Blind Source Separation of Event-Related EEG/MEG

Johanna Metsomaa*, Jukka Sarvas, and Risto Juhani Ilmoniemi

Abstract*—*Objective: **Blind source separation (BSS) can be used to decompose complex electroencephalography (EEG) or magnetoencephalography data into simpler components based on statistical assumptions without using a physical model. Applications include brain–computer interfaces, artifact removal, and identifying parallel neural processes. We wish to address the issue of applying BSS to event-related responses, which is challenging because of nonstationary data.** ***Methods:*** **We introduce a new BSS approach called momentary-uncorrelated component analysis (MUCA), which is tailored for event-related multitrial data. The method is based on approximate joint diagonalization of multiple covariance matrices estimated from the data at separate latencies. We further show how to extend the methodology for autocovariance matrices and how to apply BSS methods suitable for piecewise stationary data to event-related responses. We compared several BSS approaches by using simulated EEG as well as measured somatosensory and transcranial magnetic stimulation (TMS) evoked EEG.** ***Results:*** **Among the compared methods, MUCA was the most tolerant one to noise, TMS artifacts, and other challenges in the data. With measured somatosensory data, over half of the estimated components were found to be similar by MUCA and independent component analysis. MUCA was also stable when tested with several input datasets.** ***Conclusion:*** **MUCA is based on simple assumptions, and the results suggest that MUCA is robust with nonideal data.** ***Significance:*** **Event-related responses and BSS are valuable and popular tools in neuroscience. Correctly designed BSS is an efficient way of identifying artifactual and neural processes from nonstationary event-related data.**

***Index Terms*—Blind source separation (BSS), electroencephalography (EEG), event-related fields, event-related potentials, independent component analysis, magnetoencephalography (MEG), transcranial magnetic stimulation (TMS).**

Manuscript received August 1, 2016; revised September 8, 2016; accepted September 29, 2016. Date of publication October 12, 2016; date of current version August 18, 2017. This work was supported in part by the Academy of Finland (Decision no. 283105) and in part by the Finnish Cultural Foundation. This paper was presented in part at the BIOMAG Conference in August 24–28, 2014 [21]. *Asterisk indicates corresponding author.*

*J. Metsomaa is with the Department of Neuroscience and Biomedical Engineering, Aalto University School of Science, Espoo 02150, Finland, and also with the BioMag Laboratory, HUS Medical Imaging Center, University of Helsinki and Helsinki University Hospital, Helsinki 00100, Finland (e-mail: johanna.metsomaa@aalto.fi).

J. Sarvas is with the Department of Neuroscience and Biomedical Engineering, Aalto University School of Science.

R. J. Ilmoniemi is with the Department of Neuroscience and Biomedical Engineering, Aalto University School of Science, and also with the BioMag Laboratory, HUS Medical Imaging Center, University of Helsinki and Helsinki University Hospital.

Digital Object Identifier 10.1109/TBME.2016.2616389

I. Introduction

When event-related responses of the brain are recorded with electroencephalography (EEG) or magnetoencephalography (MEG), resulting in evoked potentials or evoked fields, respectively, several responses to the repeated stimuli are measured. Data-driven blind source separation (BSS) techniques are beneficial in extracting signal sources from the responses. BSS divides the data into components, each of which has a fixed spatial pattern in the sensor space and a waveform in the time domain. These techniques do not require complicated modeling of the physical properties of the signal sources or the head conductivity distribution. Therefore, there can be both intra- and extracranial sources originating from neural and nonneural processes. Instead, statistical properties of the underlying components need to be modeled in order to make their separation possible.

The interpretation of the evoked responses can be complicated due to several temporally overlapping neural processes. For example, a deflection in the averaged EEG may correspond to several parallel neural activations. The advantage of BSS in EEG analysis is that such overlapping processes can be separated from each other and, thus, their interpretation becomes simpler than that of the original data (see, e.g., [13] and [25]). BSS on the single-trial level can be used in online signal analysis, including brain–computer interfaces [10], [23]. In addition, BSS is an efficient way of removing artifacts, e.g., high-amplitude transcranial magnetic stimulation (TMS) evoked artifacts in EEG, and uncovering neural responses from artifactual data [12], [18], [22].

Even though evoked potentials are highly nonstationary, it has been common to apply BSS methods designed for stationary data on evoked EEG [11], [12], [16], [20], [30], [31], [34]. To get unbiased estimates for the hidden components, we present several BSS methods that all take into account the nonstationarity in the evoked data. We also make extensive comparisons between these techniques.

We introduce a BSS method called *momentary-uncorrelated component analysis (MUCA)* to decompose multitrial EEG/MEG data into momentary-uncorrelated components, i.e., components that are uncorrelated at each latency after the stimulus. MUCA is based on approximate joint diagonalization (AJD) of covariance matrices estimated at separate latencies of the evoked data. We call these the *momentary covariance matrices* (MCMs). Since the evoked responses are nonstationary, such components can be found unambiguously provided that the computed MCMs have enough variability over time. MUCA

takes the MCMs as the input data to which AJD is applied. We also introduce a two-step filtering procedure to reduce the measurement and the sampling noise in the MCMs.

AJD is a commonly used approach in the BSS of EEG and MEG. In addition to covariance matrices, the target matrices to be diagonalized can also be, for example, autocovariance matrices. The simplest form of joint diagonalization is having two target matrices [4], [33]. In many BSS applications, however, there are more than two matrices to be diagonalized, which is the case with second-order blind identification (SOBI) [2], [3] and JADE [5]. When the number of target matrices is more than two, exact joint diagonalization is, in general, not possible, and one has to resort to AJD. Several algorithms exist for AJD [9], [27], [28], [36], [37].

Nonstationary data have also previously been proposed suitable for AJD-based BSS methods. Some methods are based on dividing a long recording of nonstationary data into blocks, within which the data are stationary. Computing the autocovariance matrices in each of these blocks and applying AJD to all of them yields sufficient amount of information for separating the components [6], [7], [29]. In this paper, we also illustrate how to use this same idea of jointly diagonalizing autocovariance matrices in the case of evoked multitrial data. We further show how any BSS method suited for piecewise stationary data can be applied to event-related data.

We compared MUCA and other BSS methods suited for nonstationary data using both simulated and measured EEG. In general, the MUCA estimates were more accurate than those obtained by the other methods, even at the presence of high-amplitude TMS-induced artifacts. In this paper, we deal with EEG but the used methodology is applicable to MEG as well.

A. Notation

Let $\mathbf{A}$ be an $m \times n$ matrix, with elements $\mathbf{A}(i,j), 1 \le i \le m$, $1 \le j \le n$. We denote the ith *row* of $\mathbf{A}$ by $\mathbf{A}(i,:), i = 1, \ldots, m$, and the jth *column* by $\mathbf{A}(:,j), j = 1, \ldots, n$.

We also use three-dimensional (3-D) arrays with the following notation. Let $\mathbf{A}$ be a 3-D array with elements $\mathbf{A}(i,j,k),\ i = 1, \ldots, m,\ j = 1, \ldots, n,$ and $k = 1, \ldots, q$. For a fixed i, the notation $\mathbf{A}(i,:,:)$ denotes the 2-D matrix with elements $\mathbf{A}(i,j,k),\ j = 1, \ldots, n,$ and $k = 1, \ldots, q$. For a fixed j or k, the 2-D matrices $\mathbf{A}(:,j,:)$ or $\mathbf{A}(:,:,k)$ are defined analogously. Furthermore, for a fixed i and j, the notation $\mathbf{A}(i,j,:)$ denotes the vector with elements $\mathbf{A}(i,j,k),\ k = 1, \ldots, q$. The vectors $\mathbf{A}(:,j,k)$ and $\mathbf{A}(i,:,k)$ are defined analogously.

If x is a scalar, vector, or matrix *random variable*, we denote its *expectation value* by $\mathrm{E}(x)$ and *variance* (when x is scalar) by $\mathrm{var}(x)$. For an $m \times n$ (data) matrix $\mathbf{X}$, the sample mean is given by $\boldsymbol{\mu}_\mathrm{X} = n^{-1} \sum_{j=1}^{n} \mathbf{X}(:,j)$. The *covariance matrix* of a random vector $\boldsymbol{x}$ is denoted by $\mathrm{cov}(\boldsymbol{x}) = \mathrm{E}((\boldsymbol{x} - \mathrm{E}(\boldsymbol{x}))(\boldsymbol{x} - \mathrm{E}(\boldsymbol{x}))^\mathrm{T})$. We denote the sample covariance matrix by $\mathrm{cov}(\mathbf{X})$ and define it as

$$\mathrm{cov}(\mathbf{X}) = \frac{1}{n} \sum_{j=1}^{n} (\mathbf{X}(:,j) - \boldsymbol{\mu}_\mathrm{X})(\mathbf{X}(:,j) - \boldsymbol{\mu}_\mathrm{X})^\mathrm{T}. \quad (1)$$

In the time-series $\boldsymbol{x}$, the autocovariance matrix between two separate time instants p and q is denoted by $\mathrm{cov}_{p,q}(\boldsymbol{x}) = \mathrm{E}((\boldsymbol{x}_p - \mathrm{E}(\boldsymbol{x}_p))(\boldsymbol{x}_q - \mathrm{E}(\boldsymbol{x}_q))^\mathrm{T})$. If $\mathbf{X}_p$ and $\mathbf{X}_q$ are $m \times n$ sample matrices from $\boldsymbol{x}_p$ and $\boldsymbol{x}_q$, respectively, the sample autocovariance matrix between them is denoted by $\mathrm{cov}(\mathbf{X}_p, \mathbf{X}_q)$ and computed by

$$\mathrm{cov}(\mathbf{X}_p, \mathbf{X}_q) = \frac{1}{n} \sum_{j=1}^{n} (\mathbf{X}_p(:,j) - \boldsymbol{\mu}_{\mathrm{X}_p})(\mathbf{X}_q(:,j) - \boldsymbol{\mu}_{\mathrm{X}_q})^\mathrm{T}. \quad (2)$$

II. Materials and Methods

A. Theory

1) Multitrial Event-Related Data: We assume that the evoked responses are measured with N_c *channels*, at N_t successive *time points*, and in N_tr *trials*, resulting in a 3-D data array $\mathbf{X}$, where the matrix element $\mathbf{X}(i,j,k)$ is the EEG signal in the ith channel at the jth time instant in the kth trial. A trial refers to a single stimulus repetition.

We assume that the data obey a *linear model*, where the measured EEG data are due to N_h *hidden* (or latent) components and additive noise, defined as

$$\mathbf{X}(:,j,k) = \mathbf{A}\,\mathbf{S}(:,j,k) + \boldsymbol{\varepsilon}(:,j,k), \quad (3)$$

for $j = 1, \ldots, N_\mathrm{t},\ k = 1, \ldots, N_\mathrm{tr}$ where $\mathbf{A}$ is an $N_\mathrm{c} \times N_\mathrm{h}$ *mixing matrix*, which is independent of time and trials, $\mathbf{S}$ is an $N_\mathrm{h} \times N_\mathrm{t} \times N_\mathrm{tr}$ *time-course array*, and $\boldsymbol{\varepsilon}$ contains noise. Here, $\mathbf{A}(:,m)$ is called the *topography* of the mth component. The vector $\mathbf{S}(m,:,k)$ is called the *time-course* (or *waveform*) of mth component in trial k. We assume that $N_\mathrm{c} \ge N_\mathrm{h} = \mathrm{rank}(\mathbf{A})$.

2) Statistical Model of the Data: For a fixed time instant j, we use the notations $\mathbf{X}_j = \mathbf{X}(:,j,:)$ and $\mathbf{S}_j = \mathbf{S}(:,j,:)$. In the statistical analysis of the model (3), we assume that the column vectors $\mathbf{S}_j(:,k),\ k = 1, \ldots, N_\mathrm{tr}$, are random samples of a *momentary random variable* $\boldsymbol{s}_j$. Note that each time instant j corresponds to a different probability density function $p(\boldsymbol{s}_j)$, which allows for nonstationary statistics for the components. The linear model assumed for the momentary random variables $\boldsymbol{x}_j$ and $\boldsymbol{s}_j$ at any given j is given as

$$\boldsymbol{x}_j = \mathbf{A}\,\boldsymbol{s}_j \quad (4)$$

where $\mathbf{X}_j(:,k),\ k = 1, \ldots, N_\mathrm{tr}$, are random samples of a variable $\boldsymbol{x}_j$. For simplicity, the noise term is ignored here. We may set various statistical modeling assumptions on $\boldsymbol{s}_j$.

Due to (4), the MCMs are given as

$$\mathrm{cov}(\boldsymbol{x}_j) = \boldsymbol{\Sigma}_j = \mathbf{A}\,\boldsymbol{\Lambda}_j\mathbf{A}^\mathrm{T}, \quad \text{for } j = 1, \ldots, N_\mathrm{t} \quad (5)$$

where $\boldsymbol{\Lambda}_j = \mathrm{cov}(\boldsymbol{s}_j)$. For the components that are uncorrelated at a time instant j, $\boldsymbol{\Lambda}_j$ is diagonal.

The momentary autocovariance matrices (MACMs) can be expressed according to (4) by

$$\mathrm{cov}_{j,j'}(\boldsymbol{x}) = \boldsymbol{\Sigma}_{j,j'} = \mathbf{A}\,\boldsymbol{\Lambda}_{j,j'}\mathbf{A}^\mathrm{T}, \quad \text{for } j, j' = 1, \ldots, N_\mathrm{t} \quad (6)$$

with $\boldsymbol{\Lambda}_{j,j'} = \mathrm{cov}_{j,j'}(\boldsymbol{s})$. If the components do not correlate between separate time instants j and j', $\boldsymbol{\Lambda}_{j,j'}$ is diagonal.

If the components of $\boldsymbol{s}_j$ are independent,

$$p(\boldsymbol{s}_j) = p(s_j(1)) \cdots p(s_j(N_\mathrm{h})). \tag{7}$$

In the following, we illustrate different approaches of BSS assuming uncorrelated or independent components and taking into account the time-varying statistics of the event-related data. In all the approaches, the aim is to find estimates for the mixing matrix and the waveforms matrix in (3) based on the given multitrial data.

3) BSS by Momentary-Uncorrelated Components Analysis: We assume that the components of $\boldsymbol{s}_j$ are uncorrelated for all $j = 1, \ldots, N_\mathrm{t}$, i.e., they are momentary-uncorrelated. The sample momentary covariance matrix $\mathrm{cov}(\mathbf{X}_j) = \mathbf{C}_j$ approximates $\boldsymbol{\Sigma}_j$, and $\mathrm{cov}(\mathbf{S}_j) = \mathbf{D}_j$ approximates $\boldsymbol{\Lambda}_j$, which is diagonal. The statistical model (5) implies that

$$\mathbf{C}_j = \mathbf{A}\mathbf{D}_j\mathbf{A}^\mathrm{T} \text{ and } \mathbf{D}_j = \mathbf{A}^{-1}\mathbf{C}_j(\mathbf{A}^\mathrm{T})^{-1} \tag{8}$$

where we may assume that $\mathbf{A}$ is a square matrix obtained by data compression based on the principal component analysis (PCA), which is generally performed also prior to ICA [15]. Because $\mathbf{D}_j$s are approximately diagonal, (8) shows that $\mathbf{A}^{-1}$ is the approximate joint diagonalizer of $\mathbf{C}_1, \ldots, \mathbf{C}_{N_\mathrm{t}}$. Thus, we can use AJD algorithms to find an estimate $\widehat{\mathbf{A}}$ for the mixing matrix, whereafter $\widehat{\mathbf{S}}_j = \widehat{\mathbf{A}}^{-1}\mathbf{X}_j$.

In this paper, we applied the following AJD algorithms: FF-Diag, J-Di, ACDC, and Pham's algorithm, for details, see [37], [28], [36], [27], respectively. These algorithms assume that the covariance matrices are symmetric square matrices that change as a function of time, i.e., the data are nonstationary [1]. Pham's algorithm also assumes that the MCMs are positive definite. Note that the data do not need to be whitened.

4) Filtering the MCMs: Let us again consider (3), which, with the notation $\mathbf{X}_j = \mathbf{X}(:, j, :)$, $\mathbf{S}_j = \mathbf{S}(:, j, :)$, and $\boldsymbol{\varepsilon}_j = \boldsymbol{\varepsilon}(:, j, :)$, reads as

$$\mathbf{X}_j = \mathbf{A}\mathbf{S}_j + \boldsymbol{\varepsilon}_j, \; j = 1, \ldots, N_\mathrm{t}\,. \tag{9}$$

Because $\boldsymbol{\varepsilon}_j$ is not correlated with $\mathbf{S}_j$,

$$\mathbf{C}_j = \mathrm{cov}(\mathbf{X}_j) = \mathbf{A}\mathrm{cov}(\mathbf{S}_j)\mathbf{A}^\mathrm{T} + \mathrm{cov}(\boldsymbol{\varepsilon}_j)\,. \tag{10}$$

We assume that the measurement noise is stationary or almost stationary in the sense that $\mathrm{cov}(\boldsymbol{\varepsilon}_j)$ varies slowly with time, i.e., $\mathrm{cov}(\boldsymbol{\varepsilon}_j) \approx \mathrm{cov}(\boldsymbol{\varepsilon})$. In addition to the measurement noise, there is sampling error in $\mathbf{C}_j$ due to the limited number of trials. Before applying AJD to the MCMs, we filter them in two steps in order to suppress the sampling and the measurement noise.

We filter by moving averaging with Gaussian weights $w_{k,j}$, $1 \le k, j \le N_\mathrm{t}$. The filtered sequence of the MCMs $\mathbf{C}^\mathrm{F}_1, \ldots, \mathbf{C}^\mathrm{F}_{N_\mathrm{t}}$ is given by

$$\mathbf{C}^\mathrm{F}_k = \frac{\sum_{j=1}^{N_\mathrm{t}} w_{k,j}\mathbf{C}_j}{\sum_{j=1}^{N_\mathrm{t}} w_{k,j}}, \quad \text{where } w_{k,j} = e^{-\frac{1}{2}(\frac{k-j}{d})^2} \tag{11}$$

with $d > 0$ controlling the width of the averaging Gaussian "curve."

In the first step, we filter $\mathbf{C}_1, \ldots, \mathbf{C}_{N_\mathrm{t}}$ using weights $w_{k,j}$ with a small value for d, and get the sequence $\mathbf{C}^{(1)}_1, \ldots, \mathbf{C}^{(1)}_{N_\mathrm{t}}$.

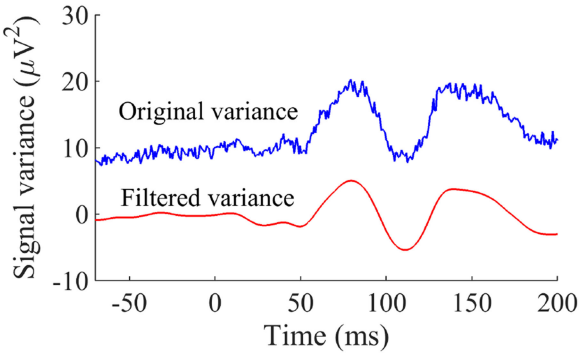

Fig. 1. Example of filtering the MCMs computed using somatosensory evoked potentials. The blue curve (upper) is the original variance in the channel 37 (between C3 and P3 in the 10–20 EEG system). The red curve is the filtered variance. Note that the filtered variance curve also gets negative values.

This filtering suppresses the high-frequency part of the sampling error in $\mathrm{cov}(\mathbf{X}_j)$. In the second step, we filter $\mathbf{C}_1, \ldots, \mathbf{C}_{N_\mathrm{t}}$ with a large value for d, and get the sequence $\mathbf{C}^{(2)}_1, \ldots, \mathbf{C}^{(2)}_{N_\mathrm{t}}$, which contains the slowly drifting part (mostly measurement noise) of the covariance matrices. The final filtered sequence is given by the differences

$$\widetilde{\mathbf{C}}_k = \mathbf{C}^{(1)}_k - \mathbf{C}^{(2)}_k, \; k = 1, \ldots, N_\mathrm{t}\,. \tag{12}$$

We can also write these filtered MCMs as

$$\begin{aligned}\widetilde{\mathbf{C}}_k &= \mathbf{A}(\mathbf{D}^{(1)}_k - \mathbf{D}^{(2)}_k)\mathbf{A}^\mathrm{T} + \mathrm{cov}(\boldsymbol{\varepsilon}_k)^{(1)} - \mathrm{cov}(\boldsymbol{\varepsilon}_k)^{(2)} \\ &\approx \mathbf{A}(\mathbf{D}^{(1)}_k - \mathbf{D}^{(2)}_k)\mathbf{A}^\mathrm{T}, \; k = 1, \ldots, N_\mathrm{t}\end{aligned} \tag{13}$$

where $\mathbf{D}^{(1)}_k$, $\mathrm{cov}(\boldsymbol{\varepsilon}_k)^{(1)}$, and $\mathbf{D}^{(2)}_k$, $\mathrm{cov}(\boldsymbol{\varepsilon}_k)^{(2)}$ are obtained by the filtering steps 1 and 2, respectively. Filtering suppresses the stationary noise term in (13), because $\mathrm{cov}(\boldsymbol{\varepsilon}_k)^{(1)} \approx \mathrm{cov}(\boldsymbol{\varepsilon}_k)^{(2)}$, and reduces the sampling error in $\mathbf{D}^{(1)}_k - \mathbf{D}^{(2)}_k$.

The sequence of filtered MCMs from (12) is the one to which we apply AJD. According to (13), the underlying structure of these target matrices has been preserved so that $\mathbf{A}^{-1}$ still jointly diagonalizes the filtered ones. It is notable that the filtered matrices are no longer positive definite. Thus, Pham's AJD algorithm is not applicable to the filtered MCMs. In practice, d in the two types of filtering is experimentally tuned to yield efficient noise suppression. An example of the effect of the filtering is presented in Fig. 1.

5) BSS of Uncorrelated Components Over Time Lags: If the hidden components are uncorrelated over time lags, we may further include MACMs for the AJD procedure described in the previous section. The symmetrized form of MACM with time index pair j, j' is approximated by $\mathbf{C}_{j,j'} = 0.5[\mathrm{cov}(\mathbf{X}_j, \mathbf{X}_{j'}) + \mathrm{cov}(\mathbf{X}_{j'}, \mathbf{X}_j)]$. The same joint diagonalizer $\mathbf{A}^{-1}$ then yields

$$\mathbf{A}^{-1}\mathbf{C}_{j,j'}(\mathbf{A}^{-1})^\mathrm{T} = \mathbf{D}_{j,j'} \tag{14}$$

where $\mathbf{D}_{j,j'} = 0.5[\mathrm{cov}(\mathbf{S}_j, \mathbf{S}_{j'}) + \mathrm{cov}(\mathbf{S}_{j'}, \mathbf{S}_j)]$, which is diagonal if the hidden components are uncorrelated over the time lag defined by j and j' [see (6)]. In practice, some subset

of time lags $(\tau_1, \ldots, \tau_P)$ is used in the estimation so that $j' = j + \tau_p$, $p = 1, \ldots, P$ for each $j = 1, \ldots, N_\mathrm{t} - \tau_p$. AJD is applied to the sequence of MACMs with all the selected time lags. The estimation tends to get more accurate with increasing number of target matrices if the assumptions of uncorrelatedness are correct. Overall, this approach is somewhat similar to the SOBI algorithm [2], [3], with the important distinction that the autocovariance matrices are allowed to vary with time j.

For noisy autocovariance matrices, $\mathrm{cov}(\mathbf{X}_j, \mathbf{X}_{j+\tau}) = \mathbf{A}\mathrm{cov}(\mathbf{S}_j, \mathbf{S}_{j+\tau})\mathbf{A}^\mathrm{T} + \mathrm{cov}(\boldsymbol{\varepsilon}_j, \boldsymbol{\varepsilon}_{j+\tau})$. Filtering by (12) can be applied to the MACMs by replacing $\mathbf{C}_j$ by $\mathbf{C}_{j,j+\tau}$ in (11). Again, we assume that noise is (almost) stationary such that $\mathrm{cov}(\boldsymbol{\varepsilon}_j, \boldsymbol{\varepsilon}_{j+\tau})$ does not change with j. When using multiple time lags, filtering is performed for each of them separately.

6) BSS by ICA and Utilizing Piecewise Stationarity: Independent component analysis can be used to estimate the hidden components if they are independent and non-Gaussian. ICA algorithms typically assume that all collected data samples originate from the same (joint) probability distribution. Thus, they are usually not suitable for nonstationary evoked EEG where only momentary independence according to (7) can be assumed.

In [22], we showed that, after a special preprocessing, FastICA [14] can uncover hidden components even from highly nonstationary evoked EEG, whereas otherwise ICA is unsuccessful with these kind of data. Here, we refer to this approach as nonstationary FastICA (FastICA-ns).

Another way to apply ICA to nonstationary data is Block EFICA, which is a modified version of the FastICA algorithm and designed for piecewise stationary data [17]. Block EFICA is based on the assumption that a long recording of nonstationary data can be divided into blocks within which the components are stationary. According to (4), multitrial data can be reorganized to fulfill the assumption of piecewise stationarity by forming concatenated data as $\mathbf{Y} = [\mathbf{X}_1, \ldots, \mathbf{X}_{N_\mathrm{t}}]$, with N_t successive stationary blocks $\mathbf{X}_j$, each of length N_tr. In fact, any BSS algorithm designed for piecewise stationary data, for example, the Barbi algorithm described in [32], can be applied to evoked responses after reorganizing the data in this manner. MUCA can also be interpreted from this perspective. Thus, it can be viewed as an extension of the BSS algorithm suggested for Gaussian piecewise stationary sources in [27].

7) Estimating the Waveforms: The estimated averaged time-courses are given by the sample means of $\widehat{\mathbf{S}}_j$ as $\widehat{\mathbf{S}}_\mathrm{ave} = [\mu_{\widehat{\mathbf{S}}_1}, \ldots, \mu_{\widehat{\mathbf{S}}_{N_\mathrm{t}}}]$. Using the estimated mixing matrix in (8), the variance waveform of component i is approximated by the sample variances of $\widehat{\mathbf{S}}_j(i,:)$ as $\widehat{\mathbf{S}}_\mathrm{var}(i,:) = [\widehat{\mathbf{D}}_1(i,i), \ldots, \widehat{\mathbf{D}}_{N_\mathrm{t}}(i,i)]$. Note that any statistical parameters of the components can be computed as a function of time j by using the samples $\widehat{\mathbf{S}}_j$.

B. Simulations

All the data simulations and analyses were performed on MATLAB (The Mathworks, Inc., Natick, Massachusetts, USA). The sample matrices $\mathbf{S}$ were generated by a two-step process. We started by generating stochastic, stationary data containing samples $\tilde{\mathbf{S}}$: For a fixed component i and trial k, the first sample was randomized from the normal distribution $\tilde{\mathbf{S}}(i,1,k) \sim \mathcal{N}(0,1)$. For the following time instants, the samples were recursively computed by

$$\tilde{\mathbf{S}}(i,j,k) = r_{i,j,k} + a\,\tilde{\mathbf{S}}(i,j-1,k), \qquad j = 2, \ldots, N_\mathrm{t} \tag{15}$$

where a is a constant defining the degree of smoothness of the time-courses $\tilde{\mathbf{S}}(i,:,k)$, and $r_{i,j,k} \sim \mathcal{N}(0,1)$ was randomized for every sample separately. This procedure was taken for all the components i and the trials k.

To make the waveforms nonstationary, the amplitudes of $\tilde{\mathbf{S}}$ were modulated with time-dependent envelopes $u_{i,j}$ as

$$\mathbf{S}(i,j,k) = \tilde{\mathbf{S}}(i,j,k)\,u_{i,j} + u_{i,j}\,. \tag{16}$$

Comparing (16) with the statistical model for $\boldsymbol{s}_j$, the components had both the variances $\mathrm{var}(s_j(i)) \propto u_{i,j}^2$ and the means $\mathrm{E}(s_j(i)) = u_{i,j}$ varying as a function of time j. Thus, the nonstationarity was regulated entirely by the selection of the envelope function. Importantly, these hidden components are mutually independent at each moment of time and they have diagonal MCMs and MACMs. The recursive process (15) does not necessarily need to be stable, since the envelope in (16) controls the amplitudes.

The envelopes were defined by

$$u_{i,j} = \exp\{(j-\mu_i)^2/(2h)^2\} \tag{17}$$

where μ_i is the time index where the amplitude of the ith component is largest on average and h determines the duration of its nonzero activity on average. The amplitudes were scaled so that $\sum_{j,k} \mathbf{S}(i,j,k)^2/(N_\mathrm{t} \cdot N_\mathrm{tr}) = 1$ for all i. We set the simulation parameters $a = 10$, $h = 15$, and the centers μ_i, $i = 1, \ldots, N_\mathrm{h}$, were equally spaced within the time axis $1 \le j \le N_\mathrm{t}$. An example trial of the simulated components is illustrated in Fig. 2.

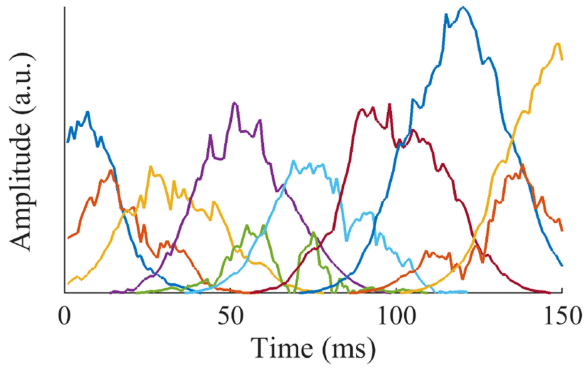

Fig. 2. Waveforms of the simulated components in an example trial. The polarities of all the components are here set positive.

The mixing matrix $\mathbf{A}$ was randomized as a square matrix, the condition number being controlled between 0 and 300. The noiseless data $\overline{\mathbf{X}}$ were constructed using (3) without the noise term. The $N_\mathrm{c} \times N_\mathrm{t} \times N_\mathrm{tr}$ noise matrix $\boldsymbol{\varepsilon}$ was added to the data as spatially colored and temporally white Gaussian noise with zero mean. The noise covariance matrix was stationary, $\mathrm{cov}(\boldsymbol{\varepsilon}_j) = \mathrm{cov}(\boldsymbol{\varepsilon})$, and its condition number was 50^2, based on the experience from measured data covariance matrices. The noise amplitude was scaled according to the predefined noise

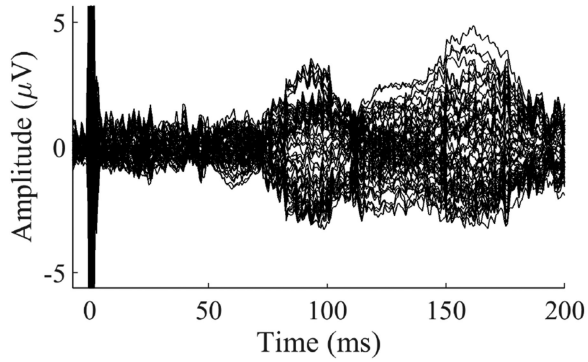

Fig. 3. Averaged evoked somatosensory EEG data after right median nerve stimulation. There is an electric artifact at the start of the signal.

level, NL, defined as

$$\mathrm{NL} = \sqrt{\frac{\sum_{i,j,k} \boldsymbol{\varepsilon}(i,j,k)^2}{\sum_{i,j,k} \overline{\mathbf{X}}(i,j,k)^2}}. \tag{18}$$

NL indicates how large the noise is relative to the noiseless signal amplitude; it is the inverse of the signal-to-noise ratio (SNR). Finally, the data were formed by using (3).

The simulations were run with several simulation parameters using ICA, MUCA, and AJD of MACMs to compare how accurately they could recover the hidden components. Generally, unless otherwise stated, the simulation parameters were chosen as $N_\mathrm{c} = N_\mathrm{h} = 20$, $N_\mathrm{t} = 150$, $N_\mathrm{tr} = 150$, $a = 10$, NL (=1/SNR) = 1, and the mixing matrix condition number was 50.

1) Measuring the Accuracy of the Simulation Results: The comparison of MUCA and ICA was based on the inner products between the true and the estimated topographies. For the ith component, we used the following coefficient for measuring the quality of the estimation:

$$c_i = \frac{|\widehat{\mathbf{A}}(:,i)^\mathrm{T}\mathbf{A}(:,i)|}{\|\widehat{\mathbf{A}}(:,i)\|\,\|\mathbf{A}(:,i)\|} \tag{19}$$

where $\widehat{\mathbf{A}}$ is the estimated mixing matrix and $\mathbf{A}$ the true one. The coefficient c_i is 0 in the worst case and 1 for a perfect match. The estimation of the ith component was considered accurate enough if c_i was greater than 0.9. The number of components exceeding this level was termed as the *number of accepted components* (NAC).

Note that the columns in $\widehat{\mathbf{A}}$ are initially estimated in a random order. Therefore, the estimated columns in $\widehat{\mathbf{A}}$ were reordered based on the greedy matching in the following way: First the columns in $\widehat{\mathbf{A}}$ and $\mathbf{A}$ with the largest coefficient, c_1, were found. These column were removed from the matrices, and the next best pair of columns was found, which had the second largest coefficient c_2. The procedure was continued similarly, always using the remaining matrices where the previously matched columns had been removed.

The data were generated 100 times for each selection of simulation parameters. The averaged NAC was computed over each set of 100 runs and for each BSS method separately.

C. Measurements and the Analysis of the Measured Data

EEG data were recorded from three healthy right-handed subjects aged 25–30 years using a 60-channel TMS-compatible Nexstim eXimia EEG device with the sampling frequency of 1450 Hz. The subjects gave their written informed consents before the experiments. The studies were approved by the Ethics Committee of Helsinki University Hospital and they were in compliance with the Declaration of Helsinki.

Somatosensory evoked potentials (SEPs) to electrical right median nerve stimulation on thenar were measured (DS7A Digitimer Ltd., Welwyn Garden City, U.K.). The duration of each rectangular pulse was 100 μs. The current was 80% above the sensation threshold to evoke clear EEG responses without causing pain. The intervals between the consecutive stimuli were randomized between 2 and 3 s. 178 (subject S1) and 100 (subject S2) trials were collected. The averaged evoked EEG data of S1 are shown in Fig. 3.

To record TMS-evoked potentials (TEPs), biphasic magnetic pulses were delivered using the Nexstim eXimia TMS stimulator, a figure-of-eight coil, and an MRI-guided navigation system (all from Nexstim Ltd., Helsinki, Finland). The stimulation hot spot of the subject's (S3) right abductor pollicis brevis muscle in the primary motor cortex and the corresponding resting motor threshold (rMT) were determined using the Nexstim eXimia EMG system. Thereafter, 167 stimuli were delivered to this same target area with random intervals of 2–3 s and the stimulus intensity of 100% of the rMT.

In addition to the neural signals, the measured data contained both stimulus-independent artifactual elements (most notably due to eye blinks) and noise. Preprocessing was performed to remove these unwanted signals. The whole preprocessing protocol was as follows:

1) The data in the interval of -500 to 500 ms with respect to the stimuli were selected.
2) The stimulus-independent artifacts were eliminated using FastICA.
3) The datasets were filtered in the time/frequency domain: for S1 and S2, we used the SOUND algorithm (see Appendix A), and for S3, filtering with the pass band of 1–45 Hz was performed. We also performed spatial filtering by SOUND.
4) The average potential over the channels was set as the reference potential.
5) Compression by PCA was used to decrease the data dimensionality.
6) The MCMs were computed.
7) The MCMs were filtered by (13).

After the preprocessing, FFDiag was used to jointly diagonalize the MCMs within the time interval of 10–120 ms for the SEP data and 0–120 ms for the TEP data. The choice of using FFDiag was based on its good performance with simulated data and short computational time. For simplicity, autocovariance matrices were not included in the estimation here since, as discussed later, with the simulated data, AJD applied to MACMs did not convincingly improve the results even though the simulated noise was temporally white. The results

with the MACMs were also variable depending on the selected time lags, and it was not clear how they should be selected optimally.

1) Assessing the Feasibility of BSS With the Evoked Responses: The aim of analyzing the SEP data was to perform feasibility testing on the BSS of evoked data by cross validating between two different BSS methods that rely on similar statistical assumptions. To this aim, after step (iv), FastICA-ns was also applied on the data as described in [22]. FastICA-ns was selected here as the comparison method for MUCA based on its good performance with simulated data (see Section III).

We compared the mixing matrices found from the SEP datasets by ICA and MUCA. The MUCA- and ICA-estimated topographies were paired by greedy matching. The goodness of match c_{m} was measured as the absolute value of the normalized dot product between the matched topographies [as in (19)].

After topographically matching the components, we also compared the corresponding waveforms estimated by MUCA and ICA. This estimation is influenced by noise unrelated to the actual waveforms: In $\widehat{\mathbf{S}}(:,:,k) = \widehat{\mathbf{A}}^{-1}\mathbf{A}\mathbf{S}(:,:,k) + \widehat{\mathbf{A}}^{-1}\boldsymbol{\varepsilon}(:,:,k)$, there is the first noiseless part and the second componentwise-mapped noise part. Because of the noise, the waveform comparison is reasonable only when the component amplitude is considerably higher than its noise level. Therefore, we determined the latencies at which the waveform amplitudes peaked above the noise level and compared whether these latencies were similar for MUCA and ICA.

The high-amplitude latencies were determined as follows. The noise level was computed within the prestimulus interval of -400 to -10 ms with respect to the stimulus. During this baseline period, the mean and the standard deviation of each averaged waveform in $\widehat{\mathbf{S}}_{\mathrm{ave}}$ were computed. Then, during the poststimulus period, we identified the time intervals during which each averaged waveform deviated from its mean by more than three standard deviations. The latencies were set at the maxima/minima waveform amplitudes during the identified time intervals. These extrema were automatically found by locally fitting a third-order polynomial within each of the time intervals. The same approach of determining the latencies was applied to the variance waveforms $\widehat{\mathbf{S}}_{\mathrm{var}}$.

We also checked how stable the obtained mixing matrix estimate was when only a subset of trials was taken into consideration. The stability test was as follows. Half of the trials were randomly chosen. Thereafter, the covariance matrices were computed using the randomized trials, and MUCA was used to find the components. The obtained mixing matrix was compared with the mixing matrix retrieved from all-trials estimation by applying greedy matching of the topographies in the previously described fashion. As a measure of the goodness of match, the coefficient c_{m} was again computed according to (19). The randomization and the matching were performed 100 times to get a histogram for the c_{m} values.

2) Comparing the BSS Methods Using Source Localization: Evaluating the utility of different BSS methods in source localization using measured data is difficult because the true locations of the underlying components are not known accurately. To make this evaluation in a realistic manner, we formed TMS–EEG test data by mixing simulated responses with the measured highly artifactual TEP data. The waveforms of five components and 167 trials were simulated as described previously. The topographies were computed based on single dipoles in a three-layer spherical head model and using the standardized locations of the EEG electrodes [26]. The radii used for the concentric spheres were 81 mm (inner skull), 85 mm (outer skull), and 88 mm (scalp). The skull conductivity was set 1/100 of that of the skin and the brain tissues. The locations and the waveforms of the dipoles as well as the resulting mixed data containing both simulated and measured background EEG are shown in Fig. 4. We used (18) to compute the noise level, now referred to as the background EEG level, by setting the measured TEP signal, containing noise, artifacts, and hidden neural components, as $\boldsymbol{\varepsilon}$.

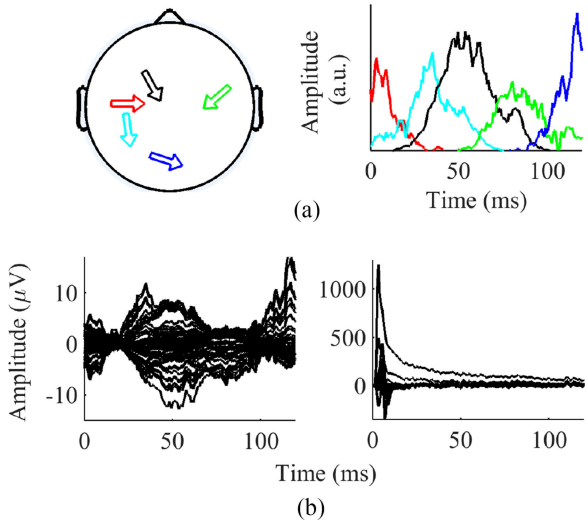

Fig. 4. Merging simulated data originating from five dipoles with measured TMS–EEG data. (a) Dipole locations and waveforms in one trial (colors correspond to each other). (b) Resulting simulated data trial (left) and the combination of the measured and the simulated EEG (right) with the background EEG level of 13.

The estimated topographies were paired with the simulated ones by greedy matching, whereafter single-dipole fitting was applied to the matched estimated topographies. We used the same spherical head model for both the simulation and the source localization. This "inverse crime" was allowed to guarantee zero localization error if the topography was perfectly uncovered by the used BSS method. Therefore, the measured error was solely due to any bias in the BSS solution. The NAC was set as the number of estimated topographies (out of five) having the dipole localization error of less than 0.5 cm. Due to the stochastic nature of the used BSS algorithms, the estimations were run 20 times over which the averaged NAC values were computed, at each chosen background EEG level separately.

III. Results

A. Simulation Results

Out of the compared BSS methods, MUCA with FFDiag and J-di yielded the most accurate BSS results at all noise levels, as

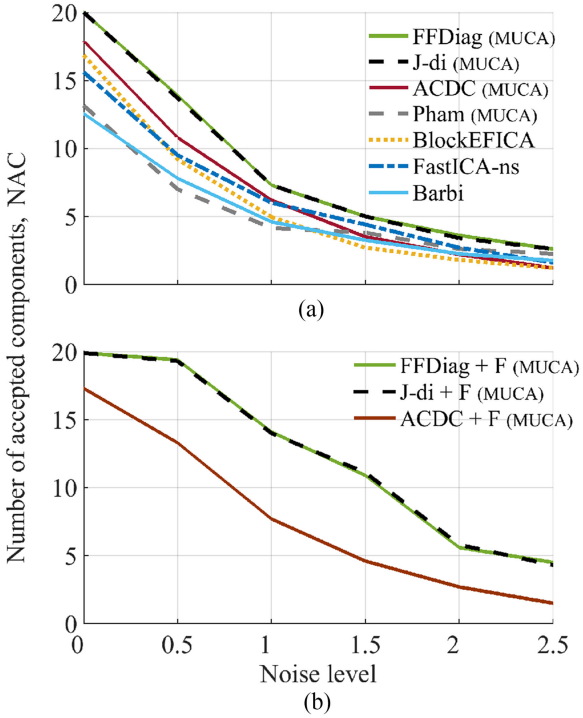

Fig. 5. NAC as a function of the noise level. (a) Comparison between the BSS methods. (b) Comparison between the AJD methods when applying MUCA to the filtered MCMs. The notation "+F" stands for filtering.

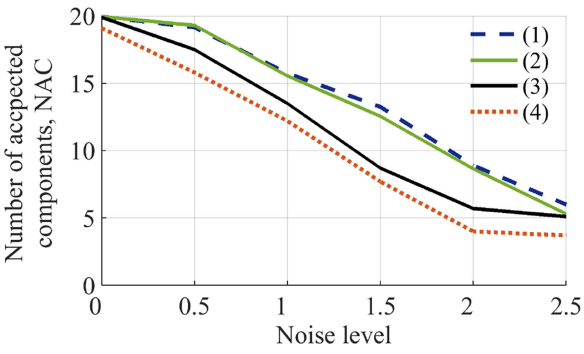

Fig. 6. Comparison between four different sets of target matrices for AJD: (1) MCMs and MACMs both filtered, (2) filtered MCMs only, (3) filtered MCMs and non-filtered MACMs, and (4) filtered MACMs.

shown in Fig. 5. The filtered MCMs could be used by MUCA with ACDC, FFDiag, and J-di, but not by the other methods [see Fig. 5(b)]. In the noiseless case, the filtering did not affect the results, but it became increasingly effective with larger noise levels. At NL = 1, MUCA with FFDiag/J-di applied to the filtered MCMs was able to uncover 14 components, whereas the other tested methods successfully found about 4–7 components. Apart from MUCA, the best-performing BSS methods were Block EFICA and FastICA-ns, which we use in the following comparisons along with MUCA.

All of the tested methods were suitable for nonstationary data. For demonstration, we computed the NAC values in the same simulation setting with SOBI and FastICA (in its usual form), which are both popular in EEG analysis. On average, SOBI was able to find only 4.0 and FastICA 5.1 components even when

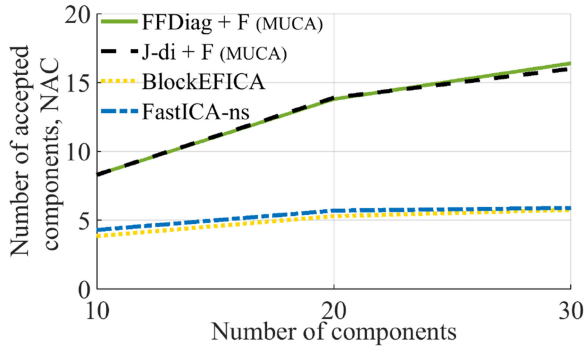

Fig. 7. Estimation accuracy with different numbers of components. The notation "+F" stands for filtering.

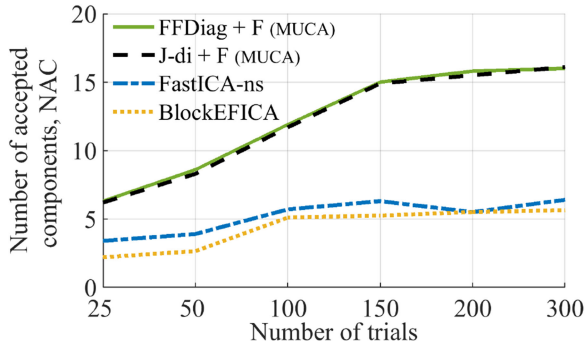

Fig. 8. Estimation accuracy as a function of the number of trials. The notation "+F" stands for filtering.

there was no measurement noise (NL = 0). This illustrates how the results deteriorate when using BSS methods designed for stationary data.

A similar comparison was made with autocovariance matrices. Including both MACMs and MCMs as target matrices for AJD (FFDiag) tended to give a minor improvement in the estimation accuracy compared to using the MCMs only. However, the results varied depending on the selected times lags. An example of the results is shown in Fig. 6, where the selected time lags were 20, 35, and 50. The filtered MACMs provided more accurate estimation as compared to using the nonfiltered MACMs.

MUCA with J-di and FFDiag performed better than the other compared methods when varying the number of the simulated components and trials, as depicted in Figs. 7 and 8, respectively. The MUCA results, with FFDiag and J-di, started to get worse when the number of trials decreased below 150 (see Fig. 8). Increasing the number of trials above 150 did not have a notable effect on the NAC of these methods. Overall, increasing the number of trials had a positive influence on the FastICA-ns and Block EFICA performances, although not as significant as with the other methods.

Increasing the condition number of the simulated mixing matrix worsened the estimation accuracy of all the tested algorithms, MUCA finding the components most accurately (see Fig. 9). When the condition number was one, MUCA found all components even at the presence of noise (NL = 1), whereas the other BSS methods found about 9 components out of 20.

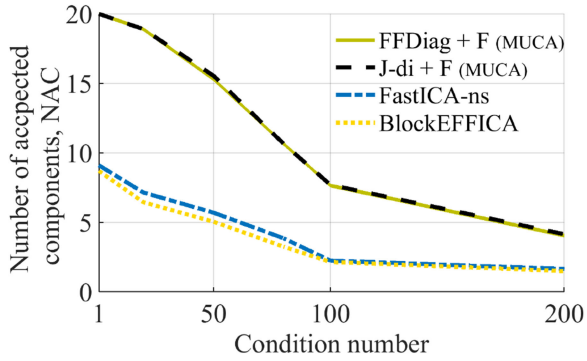

Fig. 9. Estimation accuracy at several condition numbers of the mixing matrix. The notation "+F" stands for filtering.

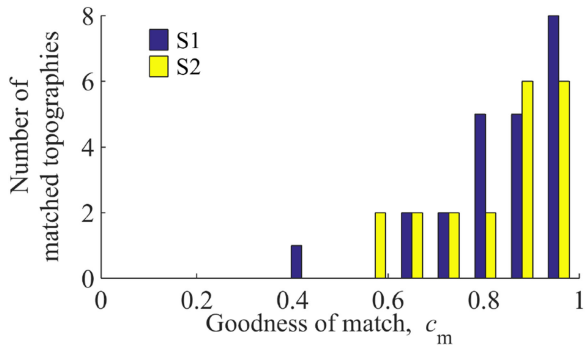

Fig. 10. Distributions of the goodness of match values, c_m, between the MUCA- and the ICA-estimated topographies.

B. Feasibility Results of the BSS Methods

The histograms of all the c_m values describing the similarity of the ICA and the MUCA estimates are illustrated in Fig. 10. The higher c_m values are more frequent than the lower ones, indicating the presence of several hidden sources which both MUCA and ICA can find. Using the S1 data, 23 components were estimated, out of which 17 had $c_m > 0.8$. With the S2 data, 14 out of 20 components had $c_m > 0.8$. The topographies and the waveforms of these components are illustrated in Appendix B. For both subjects, many of the topographies obtained by both ICA and MUCA appeared remarkably similar.

Comparing the time-courses shows that most of the latency estimates by ICA and MUCA fall close to each other (see Fig. 11). With the S1 data, the mean differences between the estimated latencies (by ICA and MUCA) were 2 and 1 ms for the averaged and the variance waveforms, respectively. For S2, these numbers were 3 and 5 ms. The standard deviations of the differences were 9 (S1) and 8 ms (S2) for the averaged waveforms, and 19 (S1) and 28 ms (S2) for the variance waveforms.

The stability of the estimated topographies was evaluated by taking half of the trials into account and comparing the results with the all-trials results. The histograms gathering the c_m values of 100 test comparisons are shown in Fig. 12. In the S1 data, there were more trials; the histogram has more counts close to $c_m = 1$. The mode is 0.93 and the mean is 0.85; 95% of the samples exceed 0.62. With the S2 data, the mode of c_m is 0.87, the mean is 0.80, and the tail away from 1 is longer; 95% of the values exceed 0.55.

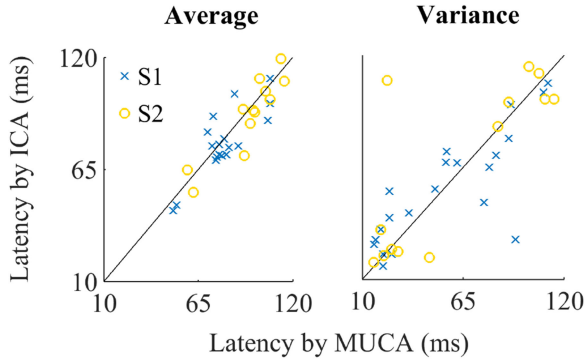

Fig. 11. Comparison of the latencies of the deflections in the average and variance waveforms estimated by MUCA and ICA.

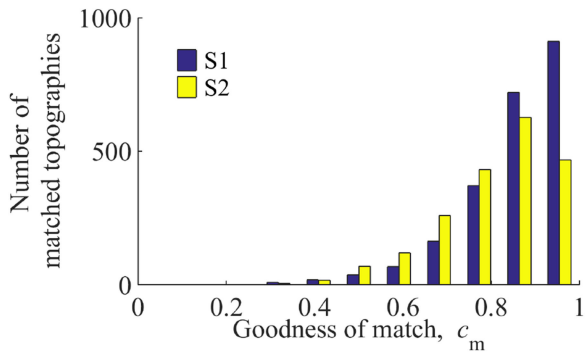

Fig. 12. Distributions of the goodness of match values, c_m, when half of the trials were randomized for MUCA 100 times. The mixing matrices were compared with the topographies obtained from the all-trials MUCA estimation.

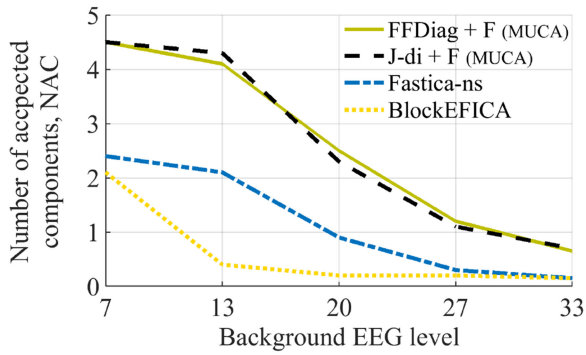

Fig. 13. Results of BSS applied to the TMS-evoked EEG data. NAC was the number of estimated components whose corresponding dipole localization errors were less than 0.5 cm. The notation "+F" stands for filtering.

C. Source Localization Results Based on the BSS Methods

Five simulated components were inserted in the TMS-evoked EEG as explained in Section II-C. MUCA, Block EFICA, and FastICA-ns were then used to estimate 30 components, among which the ones corresponding to the simulated dipoles were identified. MUCA with FFDiag and J-di applied to the filtered MCMs found the components most accurately, as shown in Fig. 13. At the background EEG levels of 7–13, NAC exceeded 4, decreasing rapidly at higher background EEG levels.

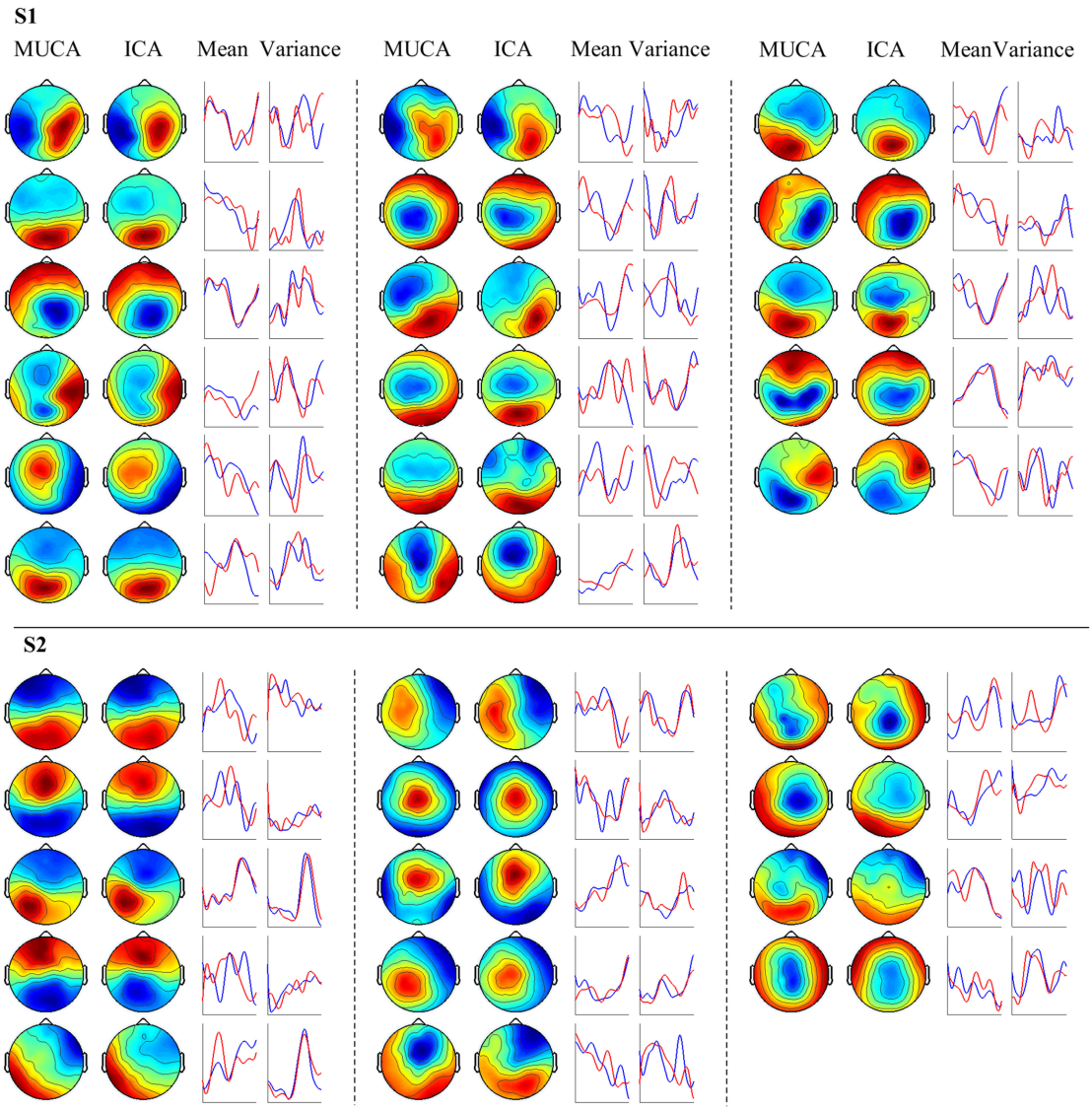

Fig. 14. Above the solid line: the best-matching MUCA- and ICA-estimated topographies ($c_{\mathrm{m}} > 0.8$) and the corresponding waveforms from the subject S1. The components are shown in three sections separated by dashed lines. Each section has four columns containing (from left to right) the MUCA-estimated topographies, the ICA-estimated topographies, the averaged time-courses, and the variance time-courses. Each row corresponds to one pair of matched components. In the waveform illustrations, blue indicates MUCA and red indicates ICA. The time interval of the waveforms is 10...120 ms, and the vertical scales are arbitrary. For visualization purposes, the time-courses have been bandpass filtered between 0.5 and 40 Hz and scaled to have unit norms. The warm and the cold colors in the topographies present positive and negative values, respectively. Below the solid line: The results from S2 are displayed as above.

IV. Discussion

The accuracy of MUCA as compared to the other BSS methods is possibly due to the simple underlying assumptions and the possibility to remove noise by filtering the covariance matrices, which makes AJD more robust to noisy data. Measured multitrial responses tend to have high noise level, and it is important to use efficient noise reduction. At the same time, one needs to be aware that the techniques that change the time-domain information of the data may also interfere with the required uncorrelatedness property. However, filtering the covariance matrices preserves this property. It is also noteworthy that MUCA does not require whitening, which is beneficial since the accuracy of whitening decreases with noisy data.

MUCA was proven tolerant to high TMS-evoked artifacts. This is partly because the trial-to-trial variability of these artifacts is relatively small although the averaged artifacts tend to be large. As the averaged response is subtracted from the trials in computing MCMs, only the intertrial variability is used by MUCA. This sort of measured data also tend to contain outlier signals, to which AJD-based BSS is known to be robust [8]. As indicated by the results, the robustness increases if the number of trials is high, preferably over 150.

Since the estimation of ill-conditioned mixing matrices is inaccurate, it would be useful to assess and/or reduce the condition number of the mixing matrix. When using PCA to compress the data dimensionality, the condition number of the estimated mix-

ing matrix can be computed as the ratio of the largest and the lowest selected eigenvalues. However, this is a highly heuristic approach, and it is usually difficult, or even impossible, only on the basis of the measured data to find the true condition number. If the sources, however, are all sufficiently strong and not very close to each other, the condition number of mixing matrix is generally not very high. If the mixing matrix is ill-conditioned due to large artifacts, one may use the suppression techniques suggested in [12] to improve the accuracy of the estimated topographies for source localization.

It is notable that stationary components cannot be found by MUCA. Stimulus-independent neural or artifact activity, such as eye blinks, is stationary. Therefore, we suggest that, prior to applying MUCA, one removes randomly occurring artifacts by some other form of BSS, like ICA [35].

Adding MACMs for AJD might enhance the BSS performance if the time lags could be chosen so that the noise would have close-to-zero autocovariance matrices and the hidden components diagonal ones. In practice, utilizing the MACMs proved difficult since, even with the simulated data having noiseless MACMs, the results improved only slightly as compared to MUCA. In measured data, the noise has autocorrelation properties, leading to noisy MACMs, which makes the BSS task more difficult.

Based only on the cross-validation results between MUCA and ICA, we cannot judge whether MUCA or ICA estimation was superior. However, the comparisons are an important feasibility test since they showed that these two methods, which are based on very different algorithms but similar assumptions, were capable of finding many highly alike components within the same SEP data. These findings suggest that evoked data contain momentary-uncorrelated components, many of which also appear independent.

BSS can also be based on some other property of the components than mutual independence or uncorrelatedness, such as sparseness in the time domain [19]. Generally, when using any BSS approach to evoked data, one should consider taking the nonstationarity of the data into account to improve the accuracy of the estimation. For example, in addition to sparseness, the components could be given an additional constraint of having similar latencies of maximum/minimum amplitude over the trials.

V. Conclusion

Event-related responses should be analyzed by methods that are designed for nonstationary data. All the methodologies presented in this work are specifically tailored for nonstationary event-related responses, which yield better estimates for the components than methods designed for stationary data.

MUCA is a novel method tailored for finding uncorrelated components from event-related multitrial EEG or MEG data. As compared to several other BSS methods, MUCA was shown more accurate at the presence of noise and artifacts.

The cross-validation results between ICA and MUCA support the idea that the evoked responses contain uncorrelated and independent processes, indicating that correctly applied BSS is a meaningful analysis tool with these data.

Appendix A
Outline of the Sound Algorithm

Here, we used the data-driven version of SOUND [24], based on least-squares estimation, to remove noise that is uncorrelated over channels or time. n samples of m-variate data are collected into an $(m \times n)$ matrix $\mathbf{Y}$. The sample correlation matrix $\mathbf{R}_Y$ of $\mathbf{Y}$ is used for estimating the noiseless data.

We use the notation $\mathbf{k}_i = [1, \ldots, i-1, i+1, \ldots, m]$ for building an index vector. The estimate for each noiseless data matrix element is then obtained by

$$\widehat{\mathbf{Y}}(i,j) = \mathbf{R}_Y(i,\mathbf{k}_i)\mathbf{R}_Y(\mathbf{k}_i,\mathbf{k}_i)^{-1}\mathbf{Y}(\mathbf{k}_i,j)\,. \qquad (20)$$

For temporal filtering of one channel i_c at a time, we set $\mathbf{Y}(:,j) = \mathbf{X}(i_\mathrm{c},:,j),\ j = 1, \ldots, N_\mathrm{tr}$. Now, applying (20) removes noise that is uncorrelated over time instants. Second, for spatial filtering at a time index i_t, $\mathbf{Y}(:,j) = \mathbf{X}(:,i_\mathrm{t},j),\ j = 1, \ldots, N_\mathrm{tr}$, after which (20) removes the part of the signal (noise) that is uncorrelated over the channels.

Appendix B
Illustrations of the Topographies and the Waveforms Estimated From the Measured SEP Data

In Fig. 14, we illustrate the best-matching components estimated by ICA and MUCA based on the somatosensory evoked potentials of subjects S1 and S2.

Acknowledgement

The authors would like to thank T. Mutanen for sharing the measured somatosensory evoked potentials for this work.

References

[1] B. Afsari, "What can make joint diagonalization difficult?," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Honolulu, HI, USA, 2007, pp. III-1377–III-1380.

[2] A. Belouchrani *et al.*, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 434–444, Feb. 1997.

[3] A. Belouchrani and M. G. Amin, "Blind source separation based on time-frequency signal representations," *IEEE Trans. Signal Process.*, vol. 46, no. 11, pp. 2888–2897, Nov. 1998.

[4] J.-F. Cardoso,, "Source separation using higher order moments," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Glasgow, U.K., 1989, pp.2109–2112.

[5] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *IEE Proc.-F—Radar Signal Process.*, vol. 140, no. 6, pp. 362–370, Dec. 1993.

[6] S. Choi and A. Cichocki, "Blind separation of nonstationary sources in noisy mixtures," *Electron. Lett.*, no. 36, vol. 9, pp. 848–849, Apr. 2000.

[7] S. Choi *et al.*, "Second order nonstationary source separation," *J. VLSI Signal Process.*, vol. 32, no. 1–2, pp. 93–104, Aug. 2002.

[8] M. Congedo *et al.*, "On the blind source separation of human electroencephalogram by approximate joint diagonalization of second order statistics," *Clin. Neurophysiol.*, vol. 119, no. 12, pp. 2677–2686, Dec. 2008.

[9] S. Dégerine and E. Kane, "A comparative study of approximate joint diagonalization algorithms for blind source separation in presence of additive noise," *IEEE Trans. Signal Process.*, vol. 55, no. 6–2, pp. 3022–3031, Jun. 2007.

[10] M. Grosse-Wentrup and M. Buss, "Multiclass common spatial patterns and information theoretic feature extraction," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 8, pp. 1991–2000, Aug. 2008.

[11] M. Hamidi *et al.*, "Brain responses evoked by high-frequency repetitive transcranial magnetic stimulation: An event-related potential study," *Brain Stimulation*, vol. 3, no. 1, pp. 2–17, Jan. 2010.

[12] J. C. Hernandez-Pavon *et al.*, "Uncovering neural independent components from highly artifactual TMS-evoked EEG data," *J. Neurosci. Methods*, vol. 209, no. 1, pp. 144–157, Jul. 2012.
[13] K. E. Hild and S. S. Nagarajan, "Source localization of EEG/MEG data by correlating columns of ICA and lead field matrices," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 11, pp. 2619–2626, Nov. 2009.
[14] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 626–634, May 1999.
[15] A. Hyvärinen *et al.*, "Principal component analysis and whitening" in *Independent Component Analysis*. Hoboken, NJ, USA: Wiley, 2004, pp. 125–144.
[16] T.-P. Jung *et al.*, "Analysis and visualization of single-trial event-related potentials," *Human Brain Mapping*, vol. 14, no. 3, pp. 166–185, Nov. 2001.
[17] Z. Koldovský *et al.*, "Blind separation of piecewise stationary non-Gaussian sources," *Signal Process.*, vol. 89, no. 12, pp. 2570–2584, Dec. 2009.
[18] R. J. Korhonen *et al.*, "Removal of large muscle artifacts from transcranial magnetic stimulation-evoked EEG by independent component analysis," *Med. Biol. Eng. Comput.*, vol. 49, no. 4, pp. 397–407, Apr. 2011.
[19] Y. Li *et al.*, "Sparse representation for brain signal processing: A tutorial on methods and applications," *IEEE Signal Process. Mag.*, vol. 31, no. 3, pp. 96–106, May 2014.
[20] S. Makeig *et al.*, "Blind separation of auditory event-related brain responses into independent components," *Proc. Nat. Acad. Sci USA*, no. 94, vol. 10, pp. 10979–10984, Sep. 1997.
[21] J. Metsomaa *et al.*, "Momentary-uncorrelated components analysis, MUCA—BSS of evoked M/EEG," in *Proc. BIOMAG Conf.*, Halifax, NS, Canada, Aug. 24–28, 2015.
[22] J. Metsomaa *et al.*, "Multi-trial evoked EEG and independent component analysis," *J. Neurosci. Methods*, vol. 228, pp. 15–26, May 2014.
[23] K.-R. Müller *et al.*, "Machine learning techniques for brain–computer interfaces," *Biomed. Eng.*, vol. 49, no. 1, pp. 11–22, Dec. 2004.
[24] T. P. Mutanen *et al.*, "Automatic and robust rejection of sensor noise in EEG: The SOUND algorithm," to be published.
[25] J. Onton *et al.*, "Imaging human EEG dynamics using independent component analysis," *Neurosci. Biobehav. Rev.*, vol. 30, no. 6, pp. 808–822, Dec. 2006.
[26] R. Oostenveld and P. Praamstra, "The five percent electrode system for high-resolution EEG and ERP measurements," *Clin. Neurophysiol.*, vol. 112, no. 4, pp. 713–719, Apr. 2001.
[27] D-T. Pham and J-F. Cardoso, "Blind separation of instantaneous mixtures of nonstationary sources," *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 1837–1848, Sep, 2001.
[28] A. Souloumiac, "Nonorthogonal joint diagonalization by combining givens and hyperbolic rotations," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2222–2231, Jun. 2009.
[29] A. Souloumiac, "Blind source detection and separation using second order non-stationarity," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Detroit, MI, USA, 1995, pp. 1912–1915.
[30] A. C. Tang *et al.*, "Contrasting single-trial ERPs between experimental manipulations: Improving differentiability by blind source separation," *Neuroimage*, vol. 29, no. 1, pp. 335–346, Jan. 2006.
[31] A. C. Tang *et al.*, "Validation of SOBI components from high-density EEG," *Neuroimage*, vol. 25, no. 2, pp. 539–553, Apr. 2005.
[32] P. C. Tichavský, "Blind separation of underdetermined linear mixtures based on source nonstationarity and AR (1) modeling," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Shanghai, China, 2016, pp. 4323–4327.
[33] L. Tong, "AMUSE: A new blind identification algorithm," in *Proc. IEEE Int. Symp. Circuits Syst.*, New Orleans, LA, USA, 1990, pp. 1784–1787.
[34] R. Vigário "Independent component approach to the analysis of EEG and MEG recordings," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 5, pp. 589–593, May 2000.
[35] R. N. Vigário, "Extraction of ocular artefacts from EEG using independent component analysis," *Electroencephalogr. Clin. Neurophysiol.*, vol. 103, no. 3, pp. 395–404, Sep. 1997.
[36] A. Yeredor, "Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation," *IEEE Trans. Signal Process.*, vol. 50, no. 7, pp. 1545–1553, Jul. 2002.
[37] A. Ziehe *et al.*, "A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation," *J. Mach. Learn. Res.*, vol. 5, pp. 777–800, Jul. 2004.

Authors' photographs and biographies not available at the time of publication.