signals, each one consists of T samples and the data can be written as the matrix Y above with dimensions $p \times T$. It is stated in [1] that the MECG can be composed as a linear combination of m statistical independent vectors, which form the MECG-subspace. Accordingly, the FECG can be composed as a linear combination of f vectors, which span the FECG-subspace. These signals that compose the ECG subspaces are the *source signals*, and can be written as the matrix X above with dimensions $q \times T$, where $q = m + f$. The linear combination coefficients form the matrix M $(p \times q)$ which represents the transfer from the bioelectric sources to the electrodes. If the additive noise that unavoidably appears in all real applications is also taken into account, represented by a matrix N with dimensions $(p \times T)$, we have exactly a problem like (1). Therefore, the separation of the foetal from the maternal ECG can be modelled as a BSS problem: given T samples of the p data signals (matrix Y), estimate the source signals (matrix X) and the coefficients of the linear transformation (matrix M). The estimations of the FECG-subspace vectors gives FECG signals free from MECG. Moreover, the coefficients of the linear transformation indicate how strongly the different electrodes capture each source signal, from which better measurement positions may be deduced. Hence,

In *Proceedings EUSIPCO '96*, Trieste, Italy, 10-13 September 1996, pp. 395-398.

the achievement is two-fold.

In this paper, three BSS methods, briefly expounded in the next section, are used on real ECG data and their performance is evaluated in terms of both simple visual inspection and the degree of statistical independence of the obtained source signals.

2 METHODS USED FOR BSS

The methods that are to be evaluated here are: Principal Component Analysis (PCA), Higher-Order Singular Value Decomposition (HOSVD) and Higher-Order EigenValue Decomposition (HOEVD). The source signals found with the PCA method are second-order independent. The last two belong to the kind of methods known as Independent Component Analysis (ICA), where the higher-order statistical independence of the source signals is maximised by resorting to the higher-order statistics of the data matrix Y.

2.1 PCA Method

It is based on the Singular Value Decomposition (SVD) of the data matrix [1]:

$$Y = U \cdot S \cdot V^T \tag{2}$$

where U $(p \times p)$ and V $(T \times p)$ are orthogonal matrices containing the left and right singular vectors, respectively, of Y, and S $(p \times p)$ is a diagonal matrix with the singular values of Y. Interesting properties of this type of matrix decomposition as well as its applications to Signal Processing can be found in [2]. As Y is derived from linear combination of the q source signals, its rank should be q and consequently S and V^T should have dimensions $p \times q$ and $q \times T$, respectively. However, because of the noise, V^T actually has dimensions $p \times T$, where its last $(p-q)$ rows are associated to the noise signals, and S dimensions $p \times p$, with $s_{kk} \neq 0$, for $k = q+1, \ldots, p$. These last elements of S can be used to form an estimate of the noise variance, from the covariance matrix of Y, and then to set up a new singular value matrix S' corresponding to the noiseless case. Additionally, a new V' matrix can be formed, which contains only the information signals, while the new U' matrix would contain only the columns of U that correspond to the desired signals. The noise-free data matrix Y' is calculated from these new matrices as $Y' = U' \cdot S' \cdot V'^T$. Finally, estimations for M and X are attained through $M = U' \cdot S'$ and $X = V'^T$.

2.2 ICA-HOSVD Method

This method, which is formally introduced in [1] and [3], makes use of both plain and higher-order SVD. It also uses the fourth-order cumulants of the observed signals. In the first place, the matrix M is factorized as $M = B \cdot Q$ where B has dimensions $p \times q$ and Q $(q \times q)$ is orthogonal. B is determined by the second-order statistics of Y, namely its covariance matrix, in combination with its SVD, which yield $B = U \cdot S$. Note that this corresponds to the matrix M found in the PCA method. Then, a four-dimensional array is computed as a function of the fourth-order cumulant of Y:

$$\phi = C_4^Y \times_1 B^* \times_2 B^* \times_3 B^* \times_4 B^* \tag{3}$$

where B^* is the pseudoinverse of B and $C_4^Y(i,j,k,l) = Cum(y_i, y_j, y_k, y_l)$ is the 4th-order cumulant of the signals $y_i(t)$, $y_j(t)$, $y_k(t)$ and $y_l(t)$. Applying now the suitable properties of higher-order array operations and cumulants [4], it is shown that Q can be calculated as the left singular matrix of the Higher-Order Singular Value Decomposition (HOSVD) of (3). Finally, once M has been obtained, the expression $X = M^* \cdot Y = Q^T \cdot V^T$ gives the separated source signals, where M^* is the pseudoinverse of M and V the right singular matrix of Y. This method enhances the quality of the separation, compared with the preceding one, because it exploits more available information, i.e., the higher-order statistical properties of the data signals.

2.3 ICA-HOEVD Method

A set of stochastic processes are statistically independent if and only if all their cross-cumulants are null. In order to achieve the maximum statistical independence at order r, the sum of the squares of all marginal cumulants of order r has to be maximised [5]:

$$\psi = \sum_{i=1}^{q} Cum_r^2(x_i) \tag{4}$$

where $Cum_r(x_i)$ is the rth-order marginal cumulant of the signal $x_i(t)$. It is also proved that it is sufficient to consider only pairwise cumulants. For two signals and considering 4th-order statistical independence, the criterion (4) becomes:

$$\psi = Cum_4^2(x_i) + Cum_4^2(x_j) \tag{5}$$

If $X = [x_i, x_j]^T$ and $X' = [x'_i, x'_j]^T$ are the signals before and after the maximisation, then $X' = Q \cdot X$, where Q is the Givens rotation matrix:

$$Q = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} \tag{6}$$

and the angle ϕ is obtained from the maximisation of (5). To extend this to more than two signals, estimations of the transfer matrix and source signals are found from the SVD of Y, as in the PCA method. Then, the procedure for the maximisation of the pairwise 4th-order independence is applied to all the $q(q-1)/2$ pairs of source signals. After every calculation of Q, the matrices M and X are updated as $M := M \cdot Q^T$ and $X := Q \cdot X$, where Q is similar to (6) but with dimensions $q \times q$. All this sweep of the source signals pairs is repeated until the convergence of the algorithm. The method is called HOEVD because the procedure is similar to a higher-order extension of the Jacobi algorithm for the computation of the EigenValue Decomposition (EVD), achieved by the diagonalization of a matrix.

3 APPLICATION TO REAL ECG DATA

3.1 Description of the Data

The experimental data consists of 8 recordings taken from a pregnant woman, which are shown in *Figure 1*. The sampling frequency is $f_s = 500$ Hz and the number of samples $T = 5000$, so the duration of the signals is 10 s. The first signals are taken from the abdominal region. Even though the foetal ECG is much weaker then the maternal one, it is detectable. The next three recordings are taken from the thoracic region. In these the foetal ECG is not detectable at all. Although the thoracic signals consist mostly of MECG, they are essential for the success of the method. Our purpose is to remove the MECG signals from the abdominal recordings and, to do this, we need *clear* MECG signals with the least possible contribution of FECG. This kind of signals are obtained from the thoracic region.

3.2 Performance of the Methods

In order to evaluate the performance of the methods, the source signals given by each of them appear in *Figures 2*, *3* and *4*. From these figures we see that the PCA method reveals only two clear MECG source signals (1st and 2nd signals in *Figure 2*). Regarding the FECG source signals, the first one is found to have a low frequency fluctuation. With the use of the second method, the ICA-HOSVD, these problem are overcome. The three MECG sources appear now clear (*Figure 3*) and the fluctuation of the FECG source signals disappears. The only disadvantage of the method is that the second FECG source signal is slightly more noisy. The third method, the ICA-HOEVD, presents almost similar behaviour to the last one. Moreover, the second FECG source signal is improved (*Figure 4*).

In addition to this visual test, a parameter to measure the quality of the separation is defined in terms of the cumulants of the resultant sources:

$$P_k = \mathop{\mathrm{E}}_{\substack{\forall(i,j)\\ i<j}} \left[\frac{|k_{40}| + |k_{04}|}{\sum\limits_{m+n=4} |k_{mn}|} \right] \qquad (7)$$

where $k_{mn} = Cum_{mn}(x_i, x_j)$. For a good separation $P_k \approx 1$, so that the cross-cumulants are very close to zero. The values of this parameter obtained from the results of each method are in *Table 1*. We observe that both the last two methods increase the value of the parameter P_k, which is consistent with the fact that they maximise the 4th-order statistical independence of the source signals. The ICA-HOEVD gives slightly better results than the ICA-HOSVD because it is specially designed to get maximum 4th-order independence. Note that the PCA and the ICA-HOSVD are not based on this criterion but, in spite of that, the comparison of the methods based on it gives the same impression as the comparison based on the figures.

Parameter P_k	PCA	HOSVD	HOEVD
Mean	0.7940	0.9049	0.9568
Std. deviation	0.1355	0.1081	0.0661

Table 1: Results of the separation in terms of the statistical independence of the obtained source signals.

4 CONCLUSIONS

In this contribution, three methods for the BSS problem have been tested on real ECG signals taken from a pregnant woman. The three of them have succeeded to separate the foetal from the maternal ECG. The overall conclusion is that the ICA methods give significantly better results than the PCA method, with the ICA-HOEVD offering a slightly better performance than the ICA-HOSVD. In short, the ICA can be seen as a higher-order refinement of the PCA which considerably improves the achieved separation.

Acknowledgements. We wish to thank L. D. Lathauwer, D. Callaerts and J. Vandewalle, from K. U. Leuven, Belgium, for providing the electrocardiogram recordings.

5 REFERENCES

1. L. D. Lathauwer, D. Callaerts, B. D. Moor and J. Vandewalle, "Fetal Electrocardiogram Extraction by Source Subspace Separation", *Proceedings IEEE/ATHOS Signal Processing Workshop on Higher-Order Statistics*, Spain, June 1995, pp. 134-138.

2. E. Biglieri and K. Yao, "Some Properties of Singular Value Decomposition and their Applications to Digital Signal Processing", *Signal Processing*, Vol. 18, No. 3, November 1989, pp. 277-289.

3. L. D. Lathauwer, B. D. Moor and J. Vandewalle, "Blind Source Separation by Higher-Order Singular Value Decomposition", *Proceedings EUSIPCO-94*, Edinburgh, Scotland, U.K., September 1994, pp. 175-178.

4. C. L. Nikias and A. P. Petropulu, *Higher-Order Spectra Analysis. A Nonlinear Signal Processing Framework*, Prentice Hall, Signal Processing Series, 1993.

5. P. Comon, "Independent Component Analysis, A new Concept?", *Signal Processing*, Vol. 36, No. 3, April 1994, pp. 287-314.

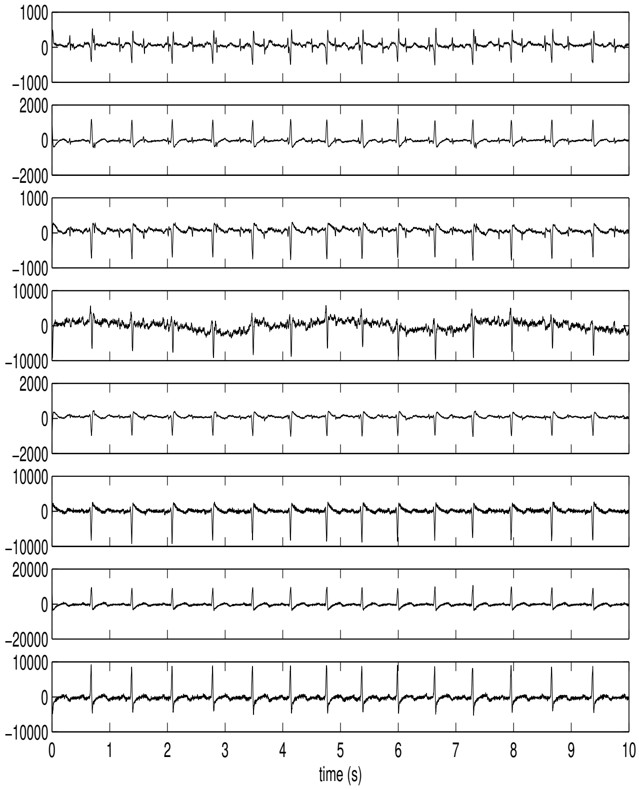

Figure 1: 8-channel electrocardiogram recording.

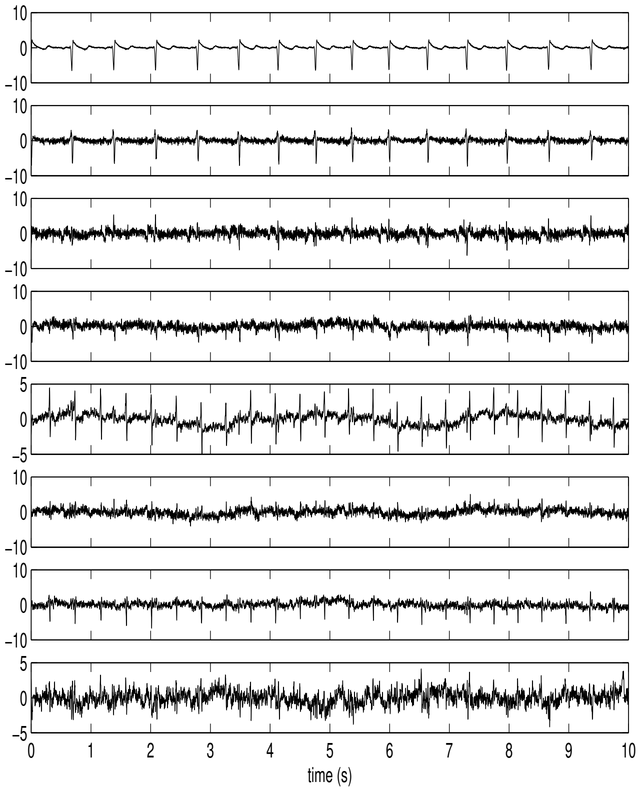

Figure 2: Source signals from the PCA method.

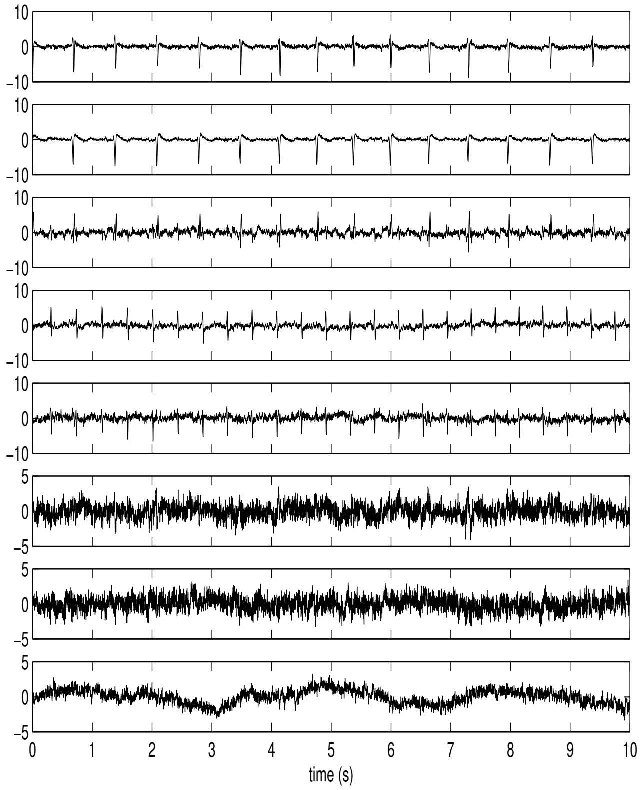

Figure 3: Source signals from the ICA-HOSVD method.

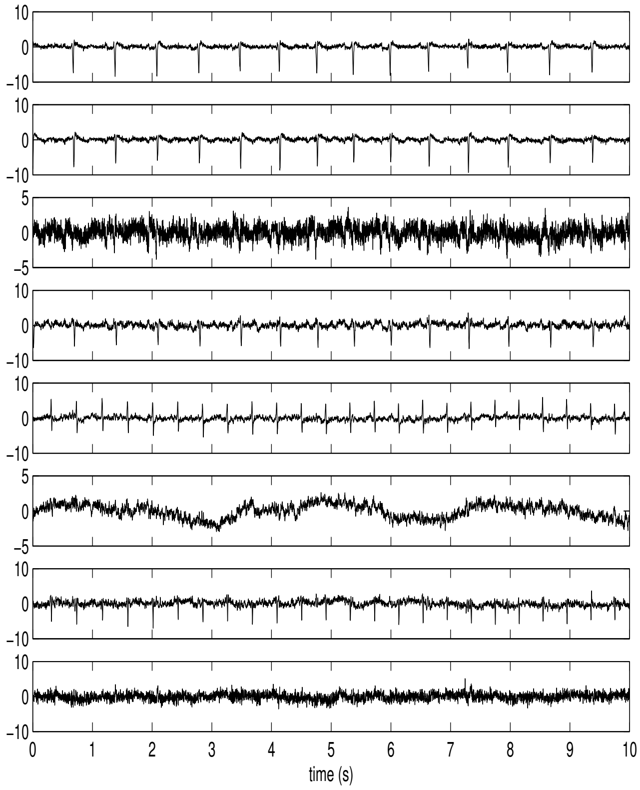

Figure 4: Source signals from the ICA-HOEVD method.