View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by HAL-UNICE

Blind source separation of underdetermined mixtures of event-related sources

Mohammad Niknazar, Hanna Becker, Bertrand Rivet, Christian Jutten, Pierre Comon

To cite this version:

Mohammad Niknazar, Hanna Becker, Bertrand Rivet, Christian Jutten, Pierre Comon. Blind source separation of underdetermined mixtures of event-related sources. Signal Processing, Elsevier, 2014, 101, pp.52-64. <10.1016/j.sigpro.2014.01.031>. <hal-00952039>

HAL Id: hal-00952039

https://hal.archives-ouvertes.fr/hal-00952039

Submitted on 26 Feb 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Blind source separation of underdetermined mixtures of event-related sources

Mohammad Niknazar[a,*], Hanna Becker[a,b], Bertrand Rivet[a], Christian Jutten[a,c,**], and Pierre Comon[a]

[a] *GIPSA-lab (UMR CNRS-5216), University of Grenoble, Grenoble, France*
[b] *I3S (CNRS UMR-7271), Sophia Antipolis, France*
[c] *Institut Universitaire de France, Paris, France*

Abstract

This paper addresses the problem of blind source separation for underdetermined mixtures (i.e., more sources than sensors) of event-related sources that include quasi-periodic sources (e.g., electrocardiogram (ECG)), sources with synchronized trials (e.g., event-related potentials (ERP)), and amplitude-variant sources. The proposed method is based on two steps: (i) tensor decomposition for underdetermined source separation and (ii) signal extraction by Kalman filtering to recover the source dynamics. A tensor is constructed for each source by synchronizing on the "event" period of the corresponding signal and stacking different periods along the second dimension of the tensor. To cope with the interference from other sources that impede on the extraction of weak signals, two robust tensor decomposition methods are proposed and compared. Then, the state parameters used within a nonlinear dynamic model for the extraction of event-related sources from noisy mixtures are estimated from the loading matrices provided by the first step.

The influence of different parameters on the robustness to outliers of the proposed method is examined by numerical simulations. Applied to clinical electroencephalogram (EEG), ECG and magnetocardiogram (MCG), the proposed method exhibits a significantly higher performance in terms of expected signal shape than classical source separation methods such as πCA and FastICA.

Keywords: quasi-periodic source, event-related source, underdetermined mixtures, robust tensor decomposition, extended Kalman filtering, fetal ECG extraction.

1. Introduction

In this study, blind separation of underdetermined mixtures of event-related sources is addressed. An event-related source is characterized by typical patterns which are elicited after some events: such patterns may vary in amplitudes and/or in inter-event intervals (IEI). In this context, an event-related source is referred to as: (i) quasi-periodic source

*Corresponding author, email: mohammad.niknazar@gipsa-lab.grenoble-inp.fr
**Member of EURASIP

(e.g., electrocardiogram (ECG)) in which IEI and amplitudes can only slightly change from a period to another; (ii) source with synchronized stimuli (e.g., event-related potentials (ERP)) in which a pattern is repeated with no assumption on IEI but with quasi-constant amplitudes; (iii) amplitude-variant source whose amplitude (even sign) can largely change from a period to another but with quasi-constant IEI (e.g., telecommunication); (iv) general source without any assumptions on amplitudes and IEI, which can thus largely vary from an event to another one (e.g., digital communications).

In the recent years, a lot of attention has been paid to blind source separation (BSS) due to its wide-ranging applications in many areas [1] such as audio and speech processing [2], telecommunications [3], biomedical engineering [4], hyperspectral imaging [5], etc. Assuming an M-dimensional observation vector, $\mathbf{y}(k)$, this problem is mathematically expressed as:

$$\mathbf{y}(k) = \mathbf{A}\mathbf{x}(k) + \mathbf{b}(k), \tag{1}$$

where $\mathbf{x}(k)$ denotes the N-dimensional source vector, $\mathbf{b}(k)$ denotes the M-dimensional additive noise vector, and $\mathbf{A}$ is the $M \times N$ mixing matrix. The BSS framework aims at identifying the mixing matrix $\mathbf{A}$, or estimating the sources $\mathbf{x}(k)$, or both, from the observation $\mathbf{y}(k)$. Unlike the determined or overdetermined cases, when the number of sources is equal to or exceeds the number of mixtures ($N > M$), i.e., in the underdetermined case, the estimation of the mixing matrix $\mathbf{A}$ does not permit to directly recover the original sources. In fact, the mixing matrix does not admit a left inverse in that case, which makes it more difficult to recover the sources even if it is known and full rank [1, 6]. It is then necessary to rely on a prior on the sources.

Sparsity of the sources in a transformed domain is a possible prior to address underdetermined BSS [7]. Indeed, most of the proposed methods in the literature of underdetermined BSS are based on the sparsity of sources in a domain, (e.g., the frequency domain [8] or the time-frequency domain [9]). In this case, even if several sources are active at the same time so that the mixture is locally overdetermined, the mixing matrix can usually be estimated by clustering methods. However, this kind of search usually requires massive computations that limit the applicability of these methods to a smaller number of observation channels and sources [10].

Separation of underdetermined sources consists of two steps: estimation of the mixing matrix and extraction of the sources. Many algebraic and geometric (clustering) methods have been developed for the first step. They employ various decompositions of different data structures such as cumulant, correlation and cross-correlation matrices or tensors [1, 10]. Then, a second step is required for recovering the original sources.

Higher-order tensors have gained increasing importance as they can be used to represent higher order cumulants that are exploited in independent component analysis [11] and have been used successfully in BSS [12]. In addition, they are natural representations of multidimensional (higher than 2) data than matrices in many practical applications (e.g., in chemistry, biomedical engineering, and wireless communications). A fundamental challenge in these applications is to find informative and sparse representations of tensors, i.e., tensor decompositions. Tensor decompositions take into account information about different variables of the data, such as, for example, spatial, temporal and spectral information, and may provide links among the various extracted factors or latent variables with physical or physiological meaning and interpretation [13].

There are many applications, in which the sources are known to be event-related. These properties are observed in digital communication, speech and some physiological signals such as electrocardiograms. The behavior of second- and fourth-order BSS algorithms in a cyclostationary context has been studied in [14]. In a recent study [10], an underdetermined separation method has been developed, which is suitable for separation of signals that are piecewise stationary, having time-varying variances. These algorithms, which exploit the cyclostationarity property resort to statistical tools. In [15], a parallel deflation procedure based on a deterministic tensor decomposition has been proposed to address the problem of underdetermined BSS in the cyclostationary context. The basic approach consists in constructing a tensor by synchronizing on the symbol rate of a certain source, and decomposing the tensor using the Canonical Polyadic (CP) decomposition [16] to extract the characteristics of the source.

In this paper, the method described in [15] has been adapted for the estimation of the mixing matrix, temporal patterns, and amplitudes of event-related sources. The method described in [15] fails to extract a source which has very little power compared to the other sources because the latter act as interferers with high amplitudes that can be considered as outliers and impede on the accurate tensor decomposition. To overcome this problem, we propose to apply a robust tensor decomposition. In the literature, one can find several methods that have been developed to this end [17, 18]. In general, these techniques are based on a modification of the classical quadratic cost function that is optimized during the tensor decomposition. For example, the authors of [18] suggested to minimize the mean absolute error, which reduces the impact of outliers in the data, but does not prevent them from influencing the results since high outliers still lead to high errors. It is also possible to introduce weights that account for different uncertainties of the tensor elements (see, e.g., [17]). In this paper, we present two robust CP decomposition methods. The first one, which we subsequently refer to as Gaussian CP (GCP) decomposition, goes a step further compared to the approach taken in [18] and optimizes a cost function that limits the maximal error to 1. The second method exploits the particular structure of the data to compute weights that discriminate outliers and employs a weighted CP (WCP) decomposition.

As the second step of the separation of underdetermined sources, a nonlinear state-space model has been developed for extracting N quasi periodic sources (or components) from M observations. This model is used within a Kalman filtering framework, whose mixing matrix and state parameters are obtained from the loading matrices of the tensor decomposition.

The robustness of the proposed tensor decomposition methods to different parameters such as the initialization, the amount of outliers, the variability of the amplitudes, and synchronization errors is analyzed by means of numerical simulations. Furthermore, the proposed method is applied to biomedical data including electroencephalogram (EEG), ECG and magnetocardiogram (MCG) to extract desired sources. In the past, tensor-based techniques using classical CP decomposition have already been applied to space-time-realization EEG data (see [19, 20, 21]). However, the newly proposed method makes the decomposition more robust to noise, enabling the correct extraction of very weak signals of interest, which is not possible with classical CP decomposition methods. Furthermore, contrary to the previous tensor-based approaches, it permits to extract the original time courses of the signals, which are not identical for all realizations, using Kalman filtering. Among several methods in the literature for multichannel fetal ECG (fECG) extraction, one can name blind source separation [22], semi-blind source separation [23], adaptive filtering [24, 25], and periodic component analysis

(πCA) [26]. All these methods exploit the redundancy of the multichannel ECG recordings to reduce maternal ECG (mECG) and other interference sources. Nevertheless, even if this reduction has been successful, the exogenous noise cannot be totally canceled in this way [27]. Moreover, these methods demand several channels to recover weak traces of fetal signal. Conversely, the proposed method, which simultaneously extracts and denoises fECG signal, is applicable to as few as two channels.

The rest of the paper is organized as follows. Equations and theories supporting our proposed method are described in section 2. Results of the proposed method applied to various data and discussion about the results are presented in section 3. Finally, our conclusions are stated in section 4.

2. Methodology

The proposed method is based on two steps: (i) a robust tensor decomposition to estimate the mixing matrix and the state parameters of the Kalman filter (subsection 2.1) and (ii) a refined Kalman filtering (subsection 2.2) for the extraction of the signals.

2.1. Rough estimation of the mixing matrix and source parameters

2.1.1. Tensor construction and CP model

In the style of [15], we exploit the event-related nature of the signals of interest to construct a data tensor with dimensions space, event-synchronized window, and time from the 2-dimensional measurements for each of Q event-related sources. To this end, for the q-th source, we identify L_q event-synchronized windows of length T_q of the corresponding time signal. This can, for instance, be achieved based on a characteristic pattern within each event-synchronized window that can be recognized in the measurements. This pattern also serves as a reference point to synchronize the signals of different patterns of the event-related source, such as the maximum amplitude in the case of impulsive signals. For each of the L_q event-synchronized windows, one can thus extract a $M \times T_q$ data matrix from the measurements. These matrices are then stacked along the second dimension of the tensor $\boldsymbol{Y}^{(q)} \in \mathbb{R}^{M \times L_q \times T_q}$ (see Figure 1).

Assuming that the q-th source can be described by $R_q \in \mathbb{N}$ components that are identical for all event-synchronized windows except for changes of amplitude, the elements of the tensor can be written as

$$Y_{ijk}^{(q)} = \sum_{r=1}^{R_q} a_{ir}^{(q)} s_{jr}^{(q)} h_{kr}^{(q)} + b_{ijk}. \tag{2}$$

The first term in the right-hand side of (2) corresponds to the CP decomposition of a tensor where $a_{ir}^{(q)}$, $s_{jr}^{(q)}$, and $h_{kr}^{(q)}$ are the elements of three loading matrices $\mathbf{A}^{(q)} \in \mathbb{R}^{M \times R_q}$, $\mathbf{S}^{(q)} \in \mathbb{R}^{L_q \times R_q}$, and $\mathbf{H}^{(q)} \in \mathbb{R}^{T_q \times R_q}$, respectively [16]. The loading matrices correspond to the mixing matrix ($\mathbf{A}^{(q)}$), the matrix of pattern amplitudes ($\mathbf{S}^{(q)}$), and the matrix containing the temporal patterns of the event's components ($\mathbf{H}^{(q)}$) that characterize the mixture of the q-th source. The second term contains noise and interference from the desynchronized signals of other sources.

Since we assume that each event-related source may consist of more than one component, the number of sources N in the model (1) corresponds to the total number of components,

i.e., $N = \sum_{q=1}^{Q} R_q \geq Q$, while the full mixing matrix $\mathbf{A}$ in the model (1) can be obtained as $\mathbf{A} = [\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(Q)}]$.

In practice, one can obtain estimates for the mixing matrix, the pattern amplitudes, and the signal patterns by decomposing the tensor using the following criterion that optimizes the classical CP cost function:

$$\left\{\hat{\mathbf{A}}^{(q)}, \hat{\mathbf{S}}^{(q)}, \hat{\mathbf{H}}^{(q)}\right\} = \underset{\left\{\mathbf{A}^{(q)}, \mathbf{S}^{(q)}, \mathbf{H}^{(q)}\right\}}{\arg\min} \sum_{i,j,k} \left\| y_{ijk}^{(q)} - \sum_{r=1}^{R_q} a_{ir}^{(q)} s_{jr}^{(q)} h_{kr}^{(q)} \right\|_F^2 . \tag{3}$$

An important advantage of the CP decomposition in comparison to matrix decompositions, such as principal component analysis (PCA), is that it is essentially unique [28, 29] up to scale and permutation indeterminacies under mild conditions on the tensor rank, without imposing additional constraints such as orthogonality or independence. In [28, 29], the following sufficient condition for essential uniqueness has been derived:

$$k_{\mathbf{A}^{(q)}} + k_{\mathbf{H}^{(q)}} + k_{\mathbf{S}^{(q)}} \geq R_q + 2. \tag{4}$$

Here, $k_{\mathbf{A}^{(q)}}$, $k_{\mathbf{H}^{(q)}}$, and $k_{\mathbf{S}^{(q)}}$ denote the Kruskal ranks of the matrices $\mathbf{A}^{(q)}$, $\mathbf{H}^{(q)}$, and $\mathbf{S}^{(q)}$, respectively. In particular, as has been shown in [15], if $\mathbf{A}^{(q)}$, $\mathbf{S}^{(q)}$, and $\mathbf{H}^{(q)}$ have full rank and $T_q \geq R_q$, $L_q \geq R_q$ (i.e., if the number of events and the number of time samples per event are larger than the number of components R_q to be extracted), then $M = 2$ sensors are enough to blindly separate R_q components.

However, errors in the decomposition are to be expected due to noise and interfering sources b_{ijk}, in particular if the source to be extracted is weak compared to the interfering sources. In this case, the other source signals can be considered as outliers and strongly influence the optimization of criterion (3) because they prevent the decomposition algorithm to concentrate on the signal of interest. To overcome this practical problem, we propose to apply a tensor decomposition that is robust to outliers. In the following, we present two different robust decomposition schemes, which are based on modifications of the CP cost function. The different cost functions are displayed in Figure 2.

2.1.2. Robust tensor decomposition

Gaussian CP (GCP) decomposition. The idea of the first method consists in resorting to a cost function that does not attribute tremendous errors to outliers as does the classical quadratic cost function used in (3). In [18], the use of an L1-norm cost function was proposed. We go a step further and employ a cost function J_G which is based on Gaussian-like functions and limits the maximal error to 1 (see Figure 2) . This leads to the following optimization criterion:

$$\min_{\left\{\mathbf{A}^{(q)}, \mathbf{S}^{(q)}, \mathbf{H}^{(q)}\right\}} \sum_{i,j,k} \psi\left(y_{ijk}^{(q)} - \sum_{r=1}^{R_q} a_{ir}^{(q)} s_{jr}^{(q)} h_{kr}^{(q)} \right), \tag{5}$$

with $\psi(u) = 1 - \exp\{-\frac{u^2}{2\sigma^2}\}$. In this case, an error value of about 3σ between a tensor element and the reconstructed tensor element is treated as an outlier since its effective error value $\psi(u)$ is very close to the maximum value. The parameter σ that adjusts the width of the Gaussian function thus allows defining a threshold between "normal" errors and large outliers and has

a high influence on the results of the decomposition. For small σ, most of the errors will be treated as outliers, which makes the identification of the model difficult, while for large σ, outliers might not be recognized and can lead to biased estimates of the loading matrices. The optimal value for σ lies in between and should be chosen according to the data. If available, estimates of the variances of the q-th source to extract, of the noise, and of the other sources can be used to determine an appropriate value for the width of the Gaussian function.

The optimization of the cost function J_G can be accomplished using a gradient descent algorithm. Starting with initial estimates of the loading matrices $\mathbf{A}^{(0)}$, $\mathbf{S}^{(0)}$, and $\mathbf{H}^{(0)}$ (here the superscript denotes the iteration; for a better readability the superscripts referring to the sources are left out in the following), the loading matrices at iteration l are updated according to

$$\mathbf{A}^{(l+1)} = \mathbf{A}^{(l)} - \mu_{\mathbf{A}}\, \mathbf{g}_{\mathbf{A}}, \tag{6}$$

$$\mathbf{S}^{(l+1)} = \mathbf{S}^{(l)} - \mu_{\mathbf{S}}\, \mathbf{g}_{\mathbf{S}}, \tag{7}$$

$$\mathbf{H}^{(l+1)} = \mathbf{H}^{(l)} - \mu_{\mathbf{H}}\, \mathbf{g}_{\mathbf{H}}, \tag{8}$$

until convergence. Here, $\mu_{\mathbf{A}} = \mu_{\mathbf{S}} = \mu_{\mathbf{H}} = \mu$ is a stepsize parameter and $\mathbf{g}_{\mathbf{A}} = \frac{\partial J_G}{\partial \mathbf{A}}$, $\mathbf{g}_{\mathbf{H}} = \frac{\partial J_G}{\partial \mathbf{H}}$, and $\mathbf{g}_{\mathbf{S}} = \frac{\partial J_G}{\partial \mathbf{S}}$ denote the gradients of the cost function J_G with respect to the three loading matrices, respectively. The r-th column of the matrix $\mathbf{g}_{\mathbf{A}} = \frac{\partial \mathbf{J_G}}{\partial \mathbf{A}}$ can be computed as

$$\mathbf{g}_{\mathbf{a}_r} = -[(\mathbf{h}_r \otimes \mathbf{s}_r) \otimes \mathbf{I}_M]^{\mathrm{T}} \left[\frac{\mathbf{e}_1}{\sigma} \boxdot \exp\left\{ \frac{\mathbf{e}_1 \boxdot \mathbf{e}_1}{2\sigma^2} \right\} \right], \tag{9}$$

with $\mathbf{e}_1 = \mathrm{vec}\left\{[\boldsymbol{Y}]_{(1)}\right\} - [(\mathbf{H} \odot \mathbf{S}) \otimes \mathbf{I}_M]\mathrm{vec}\{\mathbf{A}\}$. Here,

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1R}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{M1}\mathbf{B} & \cdots & a_{MR}\mathbf{B} \end{bmatrix}$$

denotes the Kronecker product of matrices $\mathbf{A} \in \mathbb{R}^{M \times R}$ and $\mathbf{B}$, $\odot$ denotes the Khatri-Rao column-wise Kronecker product, $\boxdot$ is the Hadamard element-wise product, $\mathrm{vec}\{\mathbf{A}\}$ is the vector obtained by concatenating the columns of the matrix $\mathbf{A}$, and $\mathbf{I}_M$ is the identity matrix of size $M \times M$. Furthermore, $[\boldsymbol{Y}]_{(1)} \in \mathbb{R}^{M \times (L_q T_q)}$ denotes the first unfolding matrix of the tensor $\boldsymbol{Y}$, which is composed of all mode-1 vectors of the tensor (a mode-1 vector is obtained by fixing the second and third index of the tensor elements and varying the first index from 1 to M) with an ordering such that the second index is varied faster than the third one (for more details and illustrations of operations on tensors see e.g., [30]). The other two gradient matrices are determined in an analogous way.

Since it does not require any information about the data except for an estimate for the threshold between noise and outliers, this robust decomposition method is rather general. It is therefore applicable to a large range of applications. However, to obtain accurate results, a good initialization is required (see Section 3.1.1). This is especially mandatory if the difference between the values of outliers and the values of the signal of interest is in the same range as the amplitude of the signal of interest.

Weighted CP (WCP) decomposition. The second proposed robust tensor decomposition method is based on the weighted cost function (see the red curve in Figure 2):

$$\min_{\{\mathbf{A}^{(q)},\mathbf{S}^{(q)},\mathbf{H}^{(q)}\}} \sum_{i,j,k} \left\| w_{ijk}^{(q)} \left(y_{ijk}^{(q)} - \sum_{r=1}^{R_q} a_{ir}^{(q)} s_{jr}^{(q)} h_{kr}^{(q)} \right) \right\|_F^2, \tag{10}$$

and exploits the particular structure of the data at hand to determine suitable weights $w_{ijk}^{(q)}$. In fact, for applications with a small variability of the amplitudes, one can exploit the desynchronization of noise and interference which manifests itself by a high variance σ_{ij}^2 over different event-synchronized windows compared to the signal of interest to identify the outliers. The influence of the outliers can then be attenuated by the attribution of low weights while weights close to 1 are assigned to the tensor elements that exhibit a small variance over event-synchronized windows. The weights are thus computed depending on the variance σ_{ij}^2 as

$$w_{ijk}^{(q)} = \exp\left\{ -\frac{(y_{ijk}^{(q)} - \mu_{ij})^2}{\sigma_{ij}^2} \right\}, \quad q = 1, \ldots, Q, \tag{11}$$

where μ_{ij} is the mean of the tensor elements over all event-synchronized windows, and can be stored into a nonnegative weight tensor, which is of the same dimensions as $\boldsymbol{Y}^{(q)}$. In order to obtain robust estimates for the variances σ_{ij}^2, we use the median absolute deviation (MAD) estimator [31] for their determination.

The optimization of (10) can, for example, be performed using a weighted Alternating Least Squares algorithm (see [17]).

As for ECG, MCG and EEG signals, the amplitudes for different event-synchronized windows are approximately the same, the WCP decomposition is especially adapted to these applications. Contrary to the GCP decomposition, which requires the manual selection of the parameter σ, all the parameters are determined automatically from the data and the technique is robust to intialization. Please note, though, that in the general case, the pattern amplitudes may change considerably from one event-synchronized window to another, which prevents the accurate estimation of the weights using the method described above.

The robust tensor decomposition methods subsequently described can be used to estimate the mixing matrix and to extract the temporal patterns and the amplitudes of the event-related sources. This is already an advantage over a matrix decomposition using the singular value decomposition (SVD), where one only obtains an estimate of the subspace spanned by the mixing matrix and no information about individual pattern amplitudes, because the temporal structure of the event-related sources is not exploited. However, as the tensor decomposition model assumes identical temporal patterns for each event, the dynamics of the sources, i.e., slight variations from one period to another, are lost. To recover these dynamics, we propose to use the estimated mixing matrix, temporal patterns and amplitudes in a Kalman filtering framework to obtain refined estimates of the sources.

2.2. EKF Framework for Extraction of Event-Related Sources

The goal of Kalman Filter (KF) is to estimate the state of a discrete-time controlled process. Consider a state vector $\mathbf{x}_{k+1}$ governed by a nonlinear stochastic difference equation

with measurement vector $\mathbf{y}_{k+1}$ at time instant $k+1$:

$$\begin{cases} \mathbf{x}_{k+1} = f(\mathbf{x}_k, \mathbf{w}_k, k+1) \\ \mathbf{y}_{k+1} = h(\mathbf{x}_{k+1}, \mathbf{v}_{k+1}, k+1), \end{cases} \tag{12}$$

where the random variables $\mathbf{w}_k$ and $\mathbf{v}_k$ represent the process and measurement noises, respectively, with associated covariance matrices $Q_k = E\left\{\mathbf{w}_k\mathbf{w}_k^T\right\}$ and $R_k = E\left\{\mathbf{v}_k\mathbf{v}_k^T\right\}$. The Extended Kalman Filter (EKF) is an extension of the standard KF to nonlinear systems $f(\cdot)$ and $h(\cdot)$, which linearizes about the current mean and covariance [32]. In order to improve the estimations, EKF can be followed by a backward recursive smoothing stage leading to the Extended Kalman Smoother (EKS). However, since EKS is a non causal method, it cannot be applied online, but can be used if a small lag in the processing is allowed.

In this framework, the state vector associated with each event-related source is defined by its phase θ_k and amplitude z_k. The phase θ_k is, in fact, a means of modeling the event-related behavior of the source. Then, each period of the source is modeled using θ_k to obtain z_k. Thereby, the state model of the event-related source, in its discrete form with a small sampling period δ, is:

$$\begin{cases} \theta_{k+1} = (\theta_k + \omega\delta)\mathrm{mod}(2\pi) \\ z_{k+1} = g(\theta_k, k) + z_k + \eta_k, \end{cases} \tag{13}$$

where θ and z are the state variables in polar coordinates and k denotes the discrete time index. ω is the phase increment, η_k is a random additive noise, and $g(.)$ models the temporal pattern and amplitude of the source. In addition to the noisy recording of the source, s_k, an observed phase ϕ_k is obtained by a linear time warping of each event interval into $[0, 2\pi)$ (see Figure 3), leading to the following system:

$$\begin{bmatrix} \phi_k \\ s_k \end{bmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{bmatrix} \theta_k \\ z_k \end{bmatrix} + \begin{bmatrix} u_k \\ v_k \end{bmatrix}, \tag{14}$$

where u_k and v_k are the corresponding observation noises with zero-mean random variable entries.

With several event-related sources in multichannel recordings, redundancy of each event-related source can be exploited to estimate the information of the desired source mixed with the other sources and background noise. In order to do so, a linear transform is assumed to decompose M mixed event-related signals into N components. In other words, we assume that all event-related sources have N components in total, which are observed in M signals. For N mixed components, the dynamic equations may be written as:

$$\begin{cases} \theta_{k+1}^{(1)} = (\theta_k^{(1)} + \omega^{(1)}\delta)\mathrm{mod}(2\pi) \\ z_{k+1}^{(1)} = g^{(1)}(\theta_k^{(1)}, k) + z_k^{(1)} + \eta_k^{(1)} \\ \vdots \\ \theta_{k+1}^{(N)} = (\theta_k^{(N)} + \omega^{(N)}\delta)\mathrm{mod}(2\pi) \\ z_{k+1}^{(N)} = g^{(N)}(\theta_k^{(N)}, k) + z_k^{(N)} + \eta_k^{(N)}. \end{cases} \tag{15}$$

The phase observations of the N components, $\mathbf{\Phi} = [\phi^{(1)},...,\phi^{(N)}]^T$, and the M noisy mixtures of the N components, $\mathbf{s} = [s^{(1)},...,s^{(M)}]^T$, are related to the state vectors $\mathbf{\Theta} = [\theta^{(1)},...,\theta^{(N)}]^T$ and $\mathbf{z} = [z^{(1)},...,z^{(N)}]^T$ at time k as follows:

$$\begin{bmatrix} \mathbf{\Phi}_k \\ \mathbf{s}_k \end{bmatrix} = \begin{pmatrix} I & 0 \\ 0 & A \end{pmatrix} \begin{bmatrix} \mathbf{\Theta}_k \\ \mathbf{z}_k \end{bmatrix} + \begin{bmatrix} \mathbf{u}_k \\ \mathbf{v}_k \end{bmatrix}, \tag{16}$$

where $\mathbf{u}_k$ and $\mathbf{v}_k$ are the corresponding observation noises.

The key step prior to the implementation of the filter is the estimation of $g^{(n)}(.)$ for the n-th component as well as the mixing matrix A:

$$A = \begin{bmatrix} a_{11} & \dots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{M1} & \dots & a_{MN} \end{bmatrix}. \tag{17}$$

In order to do so, the loading matrices provided by the previous step (Section 2.1) are used:

- the mixing matrix is directly defined as the concatenation of the loading matrices $\mathbf{A}^{(n)}$ related to all the event-related components;
- the temporal pattern of $g^{(n)}(.)$ for the n-th component is provided by the loading matrix $\mathbf{H}^{(n)}$;
- the amplitude of $g^{(n)}(.)$ for the n-th component at each period is obtained by the loading matrix $\mathbf{S}^{(n)}$ and by assuming that $\eta_k^{(1)}$,..,$\eta_k^{(N)}$ are uncorrelated, a ratio (e.g. 0.1) of standard deviation of $\mathbf{S}^{(n)}$ can be used as the estimate of $\eta_k^{(n)}$ for initializing the state covariance matrix Q_k.

3. Results

Both synthetic and clinical data have been used to study the performance of the proposed method. In the first subsection, quantitative results coming from simulations have been studied. In this subsection, the robustness of the proposed tensor decomposition methods to amplitude variations, quantity of outliers, initialization, and synchronization errors is examined based on simulations for both arbitrary tensors and tensors constructed from data of event-related sources.

In the second subsection the results of the proposed method on three sets of clinical data including EEG, ECG, and MCG have been presented. The results on ECG and MCG data have been compared with the results of πCA [26] and FastICA [33] methods. The GCP and WCP labels denote results of the first and second proposed approaches for tensor decomposition without Kalman filtering stage. GCP+EKS and WCP+EKS show the results based on the first and second proposed approaches of tensor decomposition with a Kalman filtering stage.

3.1. Synthetic data

3.1.1. Robust decomposition of arbitrary trilinear tensors in the presence of outliers

First of all, we analyzed the performance of the proposed robust tensor decomposition methods for arbitrary tensors containing outliers. To this end, we generated a set of arbitrary loading matrices $\mathbf{A} \in \mathbb{R}^{5\times 2}$, $\mathbf{S} \in \mathbb{R}^{10\times 2}$, and $\mathbf{H} \in \mathbb{R}^{10\times 2}$. The elements of $\mathbf{A}$ and $\mathbf{H}$ were random variables chosen from a uniform distribution between -1 and 1. The elements of $\mathbf{S}$

were chosen from a Gaussian distribution with mean 1 and variance $\sigma_S^2 = 0.04$. The use of other continuous distributions leads to comparable results. The tensor was then constructed from the matrices $\mathbf{A}$, $\mathbf{S}$ and $\mathbf{H}$ according to (2). To simulate outliers, we falsified a certain percentage p of randomly selected tensor values by adding or subtracting 2. The value 2 has been chosen arbitrarily such that it is large enough to be considered as an outlier, but close enough to the values of other tensor elements to remain realistic. Furthermore, we added white Gaussian noise according to a signal-to-noise ratio (SNR) of 20dB. The resulting tensor was then decomposed using the CP, WCP and GCP decompositions. For initialization, we took the original loading matrices and added zero-mean Gaussian noise with variance σ_i^2. We computed each decomposition three times for different initializations and retained the estimated loading matrices that were obtained for the minimal value of the cost function. The accuracy of the estimated loading matrices $\hat{\mathbf{A}}$, $\hat{\mathbf{S}}$, and $\hat{\mathbf{H}}$ was evaluated using the following measure:

$$E_{\mathrm{LM}} = \frac{1}{3R} \min_{\mathbf{P}} \{ ||\mathbf{A}' - \hat{\mathbf{A}}'\mathbf{D_A}\mathbf{P}||_{\mathrm{F}} + ||\mathbf{S}' - \hat{\mathbf{S}}'\mathbf{D_S}\mathbf{P}||_{\mathrm{F}} + ||\mathbf{H}' - \hat{\mathbf{H}}'\mathbf{D_H}\mathbf{P}||_{\mathrm{F}} \}, \tag{18}$$

where R denotes the number of components, $\mathbf{P}$ is a permutation matrix and the matrices $\mathbf{A}'$, $\mathbf{S}'$, $\mathbf{H}'$, $\hat{\mathbf{A}}'$, $\hat{\mathbf{S}}'$, and $\hat{\mathbf{H}}'$ correspond to the original and estimated loading matrices normalized to unit column norm. Furthermore, $\mathbf{D_A}$, $\mathbf{D_S}$, and $\mathbf{D_H}$ are diagonal matrices whose elements correspond the signs of the elements on the diagonal of the correlation matrices $(\mathbf{A}')^{\mathrm{T}}\hat{\mathbf{A}}'$, $(\mathbf{S}')^{\mathrm{T}}\hat{\mathbf{S}}'$, and $(\mathbf{H}')^{\mathrm{T}}\hat{\mathbf{H}}'$ and which are introduced to compensate for the sign ambiguity. Subsequently, we analyzed the influence of different parameters on this error for 100 Monte Carlo trials.

Influence of the percentage of outliers. Figure 4 shows the error of the loading matrices as a function of the percentage of outliers p for $\sigma_i = 0.5$. It can be seen that the error of the GCP decomposition is very small (< 0.1) over a large range of values of p ranging from 1 to 50% before the high percentage of outliers leads to an increasing error. This means that this method is very robust to even a large amount of randomly distributed outliers in the tensor. With an error that is slightly smaller than that of the GCP decomposition and which has a smaller variance, the WCP decomposition achieves a slightly better performance for small percentages of outliers ($p < 10\%$). Yet for $p > 10\%$, the error increases significantly with p, showing that the WCP is only robust to a limited number of outliers. This sensitivity of the WCP decomposition method is due to the estimation of the weights, which is not robust to a large number of outliers. However, for all examined percentages of outliers, both WCP and GCP methods show a much better performance than the classical CP decomposition, which is not robust to outliers and already exhibits a large error of 0.2 for only 1% of outliers.

In order to statistically compare the obtained results, statistical analysis was carried out for each tested value of p on Figure 4 to determine whether the distributions of the results of different methods are two-by-two significantly different. The one-way analysis of variance (ANOVA) was performed on the results provided by trials. As it is seen in Figure 4, for $p = 15\%$, WCP and GCP provide very close results. The p_value for this point has been obtained more than 0.05, which implies no significant difference between the performance of WCP and GCP methods. Nevertheless, for all the other points, p_value has been obtained less than 0.05, which means that the results of different methods are two-by-two significantly different.

Influence of initialization. To analyze the influence of the initialization on the decomposition results, we varied the variance σ_i^2 of the Gaussian noise that is added to the original loading matrices to obtain a set of initial loading matrices. The percentage of outliers was fixed to 10%. Figure 5 shows that the results of the WCP decomposition and the classical CP decomposition hardly change for different initializations, with an average error increasing only slightly from 0.04 to 0.1 for the WCP decomposition and from 0.37 to 0.4 for the CP decomposition. With 80% of the errors lying between 0.23 and 0.6, the variance of the error is high for the CP decomposition and is independent of the initialization. On the other hand, the variance of the error of the WCP decomposition increases with σ_i, attaining very small values for $\sigma_i < 0.5$ and moderate values for bad initializations. On the whole, the WCP and CP decompositions can be said to be robust to initialization. On the contrary, the GCP decomposition features a strong increase of the error for rising σ_i, in particular between $\sigma_i = 1$ and $\sigma_i = 3$. In this interval, the mean error increases from 0.1 to 0.7. For small $\sigma_i < 1$, i.e., if the initial loading matrices are close to the exact loading matrices, the GCP decomposition yields good results with a very small variance, outperforming the WCP decompositon when σ_i becomes very small, while for $\sigma_i > 2$, its results are even worse than those of the classical CP decomposition. As the Gaussian cost function assumes values close to 1 for all errors exceeding a certain threshold (see Section 2.1.2), the optimization criterion of the GCP decomposition can be expected to exhibit an increased number of local minima. If the initialization is not close to the true solution, the GCP decomposition algorithm is prone to find a local minimum. This explains the strong dependence of the results of the GCP decomposition on the initialization.

In Figure 5, all the results of different methods are statistically two-by-two different. The closest results are for WCP and GCP methods at $\sigma_i = 0.5$. Nevertheless, even for this point p_value has been obtained equal to 0.0048 that is less than 0.05.

3.1.2. Robust decomposition of tensors of event-related data

In the present paper, we focus on event-related sources. Therefore, we consider in the following tensors constructed from data containing a mixture of two event-related sources with one component per source. The elements of the mixing matrices $\mathbf{A}^{(q)} \in \mathbb{R}^{5\times 1}$, $q = 1, 2$, were chosen from a uniform distribution between -1 and 1. To introduce some amplitude variability, the amplitudes $\mathbf{S}^{(q)}$ were chosen from a continuous distribution. Here, we used a Gaussian distribution with mean 1 and variance σ_S^2. Different choices of distributions lead to similar results. The temporal patterns of the first and second source were given by $\exp\{-\frac{(x_1-0.5T_1)^2}{2\alpha_1^2}\}$ with $\alpha_1 = 0.1$ and $\exp\{-\frac{(x_2-0.5T_2)^2}{2\alpha_2^2}\}$ with $\alpha_2 = 0.15$ where x_1 and x_2 are the indices of the time samples and $T_1 = 101$ and $T_2 = 59$ correspond to the lengths of the patterns of the first and second source, respectively. On the whole, we considered 2020 time samples of data, containing 20 periods of the first source and 34 periods of the second source, in the presence of white Gaussian noise for a SNR of 20dB. The tensors were then constructed as described in Section 2.1.1 for 100 different trials. For the decomposition, the loading matrices were initialized as described in the previous section with $\sigma_i = 0.2$.

Influence of the amplitude variability. The objective of the first simulation consisted in evaluating the influence of the variance of the amplitudes $\mathbf{S}^{(q)}$. To this end, σ_S was varied between 0.1 and 4 for both sources such that the variances of their amplitudes were equal. Figure 6 shows the error of the loading matrices of the two tensors as a function of the standard

deviation σ_S of the amplitudes. For the classical CP and GCP decompositions, the errors of the loading matrices decrease with increasing σ_S. This can be explained by the fact that for small σ_S, the tensors are close to the degenerate case, while a higher amplitude variability facilitates the tensor decomposition and the separation of different components. For the WCP decomposition, the best performance is achieved for small σ_S. In this case, the amplitude variation is small enough so that the peaks of the second source in the tensor of the first source and vice versa are outlying values and do not influence the computation of the weights. However, for increasing σ_S, the amplitudes become high enough for the peaks of the interfering sources to lie within the range of amplitudes that are attained by the source to be extracted. In this case, the interfering sources do not enter as outliers in the computation of the weights, which will corrupt the estimation of the weights. This leads to an increase of the errors of the loading matrices with rising σ_S.

In Figure 6, all the results are statistically two-by-two different, because all the p_values have been obtained less than 0.05 for both tensors.

Influence of synchronization errors. To analyze the influence of synchronization errors (after the tensor construction, which was accomplished with perfect synchronization), we introduced artificial delays for each event period to model synchronization errors. The delays were uniformly distributed between $-\tau_{max}$ and τ_{max}, with τ_{max} chosen between 0 and $T_q\alpha_q$ samples, where α_q determines the width of the Gaussian function that characterizes the temporal pattern of the q-th source, $q = 1, 2$. The standard deviation of the amplitudes was fixed to $\sigma_S = 0.2$. Figure 7 shows that the error of the loading matrices increases rapidly with rising maximal delay. This means that the good synchronization of the signals in each tensor is crucial for the accuracy of the proposed method.

There are many points in Figure 7, in which the results obtained by different methods are not statistically two-by-two different, because the variance is high over the plotted mean values (see the 10% and 90% quantiles). Therefore, many of p_values have been obtained more than 0.05 for both tensors.

3.2. Clinical Data

3.2.1. ERP Extraction

Event-related potentials (ERPs) are the responses to brain stimulation measured by the scalp EEG. The measured responses are induced by multiple brain generators active in association with the eliciting event. However, they are mixed with background activity of the brain that is not related to the stimulus and also other interferences from non-neural sources, such as eye blinks [34] and muscle artefacts. Due to the much lower power of ERPs compared with background EEG, it is difficult to estimate them even though they are dominant in lower frequencies. The most common way to extract ERPs involves averaging time-locked sections of the EEG signal over many trials. This method assumes a simple model for ERPs that consists of the sum of an invariant signal and a random process that will be attenuated by averaging over trials [35]. However, there is evidence that ERP waves may vary considerably over time [36]. Furthermore, in [37], it has been shown in the context of neonatal seizure activity that tensor-based methods that exploit the repetitive nature of EEG signals exhibit an improved performance to time-locked averaging.

The ERP database consists of EEG signals recorded during a P300 speller brain-computer interface (BCI) experiment [38]. In such a BCI, the paradigm, which consists of visual stimuli divided between target and non-target stimuli, suggests that a positive deviation about 300ms is elicited after a target stimulus, while non-target stimuli do not elicit specific brain response. The EEG signals have been sampled at 1200Hz using 16 scalps electrodes. First, a three-way tensor is built by stacking the data of sixteen channels windowed from 1 second before the stimuli and 2 seconds after the stimuli. In this experiment, The ERP data was considered to be composed of 2 components. Decomposition of this tensor yields estimates of the temporal patterns of ERPs along with their amplitudes and the mixing matrix (i.e., the spatial projection on scalp). Then, the ERP estimates are improved by the proposed KF to preserve dynamics of ERPs over time. The extracted temporal patterns of ERPs via WCP, and single-trial estimates of WCP and WCP+EKS from the namely S6 dataset are shown in Figure 8. As it is seen, WCP+EKS outperforms WCP because it better preserves the dynamics of ERPs which can change from a trial to another. GCP and GCP+EKS also provide rather equivalent results to those of WCP and WCP+EKS.

3.2.2. Fetal ECG Extraction

The ECG data used in this subsection is the DaISy fetal ECG database [39], which consists of a single dataset of cutaneous potential recordings of a pregnant woman. A total of 8 channels (5 abdominal and 3 thoracic) are available, sampled at 250 Hz and lasting 10 seconds. The extracted maternal ECG and fetal ECG using the first and second channels of this database is shown in Figure 9. The mECG data was considered to be composed of 2 components, while we used only one component for the fECG data. The mixture of the first channel and extracted mECG and fECG signals using πCA, FastICA and the proposed GCP and GCP+EKS are plotted, respectively. The results of WCP and WCP+EKS are also similar to those of GCP and GCP+EKS. As it is seen, πCA and FastICA methods fail to extract fECG when only two electrodes are available, since the mixtures are underdetermined. There is neither ground truth nor golden standard on clinical fetal ECG recordings to be used as a reference for comparing the performance of the different methods. Nevertheless, in order to quantify the performance of each method on clinical data, the mean values of the contaminating and desired ECGs have been measured at their R-peak positions in the estimated ECG. This can provide an estimate for the residual of the contaminating mECG in the estimated fECG. If the contaminating mECG has been successfully canceled, the values of this measure should be low, meanwhile, the values of the estimated fECG at its R-peak positions should be close to values of the corresponding points in the original mixture. Table 1 reports values of this measure on the fECG estimated by the different methods. The results show that the proposed GCP, WCP, GCP+EKS, and WCP+EKS significantly outperform πCA and FastICA. Although GCP (or WCP) and GCP+EKS (or WCP+EKS) provided close quantitative results, it should be noted that valuable inter-beat dynamics of mECG and fECG are lost in the GCP (or WCP) estimate, because as it was explained in the previous section all beats of the reconstructed ECGs have exactly the same temporal pattern up to their amplitudes. Nonetheless, these valuable inter-beat dynamics of ECG signals are recovered using GCP+EKS (or WCP+EKS).

3.2.3. Twin MCG Extraction

In this subsection, twin fetal cardiac magnetic signals are extracted. The dataset used in this subsection consists of three sets of twin magnetocardiogram signals and other signals,

in arrays of 208 channels recorded over 30 minutes, with a sampling rate of 1025 Hz[1]. Two electrodes were used in this test. The presented results have been achieved for a typical couple of channels (indexed 92 and 116) of one of the available datasets, namely the q00002252 dataset.

To extract sources by the proposed tensor decomposition, they must have different pattern rates. As long as two sources are not exactly synchronous, they can be separated even if their pattern rates are approximately the same. This enables the method to separate twin cardiac signals even if heart rates are approximately equal. However, in this case, the interference of the second source in the tensor of the first source might be concentrated within a certain interval of the temporal pattern. This happens especially if the tensor is built from a small number of events. In this case, the correct decomposition of the tensor is particularly difficult and cannot be achieved by the classical CP decomposition. The introduction of the weights in the WCP method permits us to overcome this problem and to focus on the signal of interest. This discrimination is also provided by the Gaussian function in the first approach (GCP).

There are three sources to be extracted, one maternal MCG and two fetal MCGs, while two channels are to be utilized. The maternal and fetal MCG ranks considered in the proposed method are 2 and 1, respectively. Fig.10 presents the results of πCA and the proposed WCP and WCP+EKS methods in extraction of the maternal and two fetal MCG signals from two channels. GCP and GCP+EKS also provide close results to those of WCP and WCP+EKS. As it is seen, πCA method fails to track periodic patterns related to the fetal components due to their low power and insufficient number of the utilized electrodes. Nevertheless, the proposed WCP and WCP+EKS methods could recover weak traces of fetal MCG features. However, as expected in comparison with WCP+EKS, WCP method does not completely cancel the exogenous noise. Moreover, inter-beat dynamics of MCG signals are lost in this way. The maternal and fetal R-peak values on the first fetal MCG estimate, are presented in Table 2. In this experiment, a perfect estimate should give very low value at maternal R-peak and second fetal R-peak positions.

4. Conclusions

In this paper, we presented a novel two-step method based on robust tensor decomposition and extended Kalman filtering for extraction of event-related sources that can be quasi-periodic sources, sources with synchronized trials, and amplitude-variant sources. Two robust criteria, implemented in the GCP and WCP decompositions, were used in this work to capture the desired event-related sources even if their powers are much lower compared with other sources (e.g., large impulse concurrent sources). A comparative study of these criteria showed that while the GCP decomposition can handle a larger amount of outliers (interference of other sources) compared to the WCP decomposition, it also requires a good initialisation, which is not necessary for the WCP decomposition. For a limited number of outliers and a good initialisation, which can, for example, be provided by prior knowledge of the expected signal form, the WCP and GCP decompositions feature a similar performance. In a second step, the estimations are improved by an extended state Kalman filtering to avoid losing

[1]This dataset has been provided by Dr. Dirk Hoyer, from the Biomagnetic Center of the Department of Neurology, in Friedrich Schiller University, Jena, Germany.

dynamics of the sources.
This method is also applicable to underdetermined mixtures and this is its main interest. It allows us to utilize the minimal number of electrodes (down to two electrodes), if needed. The latter is a crucial feature for a monitoring system because it can affect the system's price, convenience and portability. Although good synchronization of events of the desired source in its tensor format is crucial for the functioning of the proposed method, application on clinical data shows its significantly superior performance in comparison to the classical multichannel methods.
Perspectives include the development of more robust tensor decomposition methods that can, for example, handle synchronization errors, and application to other datasets.

Acknowledgement

This work is partly supported by ERC AdG-2012-320864- CHESS. H. Becker has been supported by the Conseil Régional PACA and by CNRS. P. Comon was supported by the European Research Council under the European Communitys Seventh Framework Programme FP7/2007-2013 under Grant Agreement 320594.

References

[1] P. Comon, C. Jutten, Handbook of Blind Source Separation: Independent Component Analysis and Applications, Elsevier Science, 2010.

[2] A. Ozerov, C. Févotte, Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation, Audio, Speech, and Language Processing, IEEE Transactions on 18 (3) (2010) 550–563.

[3] A. Y. Kibangou, G. Favier, Blind equalization of nonlinear channels using a tensor decomposition with code/space/time diversities, Signal Processing 89 (2) (2009) 133–143.

[4] H. Lee, Y.-D. Kim, A. Cichocki, S. Choi, Nonnegative tensor factorization for continuous eeg classification, International journal of neural systems 17 (04) (2007) 305–317.

[5] D. Letexier, S. Bourennane, J. Talon, Nonorthogonal tensor matricization for hyperspectral image filtering, Geoscience and Remote Sensing Letters, IEEE 5 (1) (2008) 3–7.

[6] P. Comon, Blind identification and source separation in 2x3 under-determined mixtures, Signal Processing, IEEE Transactions on 52 (1) (2004) 11 – 22.

[7] R. Gribonval, S. Lesage, et al., A survey of sparse component analysis for blind source separation: principles, perspectives, and new challenges, in: ESANN'06 proceedings-14th European Symposium on Artificial Neural Networks, 2006, pp. 323–330.

[8] Z. He, S. Xie, S. Ding, A. Cichocki, Convolutive blind source separation in the frequency domain based on sparse representation, Audio, Speech, and Language Processing, IEEE Transactions on 15 (5) (2007) 1551 –1563.

[9] P. Georgiev, F. Theis, A. Cichocki, Sparse component analysis and blind source separation of underdetermined mixtures, Neural Networks, IEEE Transactions on 16 (4) (2005) 992 –996.

[10] P. Tichavsky, Z. Koldovsky, Weight adjusted tensor method for blind separation of underdetermined mixtures of nonstationary sources, Signal Processing, IEEE Transactions on 59 (3) (2011) 1037 –1047.

[11] P. Comon, Contrasts, independent component analysis, and blind deconvolution, International Journal of Adaptive Control and Signal Processing 18 (3) (2004) 225–243.

[12] A. Cichocki, R. Zdunek, A. H. Phan, S.-i. Amari, Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation, Wiley. com, 2009.

[13] G. Zhou, A. Cichocki, Canonical polyadic decomposition based on a single mode blind source separation, Signal Processing Letters, IEEE 19 (8) (2012) 523 –526.

[14] P. Chevalier, A. Ferreol, L. Albera, On the behavior of current second order blind source separation methods for first and second order cyclostationary sources-application to CPFSK sources, in: Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on, Vol. 3, 2002, pp. III–3081 –III–3084.

[15] A. Almeida, P. Comon, X. Luciani, Deterministic blind separation of sources having different symbol rates using tensor-based parallel deflation, in: Proceedings of the 9th international conference on Latent variable analysis and signal separation, LVA/ICA'10, Springer-Verlag, Berlin, Heidelberg, 2010, pp. 362–369.

[16] P. Comon, L. Luciani, A. L. F. D. Almeida, Tensor decompositions, alternating least squares and other tales, Journal of Chemometrics 23 (2009) 393–405.

[17] R. Bro, Multi-way analysis in the food industry: Models, algorithms and applications, Ph.D. thesis, University of Amsterdam (NL) (1998).

[18] S. A. Vorobyov, Y. Rong, N. C. Sidiropoulos, A. B. Gershman, Robust iterative fitting of multilinear models, IEEE Transactions on Signal Processing 53 (8) (2005) 2678–2689.

[19] J. Möcks, Decomposing event-related potentials: A new topographic components model, Biological Psychology 26 (1988) 199–215.

[20] J. Möcks, Topographic components model for event-related potentials and some biophysical considerations, IEEE Transactions on Biomedical Engineering 35 (6) (1988) 482–484.

[21] A. Field, D. Graupe, Topographic component (parallel factor) analysis of multichannel evoked potentials: practical issues in trilinear spatiotemporal decomposition, Brain Topography 3 (4) (1991) 407–423.

[22] L. de Lathauwer, B. de Moor, J. Vandewalle, Fetal electrocardiogram extraction by blind source subspace separation, IEEE Trans. Biomed. Eng. 47 (5) (2000) 567 –572.

[23] J.-F. Cardoso, Multidimensional Independent Component Analysis, in: Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '98), Vol. 4, 1998, pp. 1941–1944.

[24] A. G. Favret, Computer matched filter location of fetal R-waves., Med Biol Eng 6 (5) (1968) 467–475.

[25] B. Widrow, J. Glover, J.R., J. McCool, J. Kaunitz, C. Williams, R. Hearn, J. Zeidler, J. Eugene Dong, R. Goodlin, Adaptive noise cancelling: Principles and applications, Proceedings of the IEEE 63 (12) (1975) 1692 – 1716.

[26] R. Sameni, C. Jutten, M. B. Shamsollahi, Multichannel electrocardiogram decomposition using periodic component analysis, IEEE Trans. Biomed. Eng. 55 (8) (2008) 1935–1940.

[27] M. Richter, T. Schreiber, D. Kaplan, Fetal ECG extraction with nonlinear state-space projections, IEEE Trans. Biomed. Eng. 45 (1) (1998) 133 –137.

[28] J. B. Kruskal, Three-way arrays: rank and uniqueness of trilinear decompositions with application to arithmetic complexity and statistics, Linear Algebra and Applications 18 (1977) 95–138.

[29] N. D. Sidiropoulos, G. B. Giannakis, R. Bro, Blind PARAFAC receivers for DS-CDMA systems, IEEE Transactions on Signal Processing 48 (3) (2000) 810–822.

[30] T. G. Kolda, B. W. Bader, Tensor decompositions and applications, SIAM Review 51 (3) (2009) 455–500.

[31] P. J. Huber, E. M. Ronchetti (Eds.), Robust statistics, John Wiley & Sons, Hoboken, New Jersey, 2009, Ch. Scale Estimates, pp. 105–124.

[32] M. S. Grewal, A. P. Andrews, Kalman filtering: theory and practice using MATLAB, Wiley-IEEE press, 2011.

[33] A. Hyvarinen, Fast and robust fixed-point algorithms for independent component analysis, Neural Networks, IEEE Transactions on 10 (3) (1999) 626–634.

[34] S. Georgiadis, P. Ranta-aho, M. Tarvainen, P. Karjalainen, Single-trial dynamical estimation of event-related potentials: a kalman filter-based approach, Biomedical Engineering, IEEE Transactions on 52 (8) (2005) 1397 –1406.

[35] H. Mohseni, E. Wilding, S. Sanei, Preprocrssing of event-related potential signals via kalman filtering and smoothing, in: Digital Signal Processing, 2007 15th International Conference on, 2007, pp. 179 –182.

[36] M. L. Jongsma, T. Eichele, C. M. V. Rijn, A. M. Coenen, K. Hugdahl, H. Nordby, R. Q. Quiroga, Tracking pattern learning with single-trial event-related potentials, Clinical Neurophysiology 117 (9) (2006) 1957 – 1973.

[37] W. Deburchgraeve, P. J. Cherian, M. D. Vos, R. M. Swarte, J. H. Blok, G. H. Visser, P. Govaert, Neonatal seizure localization using parafac decomposition, Clinical Neurophysiology 120 (2009) 1787–1796.

[38] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, D. J. McFarland, P. H. Peckham, G. Schalk, E. Donchin, L. A. Quatrano, C. J. Robinson, T. M. Vaughan, et al., Brain-computer interface technology: a review of the first international meeting, IEEE transactions on rehabilitation engineering 8 (2) (2000) 164–173.

[39] B. De Moor, P. De Gersem, B. De Schutter, W. Favoreel, DAISY: A database for identification of systems, Journal A, Special Issue on CACSD (Computer Aided Control Systems Design) 38 (3) (1997) 4–5.

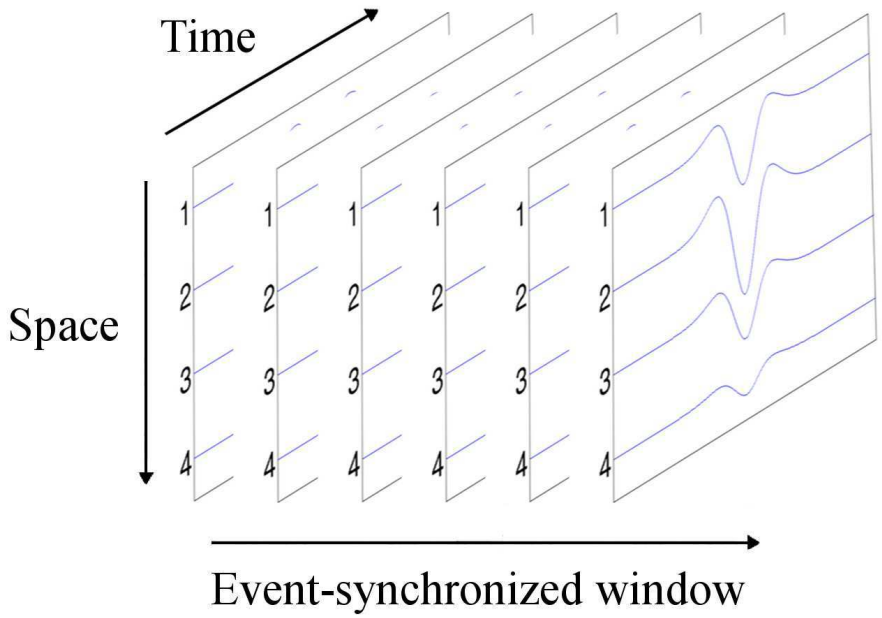

Figure 1: Illustration of a tensor with event-synchronized windows.

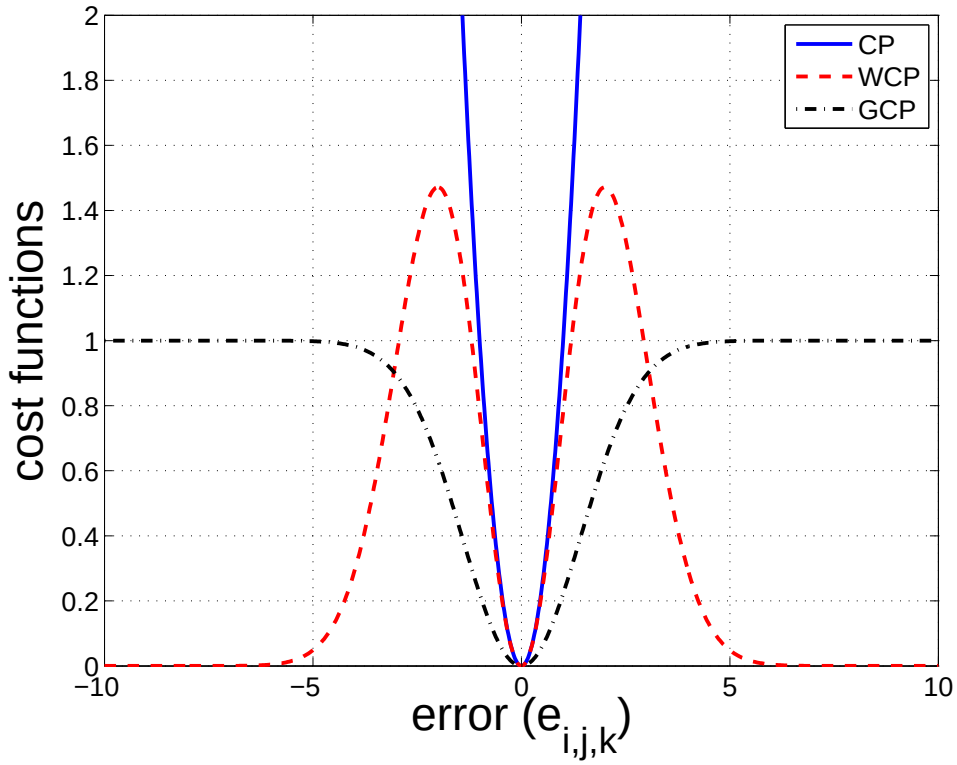

Figure 2: Cost functions applied to the error $e_{i,j,k}$ of each tensor element for the classical CP decomposition and the robust WCP and GCP methods.

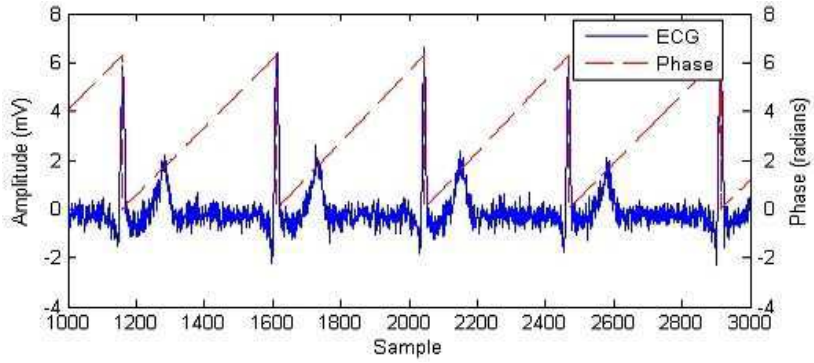

Figure 3: Illustration of linear time warping of each event interval into $[0, 2\pi)$ for a typical ECG signal.

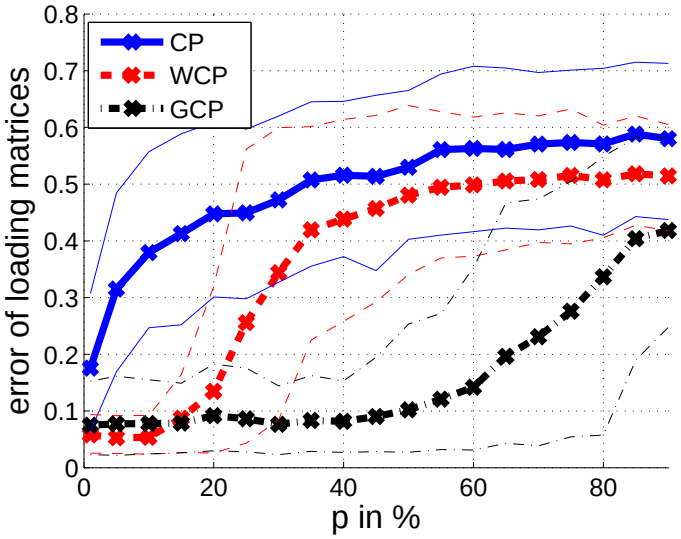

Figure 4: Error of the loading matrices as a function of the percentage of outliers. The bold curves show the results averaged over 200 trials while the 10% and 90% quantiles are represented by the thin curves.

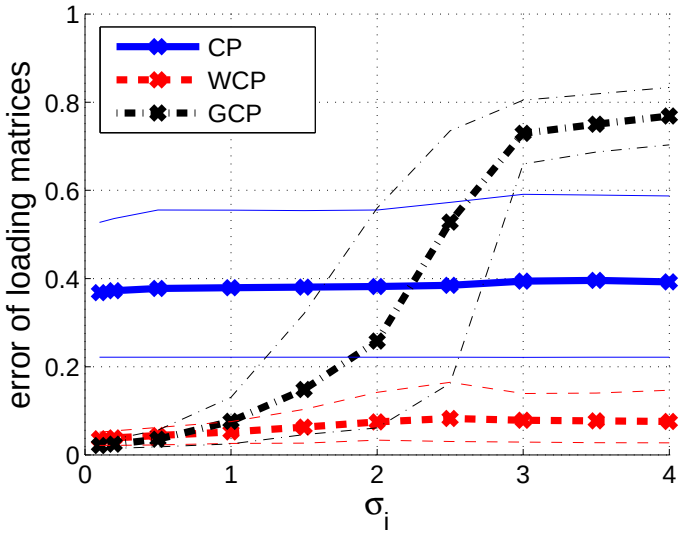

Figure 5: Error of the loading matrices as a function of the variance of the Gaussian noise added to the true loading matrices to obtain the initial loading matrices. The bold curves show the results averaged over 100 trials while the 10% and 90% quantiles are represented by the thin curves.

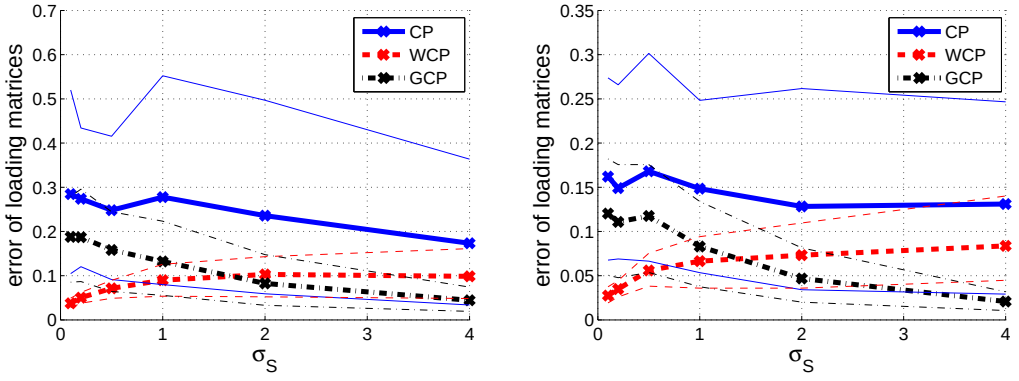

Figure 6: Error of the loading matrices as a function of the amplitude standard deviation σ_S for the tensor of the first source (left) and the tensor of the second source (right). The bold curves show the results averaged over 100 trials while the 10% and 90% quantiles are represented by the thin curves.

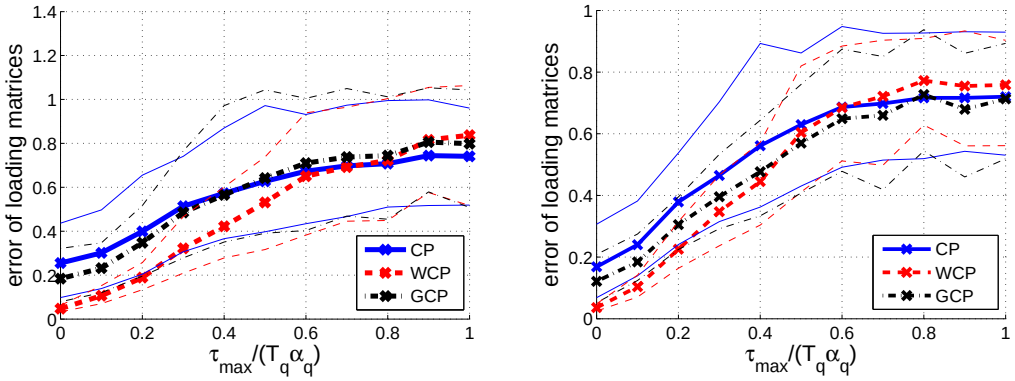

Figure 7: Error of the loading matrices as a function of the maximal relative synchronization error for the tensor of the first source (left) and the tensor of the second source (right). The bold curves show the results averaged over 100 trials while the 10% and 90% quantiles are represented by the thin curves.

Table 1: Maternal and fetal R-peak values on fECG estimate of DaISY dataset (mean ± standard deviation (SD)).

	Maternal R-peak value	Fetal R-peak value
Original mixture	43.66±2.38	17.68±2.37
FastICA	31.30±2.29	13.09±1.91
πCA	41.39±2.68	19.21±2.17
GCP	-0.90±0.91	16.04±2.72
WCP	-0.88±0.83	16.65±1.26
GCP+EKS	0.17±1.46	16.19±1.11
WCP+EKS	0.29±1.40	17.54±0.99

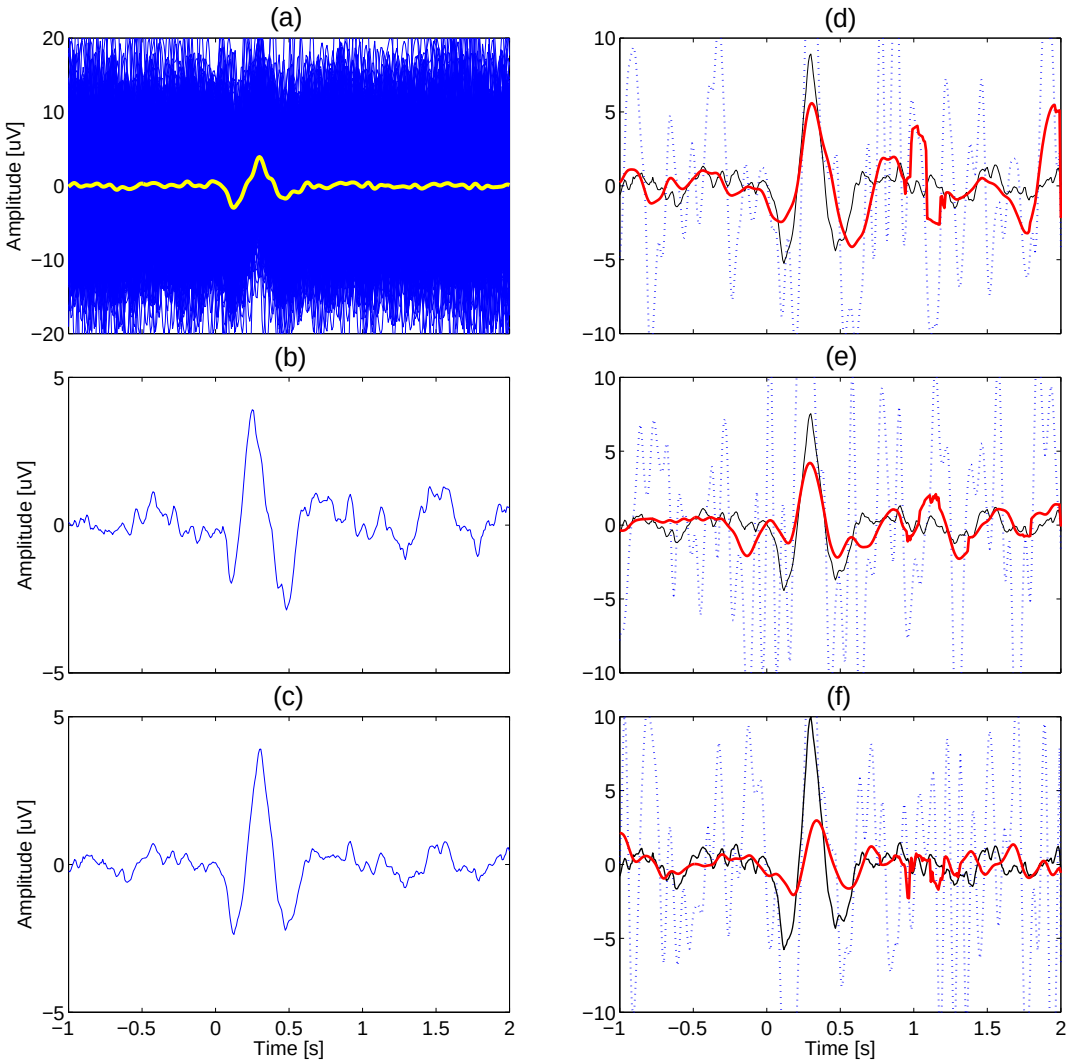

Figure 8: ERP extraction by WCP and WCP+EKS from the namely S6 dataset. (a): all the 500 measurements on channel 1 and the average in one plot; (b): the first temporal pattern extracted via WCP; (c): the second temporal pattern extracted via WCP; (d), (e), and (f): three examples of single-trial ERP extraction. Dotted lines represent the measurements, solid thin lines the estimates via WCP, and thick lines the estimates via WCP+EKS.

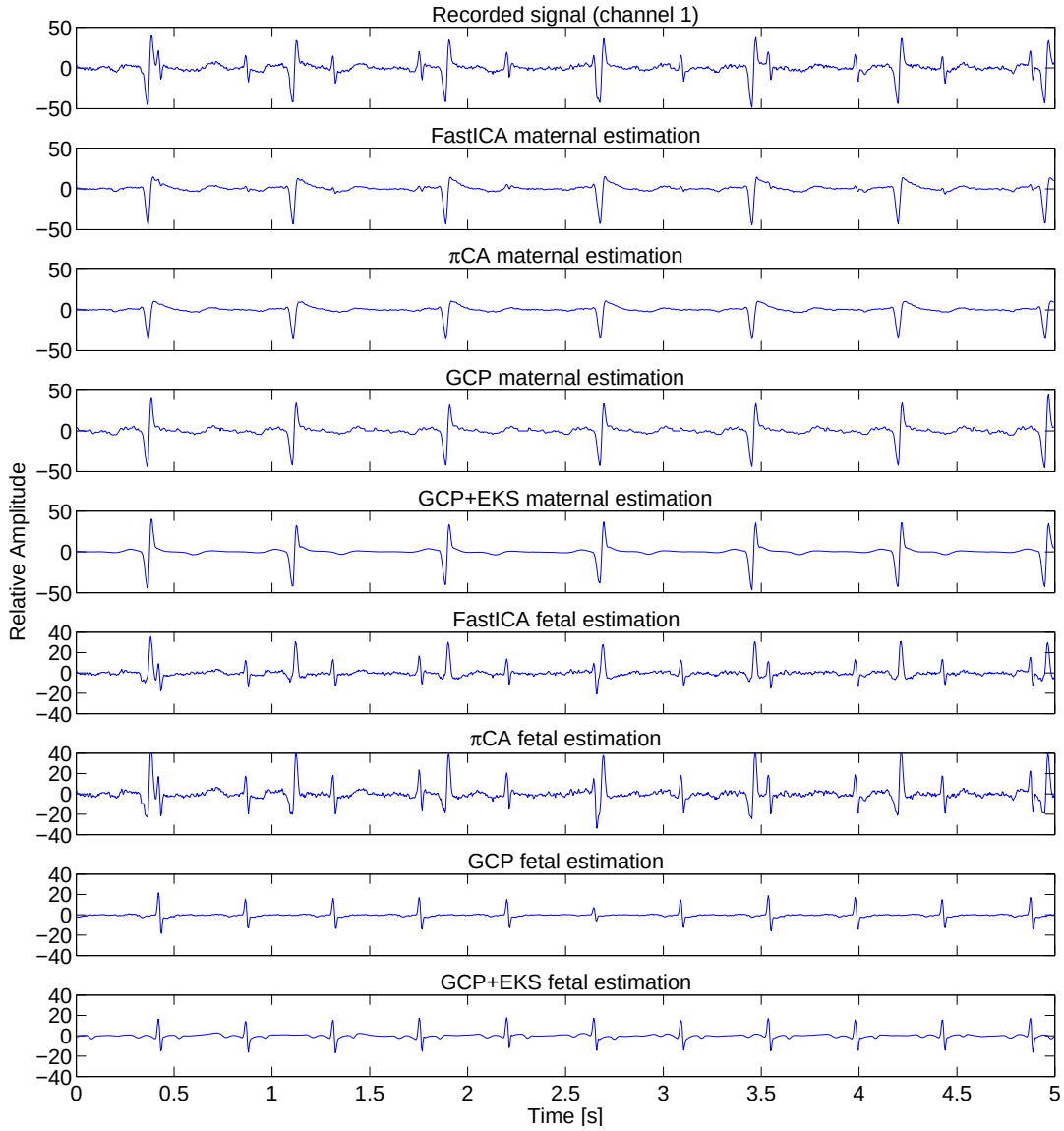

Figure 9: mECG and fECG extraction by FastICA, πCA, GCP and GCP+EKS on the first and second channels of DaISy data.

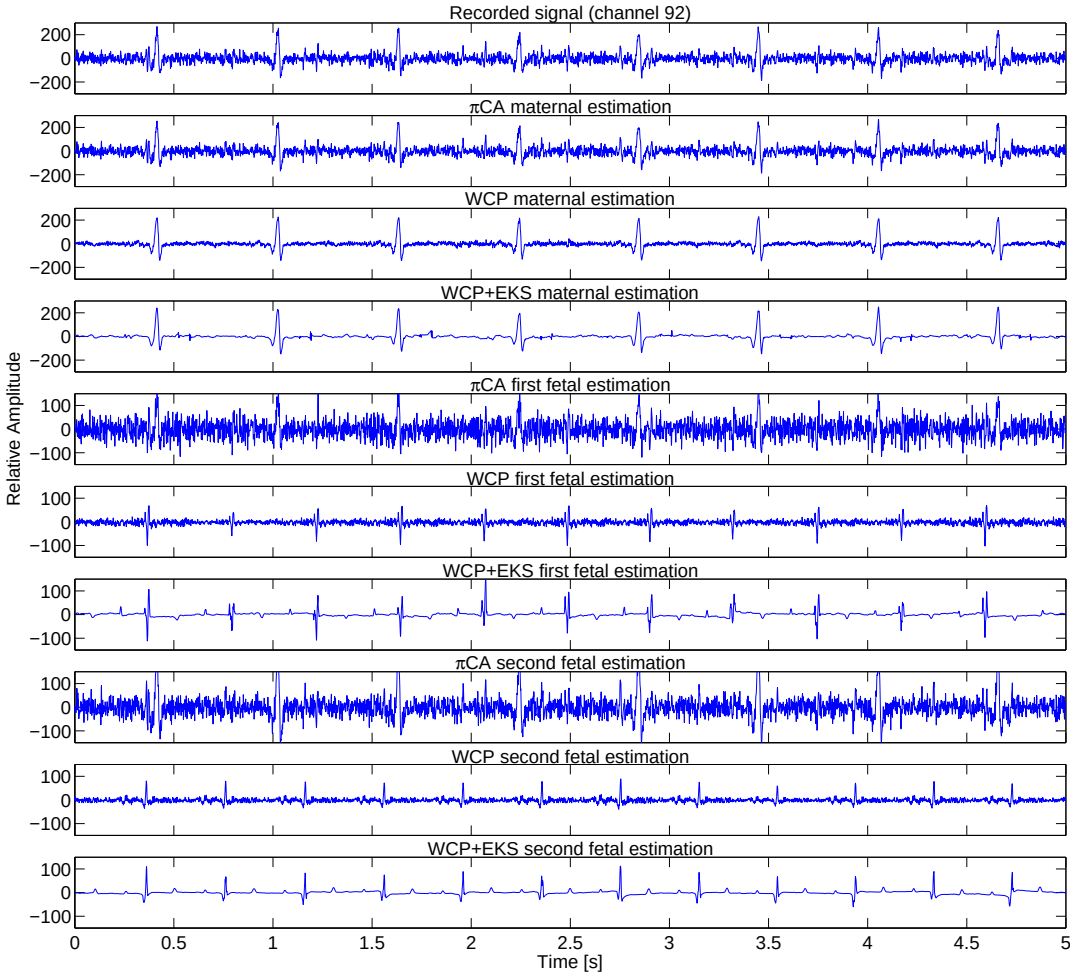

Figure 10: Maternal and fetal MCG extraction by πCA, WCP and WCP+EKS on the 92th and 116th channels of twin MCG data.

Table 2: Maternal and fetal R-peak values on the **first** fetal MCG estimate of twin MCG dataset (mean + SD).

	Maternal R-peak value	First fetal R-peak value	Second fetal R-peak value
Original mixture	210.08±31.42	66.04±40.74	74.97±29.27
πCA	159.72±25.79	63.15±36.77	21.28±24.39
GCP	-3.74±7.00	46.79±29.92	-3.08±9.55
WCP	-3.44±10.86	55.85±13.98	-2.57±8.37
GCP+EKS	1.94±8.10	65.48±33.29	1.06±8.85
WCP+EKS	1.39±6.77	71.22±28.12	0.20±6.75