MATERNAL AND FOETAL ECG SEPARATION USING BLIND SOURCE SEPARATION METHODS

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

Department of Electronic and Electrical Engineering, University of Strathclyde,

204 George Street, Glasgow G1 1XW, U.K.

Abstract

The separation of the maternal and foetal electrocardiograms (ECGs) from skin electrodes located on the mother's body may be modelled as a Blind Source Separation (BSS) problem. This consists in the reconstruction of a set of unknown mutually independent source signals from the sole knowledge of another set of linear mixtures of the sources, where the mixture pattern is also unknown. In the present paper, three BSS methods based on cumulants are considered: Principal Component Analysis (PCA), Higher-Order Singular Value Decomposition (HOSVD) and Higher-Order Eigenvalue Decomposition (HOEVD). All of them are applied to the foetal ECG extraction problem by using real ECG data. The last two methods appear to provide a more satisfactory separation than the former, with the HOEVD offering slightly better results.

KEYWORDS: Signal Processing, Blind Source Separation, Principal Components, Independent Component Analysis, Higher-Order Statistics, Higher-Order Decomposition, Electrocardiogram, Maternal and Foetal Heartbeat.

1 INTRODUCTION

One interesting problem in Biomedical Engineering of concern to Signal Processing is the **foetal electrocardiogram (ECG) extraction**. It arises when the foetus' heart condition is desired to be monitored. So-called *invasive techniques* have been shown to be very accurate, but require access to the foetal scalp and therefore are only feasible during delivery. An earlier diagnosis during pregnancy using *non-invasive techniques* is more convenient. These usually involve measurements from electrodes attached to different points of the mother's skin. As a result, the recordings pick up a mixture of foetal ECG (FECG) and maternal ECG (MECG) contributions. 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. Under normal conditions, some of these disturbances can be eliminated. For instance, a persistent EMG interference usually indicates an uneasy position of the mother. Moreover, depending on the stage of pregnancy EMG interference produced by contraction of the uterus may be observed, as well as baseline wandering during foetal movements. Nevertheless, these disturbances are rather ocassional and most of the time the main one is due to thermal noise, which is an apparently white noise signal [VAN87]. Hence, it is necessary to extract the wanted foetal contributions from the rest of non-desired components that corrupt the cutaneous recordings. This task of obtaining the FECG from skin electrode signals is another example of the so-called *inverse problems*, which can be coped by using different techniques.

Since the mother's and foetus' heartbeats show different rates, the first approach one may come up with is a simple filtering in the frequency domain. However, there exists a spectral overlap between the maternal and the foetal ECG signals, which does not make this solution very successful. Furthermore, the heart rate is not constant, but exhibits a random variation with time, the so-called *heart rate variability (HRV)*, which also does not recommend the use of frequency filters to tackle the problem. Not even a more parametric approach seems to be more suitable. In the first place, the parameters of the model would be subject to a large uncertainty owing to differences from patient to patient; in the second place, the parameters of a single patient would not be time-invariant (for instance, because of the foetal growth); and, finally and most fundamentally, because in order to be applicable to medical diagnosis and treatment, the procedure should be able to deal with unexpected ECG patterns [LAT95].

All these considerations suggest that the FECG extraction has to be formulated as a blind identification problem. Precisely the main objective of this paper is to test how several methods for **Blind Source Separation (BSS)** cope with the foetal ECG extraction problem. First of all, in the next section the BSS is mathematically introduced in its most general case. Then the FECG extraction is framed in this context by showing that it can be modelled as a particular BSS problem. In order to solve it, three cumulant-based BSS methods are briefly reviewed in section 3. The first one only employs second-order statistical tools, while the other two resort to the higher-order properties of the observed data, namely their fourth-order cumulants. The experiments carried out to assess these methods involve real ECG data and are depicted in section 4. Conclusions from all these experiments are left for section 5. For the sake of clarity and completeness, some basic results on higher-order arrays and the Higher-Order Singular Value Decomposition have been moved to the appendices. Finally, a more detailed explanation on the properties and results of *Higher-Order Statistics (HOS)* mentioned throughout the text can be sought in [BRI81], [KEN87] and [NIK93].

2 BLIND SOURCE SEPARATION

The present section presents the mathematical formulation of a general BSS problem and highlights its inherent indeterminancy. In order to bound this indeterminancy, section 2.1 also introduces the concept of waveform-preserving BSS. Later, a necessary and sufficient condition to achieve the separation when an independent source model is taken into account, as for instance in the cumulant-based approaches, is stated. Finally, it is shown how the foetal ECG extraction can be modelled as a BSS problem.

2.1 Mathematical formulation of BSS

Let $\{x_i(t),\ i = 1, 2, \ldots, q\}$ be the set of q **source signals**, assumed to be statistically independent and zero-mean. These are the signals we desire to obtain from the p-channel **observation** $\{y_i(t),\ i = 1, 2, \ldots, p\}$, which, in a first order approximation, in most practical applications can be considered as formed by unknown instantaneous linear combinations of the unknown sources, plus some measurement additive **noise** $\{n_i(t),\ i = 1, 2, \ldots, p\}$. Then the observed signals may be defined by the following matrix equality:

$$\begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_p(t) \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & \ldots & m_{1q} \\ m_{21} & m_{22} & \ldots & m_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ m_{p1} & m_{p2} & \ldots & m_{pq} \end{bmatrix} \cdot \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_q(t) \end{bmatrix} + \begin{bmatrix} n_1(t) \\ n_2(t) \\ \vdots \\ n_p(t) \end{bmatrix}$$

or, in a more simplified form:

$$Y = M \cdot X + N \tag{1}$$

where $M = (m_{ij}) \in \mathbb{R}^{p \times q}$ represents the stationary linear transformation from the sources to the observations, also referred to as **channel parameter matrix** [TON91]. The aim of BSS is to retrieve X, the source signals, and M, the transfer matrix, from the sole knowledge of the data signals contained in Y. Figure 1 summarizes graphically the context and main purpose of BSS. It seems obvious that, in order to recover a certain number of sources, a greater number of linearly independent observations must be available. That is, $p > q$. The noise signals are assumed to be Gaussian, with zero-mean and variance σ_N^2, mutually independent as well as independent from the source signals. We can also assume that $\text{cov}\{X\} = I_q$, where I_q is the $(q \times q)$ identity matrix and $\text{cov}\{X\}$ the covariance matrix of the sources: $\text{cov}\{X\}(i, j) = \mathrm{E}[x_i \cdot x_j] - \mathrm{E}[x_i] \cdot \mathrm{E}[x_j] = \mathrm{E}[x_i \cdot x_j]$. This assumption means a scaling of the columns of the coefficients matrix M and does not disturb the generality of the formulation.

Usually the observed signals are digitised in such a way that each one consists of T samples. Then the sampled data can be represented as the matrix Y above with dimensions $p \times T$ and the sources as the matrix X with dimensions $q \times T$, while the matrix M does not suffer any variation. Hence the discrete-time BSS model can be written in matrix form as:

$$Y_{p\times T} = M_{p\times q} \cdot X_{q\times T} + N_{p\times T}. \tag{2}$$

It is obvious that there is more than one pair (M, X) that fulfil (2) given a observation Y and a noise matrix N. In fact, if (M_0, X_0) represents the actual transfer parameter matrix and the actual source signals, then $(M_0 \cdot Z^{-1}, Z \cdot X_0)$ also satisfies (2) and the additional constraint on the sources covariance matrix, for any orthogonal matrix $Z \in \mathbb{R}^{q\times q}$. Hence, the general BSS problem has an inherent indeterminancy which is characterized by a matrix Z which can be almost arbitrary. However, in many situations, when a set of source signals are to be separated the following considerations are applicable [TON91]:

1. The relevant information of the source signals is contained in their waveforms, rather than in their magnitudes, so a finite scale factor is acceptable.

2. The order in which the source signals are arranged does not matter.

The above observations mean that X_0 and $P \cdot D \cdot X_0$ may be regarded as equivalent sets of source signals, as far as the preservation of the waveforms is concerned, where $P \in \mathbb{R}^{q\times q}$ is a permutation matrix and $D \in \mathbb{R}^{q\times q}$ an invertible diagonal matrix. The latter represents the scale factors mentioned in 1, while the former signifies the change in the order of the obtained sources named in 2. As a matter of fact, X_0 and $P \cdot D \cdot X_0$ are said to be equivalent in the waveform-preserving sense. Hence, the term **waveform-preserving BSS** refers to BSS whose indeterminancy is restricted to a set of solution pairs in the form $(M_0 \cdot Z^{-1}, Z \cdot X_0)$ with $Z = P \cdot D$. If the separation is carried out through the estimation of a matrix $W \in \mathbb{R}^{q\times p}$ applied on the observations, it is considered to achieve a correct solution in the waveform-preserving sense if

$$W \cdot M = P \cdot D$$

so that the matrix of the resultant signals contains in its rows the sources up to, perhaps, a scale factor and an order change:

$$S = W \cdot Y = W \cdot M \cdot X = P \cdot D \cdot X.$$

2.2 Identifiability for waveform-preserving cumulant-based BSS

> *The source signals and the channel parameter matrix can be uniquely identified up to the waveform-preserving equivalence, if and only if at most one of the components of the source signals is Gaussian* [TON91].

In other words, two independent Gaussian processes cannot be retrieved from a set of linear mixtures of them, at least by using the HOS tools developed up to now. The reason for that lies in the Gaussianity property of cumulants, which states that the cumulants of order greater than 2 of a Gaussian random variable are all zero. This restricts the amount of information that can be used in order to separate this kind of sources and therefore leads to an undetermined set of possible sources with an infinite number of non-waveform-preserving solutions. This result also follows more fundamentally from the basic property that the sum of normals is normal too. Hence, there is no way of separating them in an additive model like that in (1).

2.3 BSS model for the foetal ECG extraction problem

If BSS methods are desired to be applied to the foetal ECG extraction problem, first it must be proved that it has a BSS model like that established in section 2.1. In order to do so, let us start by assuming that the recordings consists of p ECG signals which are obtained by p electrodes attached to different points of the mother's skin. After digitisation of the signals, each one consists of T samples and the data can be written as the matrix Y of (2) with dimensions $p \times T$. Now, it is stated in [LAT95] that the maternal ECG signals can be composed as a linear combination of m statistical independent vectors ($m = 3$), which form the MECG-subspace, or set of all possible ECGs that can be recorded from the mother. Accordingly, in the same reference it is also claimed that every ECG measured from the foetus can be composed as a linear combination of f vectors ($f = 2$), which span the FECG-subspace. These vectors that compose the ECG-subspaces are the source signals and can be written as the matrix X of (2) with dimensions $q \times T$, where $q = m + f$. The linear combination coefficients form the matrix M ($p \times q$) of (2) which represents the transfer from the bioelectric sources to the electrodes. Specifically, its element m_{ij} is the coefficient by which the jth source signal is present to the ith recording, and is determined by the geometry of the body, the electrode and source positions and the conductivity of the body tissues [VAN87]. If the additive measurement noise is also taken into account, it may be represented by a $(p \times T)$-matrix N and then we have exactly a problem like (1)–(2). Further on, it can be assumed that all of these noise signals have the same power, say σ_N^2, and that they are independent from the sources and from each other (this is a good approximation if $T \gg p$), i.e. $cov\{N\} \approx \sigma_N^2 \, I_p$. As a conclusion, the separation of the foetal and maternal ECGs from skin electrodes located on the mother's body 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 sources gives FECG contributions to each recording free from MECG and other disturbances. Effectively, if f_1 and f_2 represent the position of the estimated FECG sources, then the foetal heartbeat contribution to each recording is given by:

$$Y_f = \sum_{i=f_1,\, f_2} M_i \cdot X_i \tag{3}$$

where M_i stands for the ith column of the estimated transfer matrix and X_i the ith obtained source signal. In addition, the coefficients of M indicate how strongly the different electrodes capture each source signal, from which better measurement positions may be deduced. Hence, the achievement is two-fold. A more detailed explanation on this problem and its relationship with BSS can be found in [BAC95]. The main point is that BSS techniques can be employed to tackle it.

3 BSS METHODS

Three methods which aim at achieving BSS are to be evaluated here. The first of them is called **Principal Component Analysis (PCA)** and it only searches the removal of the correlation from the observed signals. The key tool to achieve this comes from "classical" linear algebra: it is the matrix Singular Value Decomposition (SVD). The other two methods are named **Higher-Order Singular Value Decomposition (HOSVD)** and **Higher-Order Eigenvalue Decomposition (HOEVD)**, and exploit the statistical independence of the sources up to the fourth order. Both belong to the sort of techniques known as *Independent Component Analysis (ICA)*, whose main objective is the elimination of higher-order dependence, which is a problem of HOS. In general, the ICA of a random vector consists in searching for a linear transformation that minimises the statistical dependence between its components [COM94]. The concept of ICA may actually be seen as a higher-order refinement of the PCA [BIG89], [COM91], [COM94], which can only impose independence up to second order (uncorrelation).

As a result, the quality of the separation is improved by employing the former, as it will be tested from the experimental results.

3.1 PCA method

This is the standard statistical technique based on the decomposition of the covariance matrix of the observed variables [MAR79]. The aim is to look for a few linear combinations of the observations such that they may be used to summarize the data, losing in the process as little information as possible. Furthermore, if the transformation matrix is properly chosen, these linear combinations can easily be made uncorrelated, yielding a estimation of the source signals. Nevertheless, the main problem that arises when using PCA for BSS is that it does not assume a model like (1), and so the noise signals will affect the estimation of the sources obtained through this procedure. In order to get rid of the noise effects, the covariance matrix of the noiseless model

$$Y' = M \cdot X$$

is estimated from that of the original model (1). Figure 2 sketches an algorithm for the computation of the PCA based on the SVD of the observations matrix Y. The SVD is a numerically more reliable way of factorizing the covariance matrix when it is estimated through the sample matrix Y:

$$C_2^Y = cov\{Y\} \approx \frac{1}{T} Y \cdot Y^t.$$

Moreover, it avoids the explicit computation of C_2^Y and improves the numerical accuracy [GOL89]. In figure 2 it can be seen how the singular values of the noiseless covariance matrix are obtained by substracting the singular values due to the noise. These usually (not always, as we shall see in the experimental results) correspond to the singular values with the lowest magnitude, i.e. the last elements of the diagonal matrix of the SVD of Y. The major difficulty stems from the fact that in reality the singular values of the noise, s_0, are not equal. In such a case, s_0 would be estimated as their mean value and at the new set of singular values they would all be set to zero. However, in some critical though real cases, the singular values of the noise cannot be distinguished from those of the information or desired signals, and so the estimation of the noiseless covariance matrix cannot be carried out. A thorough description of this method and some variations for the particular context of the problem we are dealing here is made in [VAN87] and [CAL90].

On the other hand, one could think that *Factor Analysis (FA)*, for instance, would be more suitable for this particular problem. As a matter of fact, the latter technique assumes a model exactly like (1) (in the FA framework, X would represent the "common factors", M the "factor loadings" and N the "specific" or "unique factors"). Further on, it does not assume that all the noise signals have the same power. However, only under certain conditions on the number of observed and source variables [MAR79] the factor model is well-defined and several methods can then be applied to solve it, such as Principal Factor Analysis or Maximum Likelihood Factor Analysis.

3.2 ICA-HOSVD method

This method, which is formally introduced in [LAT94] and [LAT95], makes use of both plain and Higher-Order SVD (see appendix B). The motivation behind it is to improve the result of the PCA by resorting to the HOS of the available signals in order to remove higher-order dependence from them. Let us assume that the channel parameter matrix M may be factorised as

$$M = B \cdot Q \tag{4}$$

where B is a non-singular $(p \times q)$-matrix and Q is orthogonal with dimensions $q \times q$. The matrix B is determined from the second-order statistics of Y, its covariance matrix, in combination with its SVD. From the covariance matrix of the observations, B turns out to be the matrix M found in the PCA method. The matrix Q will be obtained with the help of the fourth-order cumulants array of the matrix Y, and it is chosen to be orthogonal not to alter the already achieved uncorrelation. The reason why third-order cumulants are not used is simply because many practical probability distributions are almost symmetrical and so their cumulants of that order are usually bad conditioned. The 4th-order cumulants array of the observed signals is given by:

$$C_4^Y(i, j, k, l) = \mathrm{Cum}(y_i, y_j, y_k, y_l), \qquad C_4^Y \in \mathbb{R}^{p \times p \times p \times p}. \tag{5}$$

From the BSS definition equation $Y = M \cdot X + N$ and the independence property of cumulants:

$$C_4^Y = C_4^{MX} + C_4^N$$

since sources and noise are, by hypothesis, mutually independent, where C_4^Y, C_4^{MX} and C_4^N contain the fourth-order cumulants of Y, $M \cdot X$ and N, respectively. Furthermore, from the

Gaussianity property of cumulants it follows that $C_4^N = 0$, and from the multilinearity property:

$$C_4^{MX} = C_4^X \times_1 M \times_2 M \times_3 M \times_4 M \tag{6}$$

in which $C_4^X \in \mathbb{R}^{q\times q\times q\times q}$, $M \in \mathbb{R}^{p\times q}$ and "$\times_n$" denotes the n-mode product (appendix A).

Let us define

$$\Phi = C_4^Y \times_1 B^* \times_2 B^* \times_3 B^* \times_4 B^* \tag{7}$$

where $B^* \in \mathbb{R}^{q\times p}$ is the pseudoinverse [BIG89], [GOL89] of B. From (4), (6) and the corollaries 1 and 2 of appendix A, (7) becomes:

$$\Phi = \left(C_4^X \times_1 M \times_2 M \times_3 M \times_4 M\right) \times_1 B^* \times_2 B^* \times_3 B^* \times_4 B^* = \\ = C_4^X \times_1 Q \times_2 Q \times_3 Q \times_4 Q. \tag{8}$$

By definition, Q is orthogonal and also C_4^X is all-orthogonal in the sense of appendix B, owing to the sources' independence. Hence, from the results established in such an appendix, equation (8) is nothing else but the Higher-Order SVD of the 4th-order cumulants array Φ. Finally, in order to relieve the computational load of this method, Q may be calculated as the left singular matrix of the unfolding of Φ, defined as:

$$\Phi_{\text{unfold}}(i,\, j + q\,(k-1) + q^2\,(l-1)) = \Phi(i,\, j,\, k,\, l), \qquad i,\, j,\, k,\, l = 1,\, 2, \ldots,\, q. \tag{9}$$

Φ_{unfold} consists of q^2 matrices of dimensions $q \times q$ placed in a row in such a way that its overall dimensions are $q \times q^3$. After B and Q have been found, the linear transformation coefficients are calculated through (4), and the source signals as $X = M^* \cdot Y'$, being M^* the pseudoinverse of M. Equivalently:

$$X = M^* \cdot Y' = \left(\frac{1}{\sqrt{T}}\, U' \cdot S' \cdot Q\right)^* \cdot (U' \cdot S' \cdot V'^t) = \sqrt{T}\, Q^t \cdot V'^t$$

since Q is orthogonal, where $Y' = U' \cdot S' \cdot V'^t$ is the SVD of the noiseless observations matrix. From all the explained above, the algorithm for the ICA-HOSVD method can be outlined as done in figure 3.

3.3 ICA-HOEVD method

This method was originally established in [COM94] and it mainly relies on the independence criterion based on cumulants. This criterion can be briefly stated as follows: a set of random variables are statistical independent up to order r if and only if all their sth-order cross-cumulants are null, for $s \leq r$. The sources estimated through this procedure correspond to realizations of a random vector whose components are the "most independent possible", in the sense of the maximization of a given contrast function. This constrast function is basically the average mutual information of the estimated sources:

$$\Psi(p_x) = -I(p_x); \qquad I(p_x) = \int p_x(u) \log \frac{p_x(u)}{\prod p_{x_i}(u_i)}\, du; \qquad u \in \mathbb{R}^n,$$

where $p_x(x)$ is the probability density function (p.d.f.) of the random vector x and $p_{x_i}(x_i)$ the marginal p.d.f. of its ith component. As $\Psi(p_x)$ involves, in practice, an unknown density function, p_x is approximated by its Edgeworth expansion as a function of the cumulants of x. After some simplifications, the contrast function becomes:

$$\psi = \sum_{i=1}^{q} \mathrm{Cum}_r^2(x_i) \tag{10}$$

in which $\mathrm{Cum}_r(x_i)$ represents the rth-order marginal standardized cumulant of the variable x_i. The term "standardized" means that the initial set of observed variables has already been transformed into another set with unit covariance matrix, i.e. a PCA process has been carried out previously. It is surprising that a criterion only based on marginal cumulants works properly. However, taking into acount that

$$\Omega = \sum_{i_1 i_2 \ldots i_r} \mathrm{Cum}^2_{i_1 i_2 \ldots i_r}, \qquad \mathrm{Cum}_{i_1 i_2 \ldots i_r} = \mathrm{Cum}(x_{i_1},\, x_{i_2}, \ldots\, x_{i_r}),$$

is invariant under linear and invertible transformations preserving standardization (orthogonal transformations) then the maximization of ψ is equivalent to the minimization of $\Omega - \psi$, i.e. of the sum of the squares of all rth-order cross-cumulants. These cumulants are precisely the measure of statistical dependence at that order. In short, the maximization of ψ leads to the maximum statistical independence at order r, as claimed at the beginning. In particular, an orthogonal transformation Q will be sought in order to maximize ψ in (10) with respect to the set of orthogonal matrices. Although this criterion has been derived after simplifications, has the major advantage that it leads to an algorithm with polynomial execution time. It is also shown in [COM94] that when a contrast of polynomial form is utilized, it suffices to consider only pairwise independence. Thus a pairwise processing will provide the desired results while reducing the computational complexity of the method. Then, for two variables and considering 4th-order statistical independence the criterion (10) becomes:

$$\psi = \mathrm{Cum}_4^2(x_i) + \mathrm{Cum}_4^2(x_j). \tag{11}$$

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

$$Q = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} = \frac{1}{\sqrt{1+\theta^2}} \begin{bmatrix} 1 & \theta \\ -\theta & 1 \end{bmatrix}, \qquad \theta = \tan\phi \tag{12}$$

and the angle ϕ is obtained from the maximisation of (11). Expanding the expression (11) as a function of θ, one can arrive at:

$$\Psi(\xi) = (\xi^2 + 4)^{-2} \sum_{k=0}^{4} b_k\, \xi^k \tag{13}$$

where $\xi = \theta - 1/\theta$ and the coefficients b_k are given from non-linear combinations of the pairwise 4th-order cumulants of x_i and x_j. In order to maximise (13), its derivative must be found and set equal to zero:

$$\omega(\xi) = \sum_{k=0}^{4} c_k\, \xi^k = 0 \tag{14}$$

in which the coefficients c_k can also be written as a function of the 4th-order cumulants of the same pair of signals. The exact expressions for b_k and c_k can be found in the appendices of [COM94]. In particular, in the noiseless case this last polynomial degenerates into another one of degree 2 only. Then, extracting the roots of (14) and substituting them in (13) to verify which one yields the maximum value, the optimum rotation angle is obtained. To extend this to more than two signals, estimations of the transfer matrix and the source signals are found from the SVD, like in the PCA method. Then this procedure for the maximisation of the pairwise 4th-order independence is applied to all the $q\,(q-1)/2$ pairs of estimated sources. After every calculation of the optimum rotation angle, the matrices M and X are updated as $M := M \cdot Q^t$ and $X := Q \cdot X$, where Q has similar structure to that in (12) but with dimensions $q \times q$. All this sweep over the source signals pairs is repeated until the convergence of the algorithm, which is estimated to take at most $k_{\max} = 1 + \sqrt{q}$ sweeps. Therefore, the algorithm for the ICA-HOEVD method can be summarised as in figure 4. An interesting algebraic interpretation of this BSS method is given in [LAT94]. The procedure is called HOEVD because it may be seen as a higher-order extension of the Jacobi method for the computation of the Eigenvalue Decomposition (EVD) of a Hermitean matrix. Effectively, the statistical dependence is minimised by diagonalising, as far as possible, the 4th-order cumulants array of the output signals, which agrees with the cumulant-based independence criterion. Finally, note that the rotation procedure is reminiscent of the *varimax* method of FA. However, the purposes are different in each case, and so are the statistical tools employed. The varimax method aims at providing a factor loadings matrix (transfer matrix) with a few large elements and as many near-zero elements as possible. This is accomplished by an iterative maximization of a quadratic function of the loadings [MAR79].

3.4 Description of the ECG data

The experimental data consists of 8 recordings taken from 8 electrodes attached to different points of a pregnant woman's skin. This multi-channel ECG recording is displayed in figure 5. The signals have been sampled for 10 seconds at a sampling rate of $f_s = 500$ Hz, so that each signal consists of $T = 5000$ samples. The actual magnitude of the y-axis is not important, but only the relative amplitudes between the recordings. The first five signals have been measured from the abdominal region and even though the foetal ECG is much weaker than the maternal one, it can still be detected. The next three recordings are taken from the mother's thoracic region and, therefore, in them the foetal ECG is not detectable at all. Although the thoracic region signals consist mostly of MECG, they are essential for the success of the extraction. Our purpose is to remove the MECG signals from the abdominal region. In order to do this, we need clear MECG signals with the least possible contribution of FECG so that, somehow, they can be "substracted" from the abdominal recordings, leaving only the desired FECG components. There is experimental evidence that the extraction of thoracic signals as described guarantees MECG supression and simultaneously an optimal FECG to noise ratio. It has also been shown that supplementary abdominal electrodes improve the FECG to noise ratio [VAN87]. This real 8-channel ECG data will be used as an input matrix to the algorithms which implement the methods.

3.5 Results

The resulting source signals can be seen in figures 6, 8 and 10 for the PCA, the ICA-HOSVD and the ICA-HOEVD, respectively. In order to allow a clearer and more detailed comment on the results, figures 7, 9 and 11 show the five relevant ECG sources with a larger timescale (first 5 seconds) for the three methods. In the latter figures, the first three waveforms correspond to the retrieved maternal ECG sources, whereas the last two are the foetal ones. The **PCA** method reveals only two clear MECG signals (1st and 2nd waveforms in figs. 6 and 7), while the other one (4th signal in fig. 6 and 3rd one in fig. 7) is heavily contaminated by noise. As far as the FECG sources are concerned (5th and 7th signals in fig. 6 and 4th and 5th ones in fig. 7), the first one presents an undesirable low-frequency fluctuation, which is possibly related to the mother's respiration. It is also observed that the obtained noise sources (3rd, 6th and

8th waveforms in fig. 6) exhibit a series of peaks that correspond to the MECG and FECG signals, and so they cannot be considered as pure noise. Note that one out of the last three recovered sources in fig. 6 corresponds to a FECG source, and therefore its singular value could not be taken to estimate the noise singular values as done in the algorithm of fig. 2. This experimental result agrees with the comments made in section 3.1. With the second method, the **ICA-HOSVD**, the difficulties found in the PCA seem to be overcome. In the first place, the noise sources (last three waveforms in fig. 8) do not have any contribution from the ECG subspaces, and therefore they can be regarded as pure noise. Secondly, the third MECG signal (3rd waveform in figs. 8 and 9) appears much neater. The same occurs with the 1st MECG source, which is less noisy than its equivalent 2nd MECG source found by the PCA. Hence, the MECG subspace is now clearly revealed. Finally, the fluctuation that was present at the first of the FECG source signals found with the PCA method, now vanishes and forms a separate noise signal, that of the 8th recovered source in fig. 8. However, the other FECG signal (5th waveform in figs. 8 and 9) is slightly more noisy than its equivalent found with the PCA (7th signal in fig. 6 and 5th one in fig. 7). The third procedure, the **ICA-HOEVD**, presents almost similar behaviour to the previous one, but with a remarkable improvement in the second FECG source (7th waveform in fig. 10 and 5th one in fig. 11). Note that the application of the ICA-HOSVD had worsened this signal in comparison with the PCA, but the application of the ICA-HOEVD leaves it unchanged. For all the three methods, once the source signals have been estimated the foetal ECG contribution to each recording can be obtained through (3).

4 CONCLUSIONS

After introducing briefly the problem of the maternal-foetal ECG separation, it has been seen how it may be modelled from the BSS point of view. Then, three cumulant-based BSS methods (namely PCA, ICA-HOSVD and ICA-HOEVD) have been reviewed and applied to real ECG recordings taken from a pregnant woman. From a simple visual test on the obtained source waveforms, the first conclusion is that all the three procedures are able to extract the foetal sources from the cutaneous electrode signals and hence to obtain the foetal heartbeat contributions to each recording. Comparatively, the ICA methods have shown to provide a better behaviour than the PCA. This result corroborates the fact that ICA may be seen as a higher-order refinement of the PCA, which improves the quality of the attained separation by exploiting the statistical independence to fourth order of the source signals and by resorting to the higher-order statistical properties of the observations. More specifically, the ICA-HOEVD turns out to give the best performance out of the three procedures. This may occur on the account that this last method is specially designed to maximize the higher-order statistical independence of the estimated sources.

APPENDICES

Appendix A. Higher-order arrays

An Nth-order array Φ with complex elements is defined as a multidimensional matrix of dimensions $I_1 \times I_2 \times \ldots I_N$: $\Phi \in \mathbb{C}^{I_1 \times I_2 \times \ldots I_N}$. Some operations and properties on higher-order arrays are set out below [LAT94].

SCALAR PRODUCT. The scalar product $\langle \Phi, \Psi \rangle$ of two arrays $\Phi, \Psi \in \mathbb{C}^{I_1 \times I_2 \times \ldots I_N}$ is defined as:

$$\langle \Phi, \Psi \rangle \triangleq \sum_{i_1} \sum_{i_2} \cdots \sum_{i_N} \Psi^*_{i_1 i_2 \ldots i_N} \Phi_{i_1 i_2 \ldots i_N}$$

where the symbol $*$ denotes the complex conjugation.

ORTHOGONALITY. If $\langle \Phi, \Psi \rangle = 0$, the arrays Φ and Ψ are said to be mutually orthogonal.

FROBENIUS-NORM of an array Φ: $\quad ||\Phi|| \triangleq \sqrt{\langle \Phi, \Phi \rangle}$.

MULTIPLICATION OF A HIGHER-ORDER ARRAY BY A MATRIX. The n-mode product of the Nth-order array $\Phi \in \mathbb{C}^{I_1 \times I_2 \times \ldots I_N}$ by the matrix $U \in \mathbb{C}^{J_n \times I_n}$, denoted by $\Phi \times_n U$, is an $(I_1 \times I_2 \times \ldots J_n \times \ldots I_N)$-array whose entries are given by:

$$(\Phi \times_n U)_{i_1 i_2 \ldots j_n \ldots i_N} \triangleq \sum_{i_n} \Phi_{i_1 i_2 \ldots i_n \ldots i_N} U_{j_n i_n}.$$

Corollary 1. Given the array $\Phi \in \mathbb{C}^{I_1 \times I_2 \times \ldots I_N}$ and the matrices $F \in \mathbb{C}^{J_n \times I_n}$, $G \in \mathbb{C}^{J_m \times I_m}$, it follows that

$$(\Phi \times_n F) \times_m G = (\Phi \times_m G) \times_n F = \Phi \times_n F \times_m G.$$

Corollary 2. Given the array $\Phi \in \mathbb{C}^{I_1 \times I_2 \times \ldots I_N}$ and the matrices $F \in \mathbb{C}^{J_n \times I_n}$, $G \in \mathbb{C}^{K_n \times J_n}$, it follows that

$$(\Phi \times_n F) \times_n G = \Phi \times_n (G \cdot F).$$

Example. For the real matrices $S \in \mathbb{C}^{I_1 \times I_2}$, $U \in \mathbb{C}^{J_1 \times I_1}$ and $V \in \mathbb{C}^{J_2 \times I_2}$:

$$U \cdot S \cdot V^t \equiv S \times_1 U \times_2 V.$$

Appendix B. The Higher-Order Singular Value Decomposition

Every complex $(I_1 \times I_2 \times \dots I_N)$-array Φ can be written as the product [LAT94]

$$\Phi = \Sigma \times_1 U^{(1)} \times_2 U^{(2)} \cdots \times_N U^{(N)}$$

where:

- $U^{(n)} = \left[u_1^{(n)}, u_2^{(n)}, \dots, u_{I_n}^{(n)} \right]$ is a complex orthogonal $(I_n \times I_n)$-matrix
- the core array Σ is a complex $(I_1 \times I_2 \times \dots I_N)$-array whose subarrays $\Sigma_{i_n=\alpha}$ obtained by fixing the nth index to α have the properties of:
  - all-orthogonality: two subarrays $\Sigma_{i_n=\alpha}$ and $\Sigma_{i_n=\beta}$ are orthogonal, for all possible values of n, α and β subject to $\alpha \neq \beta$:

    $$\langle \Sigma_{i_n=\alpha}, \Sigma_{i_n=\beta} \rangle = 0, \qquad \alpha \neq \beta$$

  - ordering:

    $$||\Sigma_{i_n=1}|| \geq ||\Sigma_{i_n=2}|| \geq \cdots \geq ||\Sigma_{i_n=r_n}|| > 0$$

    and

    $$||\Sigma_{i_n=r_n+1}|| = \cdots = ||\Sigma_{i_n=I_n}|| = 0$$

    for all possible values of n.

The Frobenius-norms $||\Sigma_{i_n=i}||$, symbolized by $\sigma_i^{(n)}$ are the n-mode singular values of Φ and the vectors $u_i^{(n)}$ its ith n-mode singular vectors.

The SVD [BIG89], [GOL89] of a real $(m \times n)$-matrix A turns out to be a particular case of the Higher-Order SVD, with $n = 2$:

$$A = U \cdot S \cdot V^t = S \times_1 U \times_2 V, \qquad S \in \mathbb{R}^{m \times n},\ U \in \mathbb{R}^{m \times m},\ V \in \mathbb{R}^{n \times n}$$

as deduced immediately from the example of appendix A.

ACKNOWLEDGEMENTS. The authors wish to thank L. De Lathauwer, D. Callaerts and J. Vandewalle, from K. U. Leuven, Belgium, for providing the electrocardiogram recordings.

This work is partly funded by a grant from the Defence Research Agency of the UK. The authors are also indebted to the referees of the paper, whose comments have been extremely useful in improving the presentation of the material.

REFERENCES

[BAC95] E. Bacharakis, "Separation of Maternal and Foetal ECG using Blind Source Separation Methods", *MSc Thesis*, University of Strathclyde, Glasgow, Scotland, U.K., September 1995.

[BIG89] 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.

[BRI81] D. R. Brillinger, *Time Series. Data Analysis and Theory*, Holden-Day Inc., 1981

[CAL90] D. Callaerts, B. D. Moor, J. Vandewalle, W. Sansen, G. Vantrappen and J. Janssens, "Comparison of SVD methods to extract the foetal electrocardiogram from cutaneous electrode signals". *Medical & Biological Engineering & Computing*, May 1990, No. 28, pp. 217–224.

[COM91] P. Comon, C. Jutten and J. Herault, "Blind Separation of Sources, Part II: Problems Statement", *Signal Processing*, Vol. 24, No. 1, November 1991, pp. 11–20.

[COM94] P. Comon, "Independent Component Analysis, A new concept?", *Signal Processing*, Vol. 36, No. 3, April 1994, pp. 287–314.

[GOL89] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, Second Edition, 1989.

[KEN87] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics*, Vol. I, Charles Griffin & Company Ltd., Fifth Edition, 1987.

[LAT94] L. D. Lathauwer, B. D. Moor and J. Vandewalle, "Blind Source Separation by Higher-Order Singular Value Decomposition", in: M. Holt et alt. (Eds.), *Proceedings EUSIPCO-94*, Edinburgh, Scotland, U.K., September 1994, pp. 175–178.

[LAT95] L. D. Lathauwer, D. Callaerts, B. D. Moor and J. Vandewalle, "Fetal Electrocardiogram Extraction by Source Subspace Separation", *Proceedings IEEE/ATHOS Signal Processing Conference on Higher-Order Statistics*, Spain, June 1995, pp. 134–138.

[MAR79] K. V. Mardia, J. T. Tent and J. M. Bibby, *Multivariate Analysis*, Academic Press, 1979.

[NIK93] C. L. Nikias and A. P. Petropulu, *Higher-Order Spectra Analysis. A Nonlinear Signal Processing Framework*, Prentice-Hall, Signal Processing Series, 1993.

[TON91] L. Tong, S. Yu, Y. Inouye and R. Liu, "A Necessary and Sufficient Condition of Blind Identification", *International Signal Processing Workshop on Higher-Order Statistics*, Chamrousse, France, July 1991, pp. 261–264.

[VAN87] J. Vanderschoot, D. Callaerts, W. Sansen, J. Vandewalle, G. Vantrappen and J. Janssens, "Two methods for optimal MECG elimination and FECG detection from skin electrode signals", *IEEE Transactions on Biomedical Engineering*, Vol. BME-34, No. 3, March 1987, pp. 233–243.

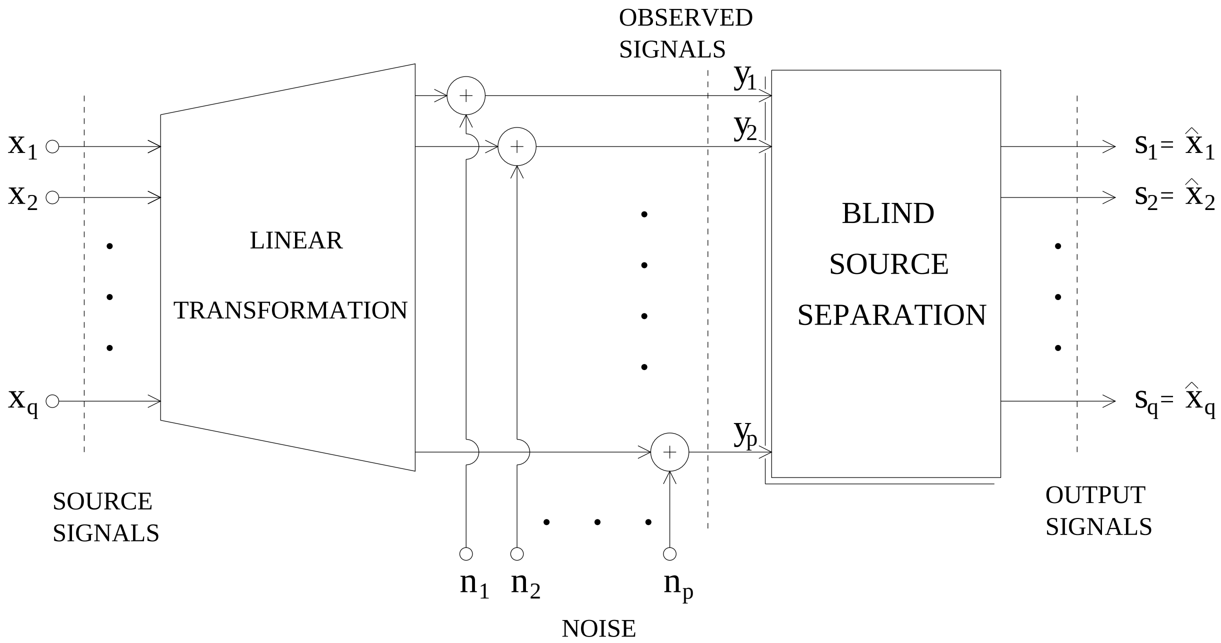

Figure 1: Context and purpose of Blind Source Separation (BSS).

1. Compute the SVD of the data matrix Y: $\quad Y = U \cdot S \cdot V^t$.

2. Estimate the singular values corresponding to the noise: s_0.

3. Obtain the singular values matrix of the noiseless problem:

$$S' = \mathrm{diag}\{s'_k\}, \qquad \begin{cases} {s'}_k^2 = s_k^2 - s_0^2 \\ s'_k = 0, \quad \text{for the noise singular values.} \end{cases}$$

4. Estimation of the channel parameter matrix: $M = \frac{1}{\sqrt{T}} U \cdot S'$.

5. Estimation of the source signals: $X = \sqrt{T}\, V_q^t$, where V_q contains only the q right singular vectors corresponding to the sources (i.e., all except for those corresponding to the noise) and the scale factor $\sqrt{T}$ appears in order to preserve the covariance matrix of the sources equal to the identity matrix: $C_2^X = I_q$.

Figure 2: Algorithm for the PCA method.

1. Obtain B as the matrix M achieved through the application of the PCA method over the data signals in Y.
2. Compute the 4th-order cumulant array of Y, C_4^Y (5).
3. Create the 4th-dimensional array Φ by using equation (7).
4. Form the unfolding matrix of Φ, Φ_{unfold}, as set out by expression (9).
5. Obtain the SVD expansion of Φ_{unfold}. Its left singular vectors form the matrix Q.
6. Estimation of the transfer matrix: $M = B \cdot Q$.
7. Estimation of the source signal: $X = \sqrt{T}\, Q^t \cdot V^t$, where V comes from the SVD of the observations matrix.

Figure 3: Algorithm for the ICA-HOSVD method.

1. Initialise: $M = M_{\text{PCA}}$, $X = X_{\text{PCA}}$, i.e. the transfer matrix and source signals estimated by the PCA method.
2. Main loop: for $k = 1, 2, \ldots, k_{\max}$, $\quad k_{\max} = 1 + \sqrt{q}$.
3. Sweep the $q\,(q-1)/2$ pairs of estimated sources $[x_i = X(i, :), \, x_j = X(j, :)]^t$ according to a fixed ordering. For each pair:
   3.1. Find the angle θ from (13)–(14) and set up the rotation matrix Q.
   3.2. Update transfer matrix and source signals: $M := M \cdot Q^t, \quad X := Q \cdot X$.
4. End k-loop.

Figure 4: Algorithm for the ICA-HOEVD method.

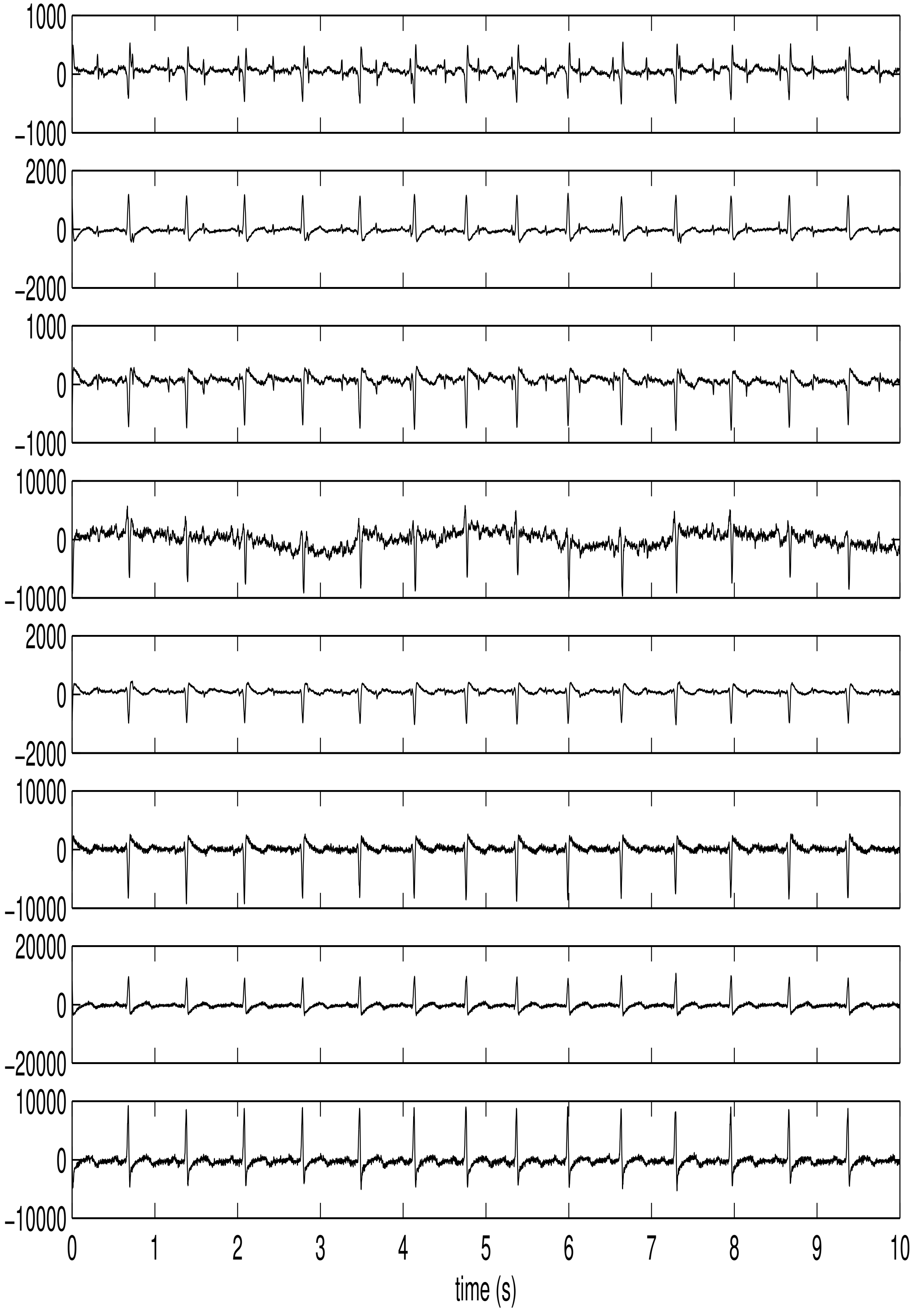

Figure 5: 8-channel electrocardiogram recording.

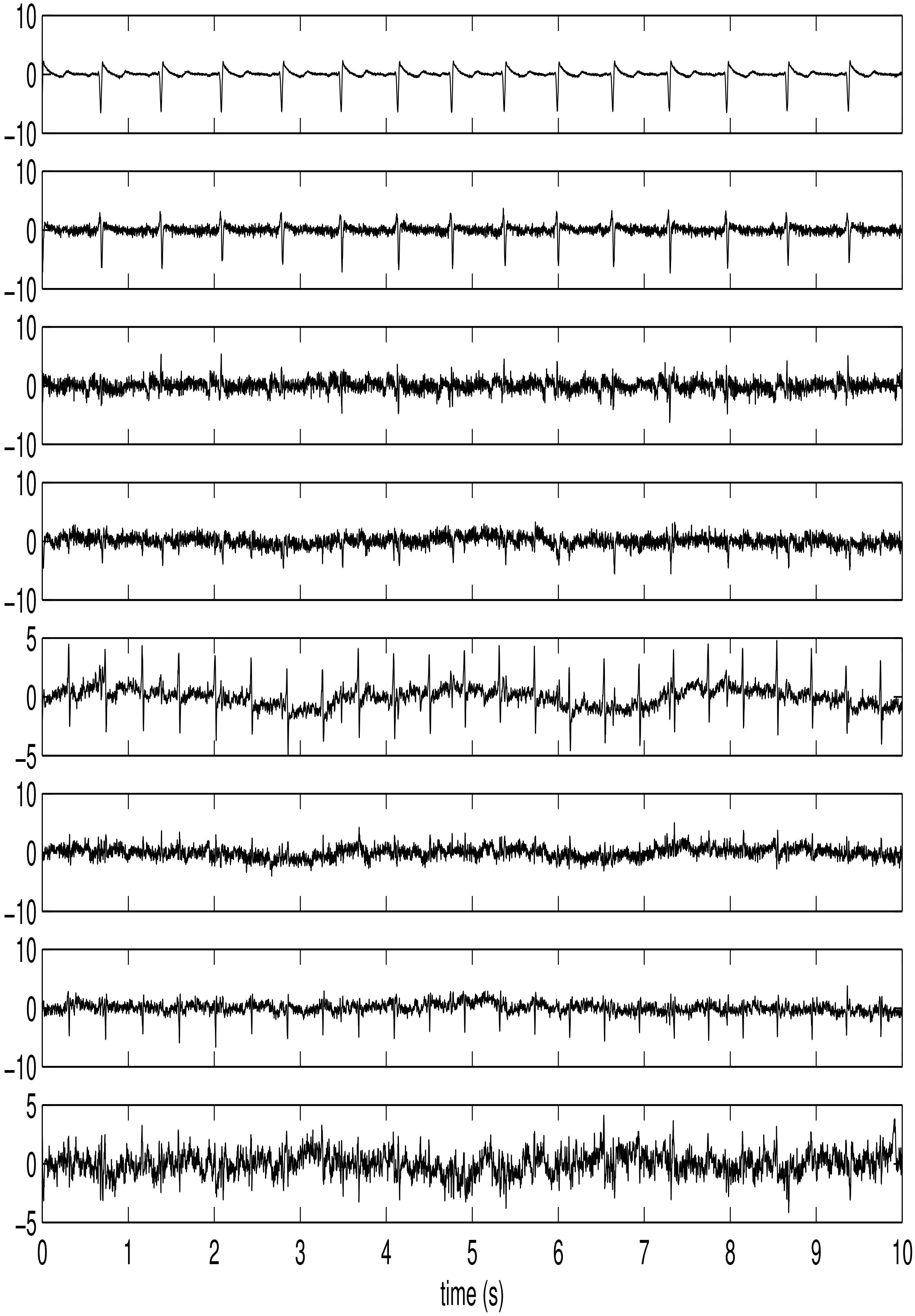

Figure 6: Source signals from the PCA method.

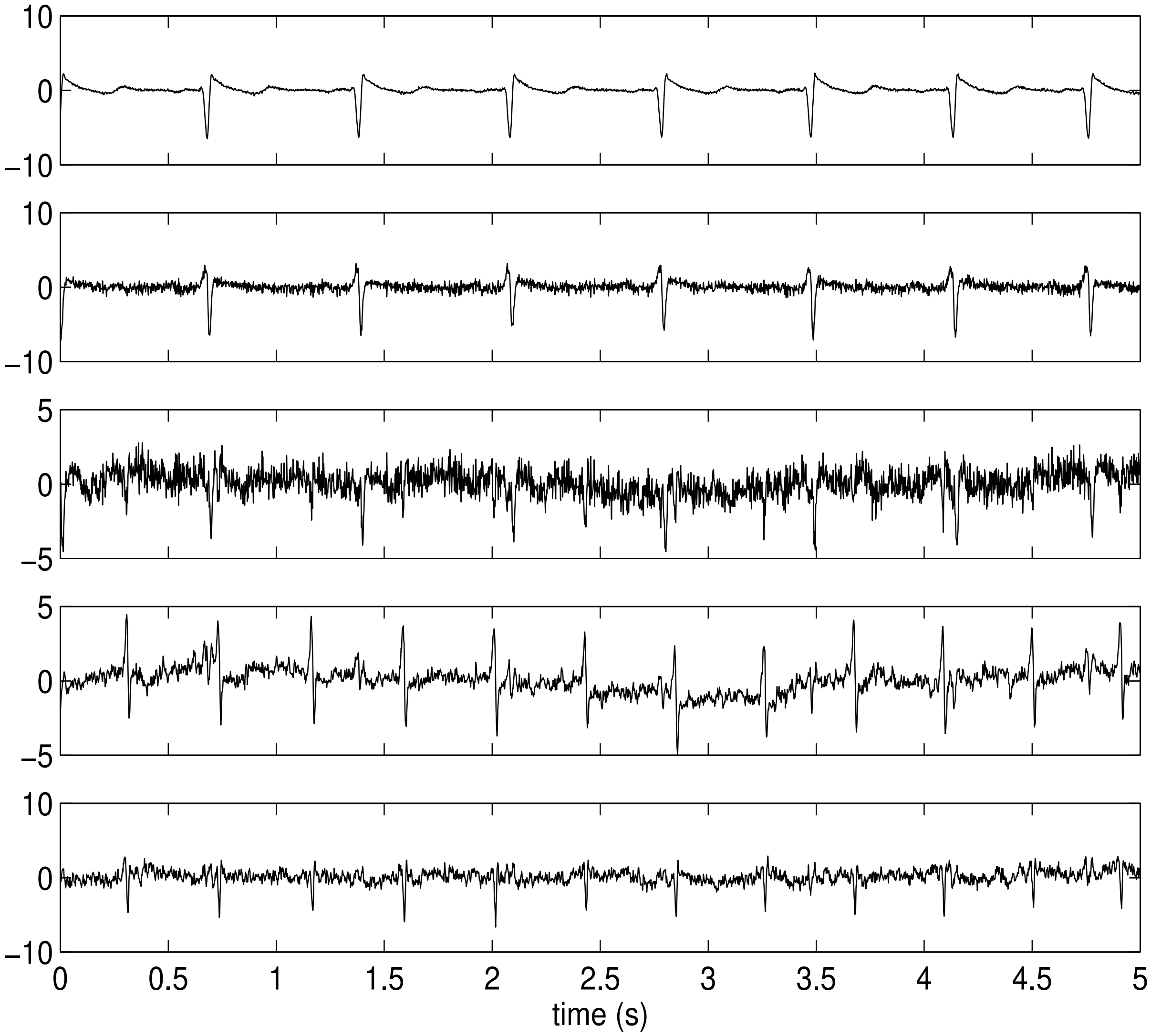

Figure 7: ECG sources from the PCA method.

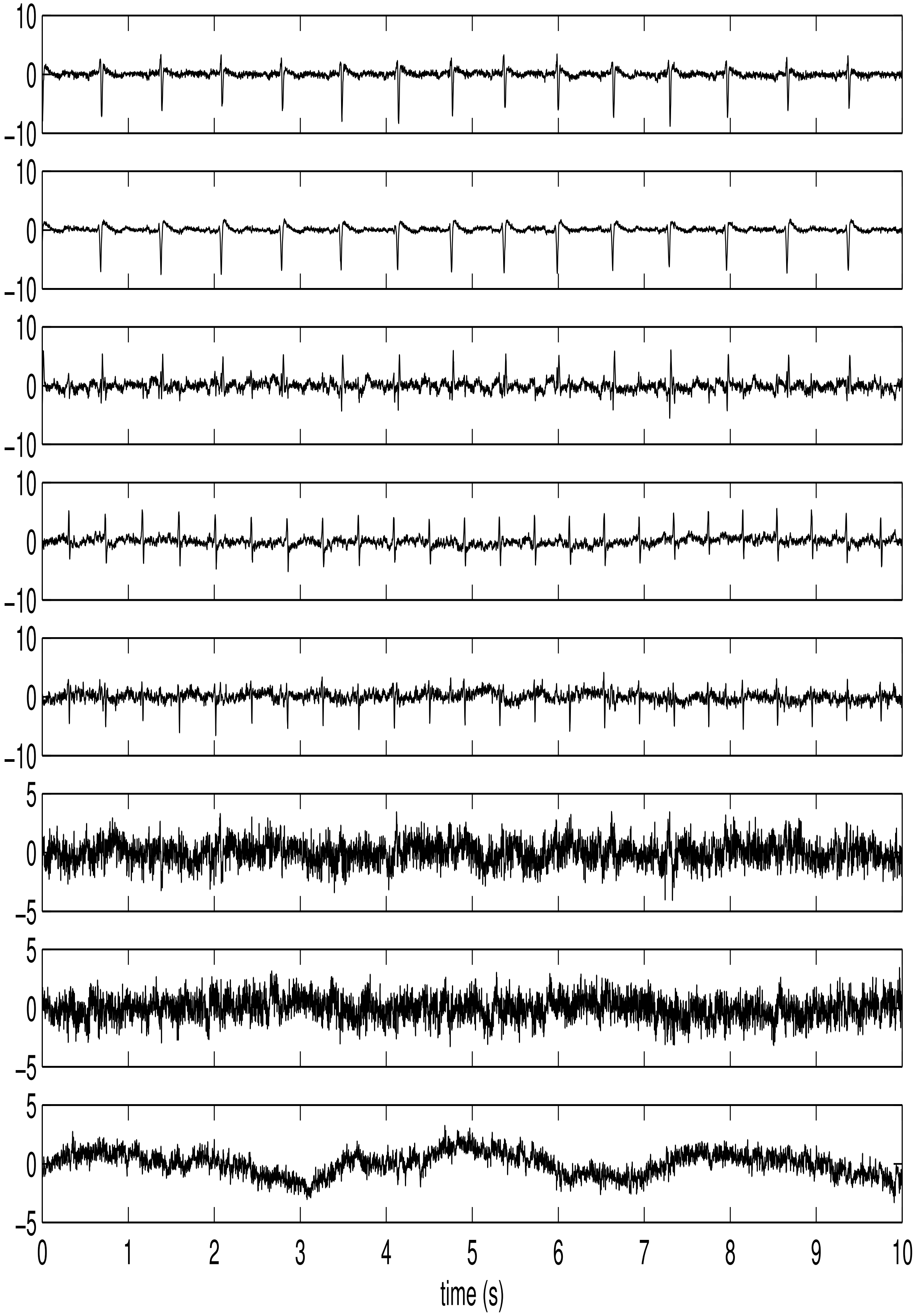

Figure 8: Source signals from the ICA-HOSVD method.

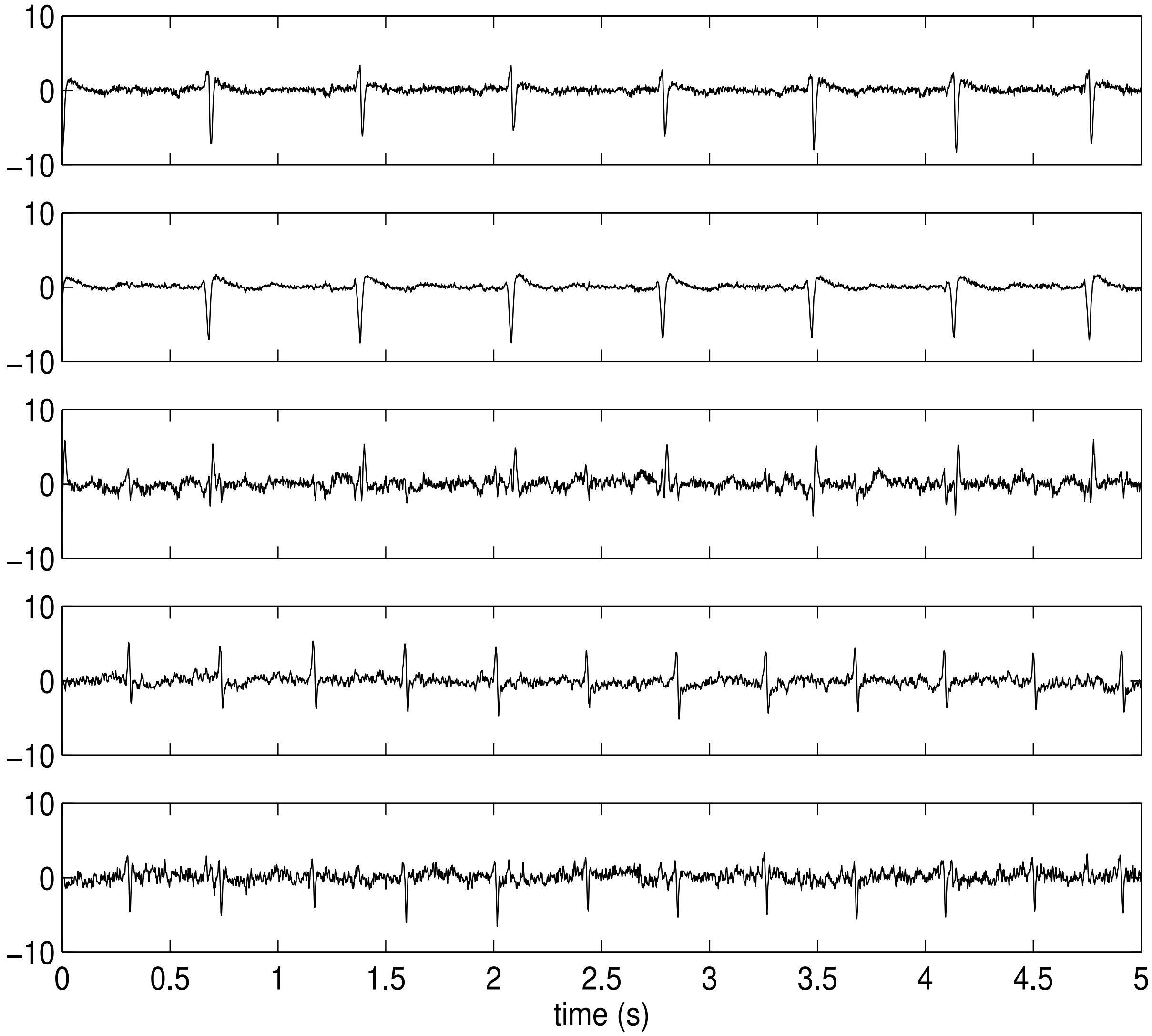

Figure 9: ECG sources from the ICA-HOSVD method.

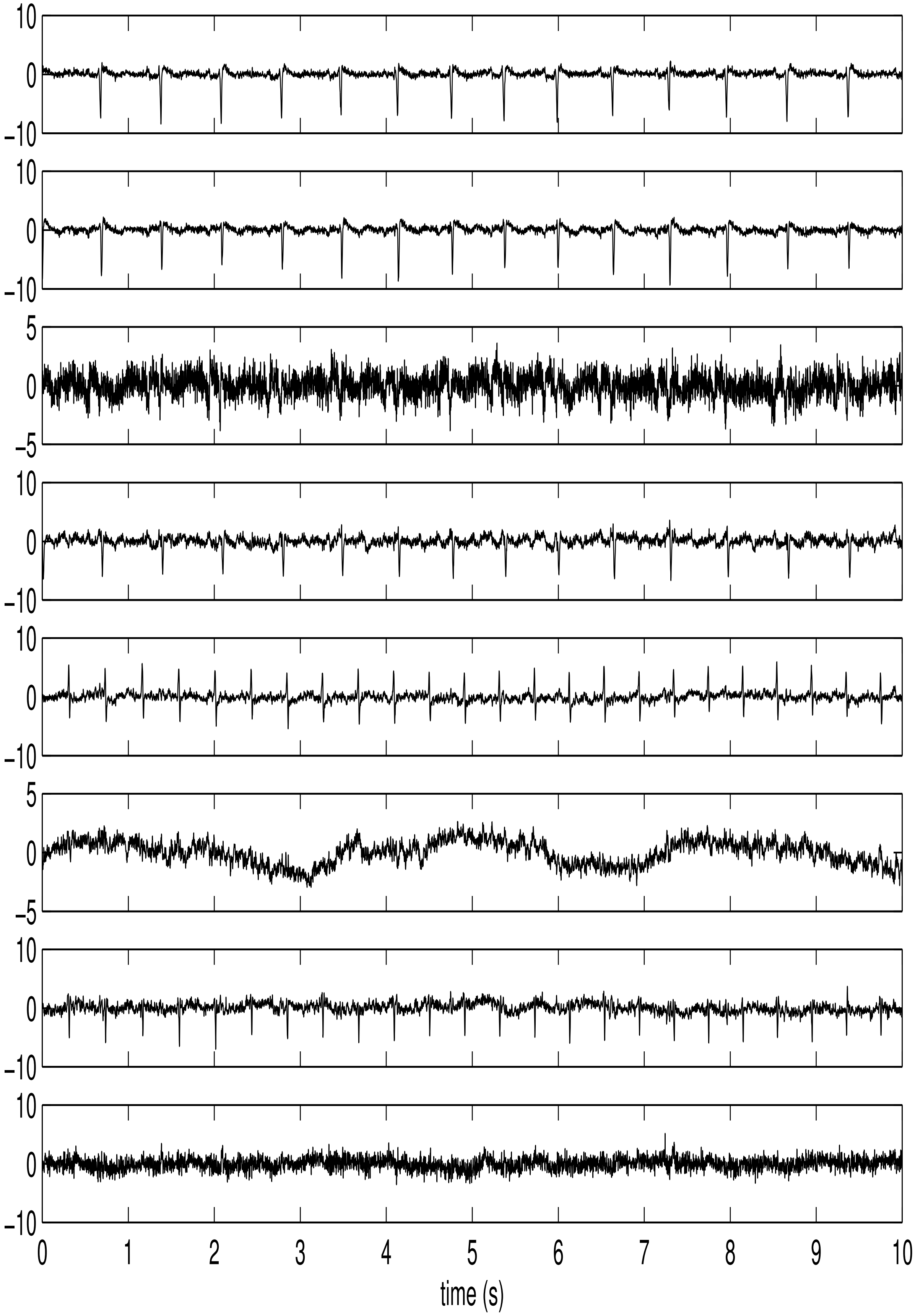

Figure 10: Source signals from the ICA-HOEVD method.

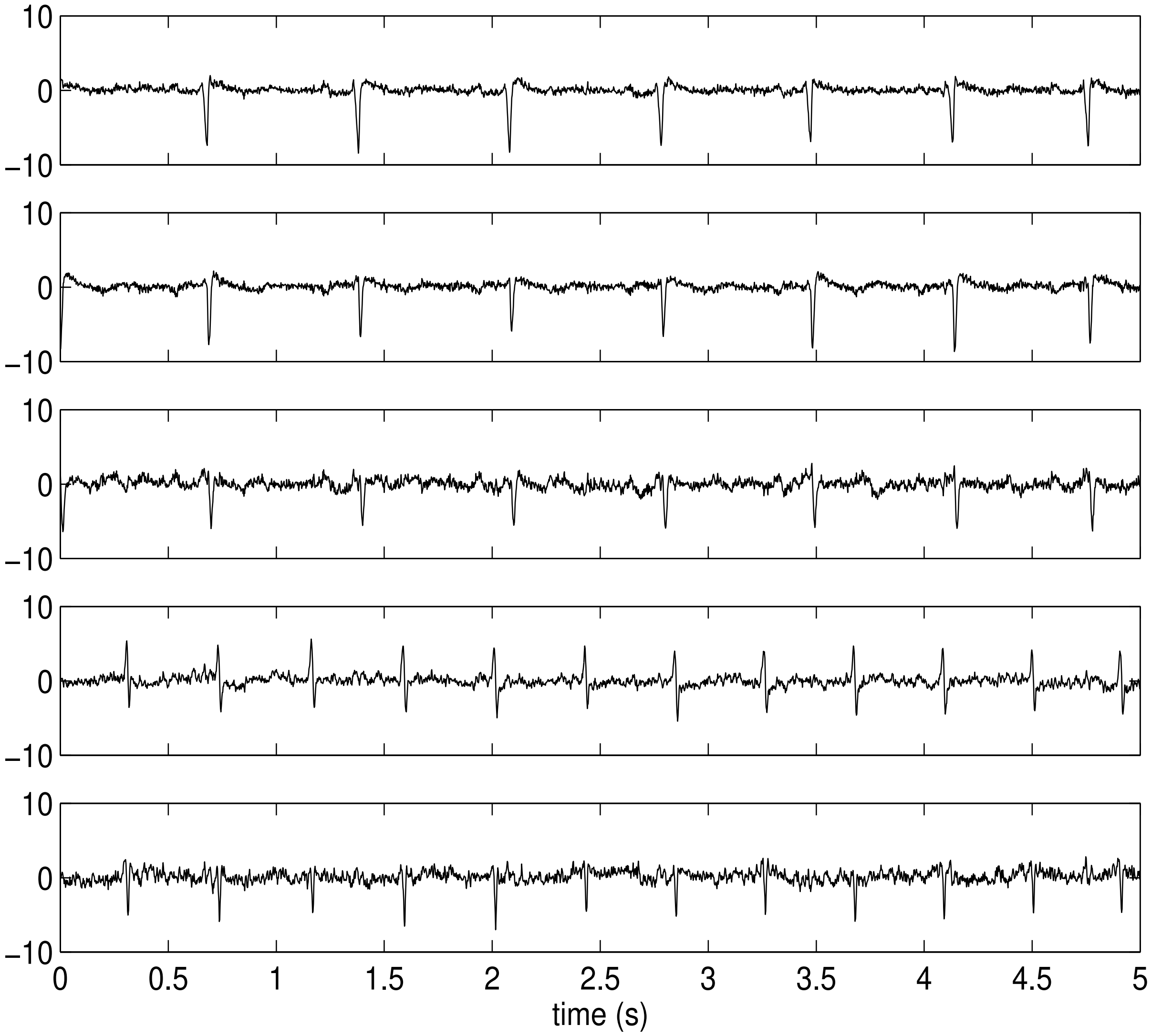

Figure 11: ECG sources from the ICA-HOEVD method.