Tensor Based Singular Spectrum Analysis for Automatic Scoring of Sleep EEG

Samaneh Kouchaki, *Student Member, IEEE*, Saeid Sanei, *Senior Member, IEEE*, Emma L. Arbon, and Derk-Jan Dijk

***Abstract*—A new supervised approach for decomposition of single channel signal mixtures is introduced in this paper. The performance of the traditional singular spectrum analysis algorithm is significantly improved by applying tensor decomposition instead of traditional singular value decomposition. As another contribution to this subspace analysis method, the inherent frequency diversity of the data has been effectively exploited to highlight the subspace of interest. As an important application, sleep electroencephalogram has been analyzed and the stages of sleep for the subjects in normal condition, with sleep restriction, and with sleep extension have been accurately estimated and compared with the results of sleep scoring by clinical experts.**

***Index Terms*—Electroencephalogram (EEG), empirical mode decomposition (EMD), single channel source separation, singular spectrum analysis (SSA), sleep, tensor factorization.**

I. Introduction

In BIOMEDICAL signal processing, mixtures are often recorded from the source signals. Hence, our task is to unmix them and retrieve the underlying sources. For instance, in the analysis of electroencephalogram (EEG) signals, it is necessary to extract neurophysiologically meaningful information in applications such as characterizing event related potentials, brain–computer interfacing (BCI), seizure detection, and sleep analysis. In multichannel data, this problem is efficiently handled by employing blind source separation (BSS) techniques, which unmix the given signal mixtures into their constituent sources [1]–[3]. BSS is traditionally applied when the number of sources is equal or less than the number of electrodes. It fails for single channel or generally underdetermined recordings. Some methods have also been proposed for multichannel underdetermined source separation [4], [5].

However, there are several applications where just one channel is used, for example, restoring the electromyogram (EMG) signal contaminated by electrocardiogram (ECG) artefact [6], [7], single channel deep brain recordings, where the neuronal spikes are not easy to separate from noise [8], many BCI applications, and different sleep stages with various dominant frequency bands where each frequency band is prevailing in a special recorded channel [9], [10]. In such applications, even in the case of multichannel recording, the desired information can be retrieved just from some special recorded channels. On the other hand, in the recordings with a small number of channels, the separation using BSS is often poor due to the system being underdetermined.

Several methods have been proposed for single channel source separation. The extension of independent component analysis (ICA) to single channel signals or single channel ICA (SCICA), has been proposed in [11]. This algorithm requires stationarity of the data, independency of the sources, and also disjoint frequency domain to perform well. These conditions do not always hold in all applications.

Another approach is to create multichannel data from a single channel signal. Then, ICA is used to decompose the signal into independent components. In several methods, a single channel signal is decomposed into multiple spectral modes using wavelet transform (WT) or empirical mode decomposition (EMD) [12], [13]. In both cases, the signal is first decomposed using wavelet or EMD and then, ICA is applied to the resulting components. The problem with WT is that the extracted components vary significantly with changes in the mother wavelet. However, EMD automatically decomposes the signal to its constituent components intrinsically matched with the signal structure. Moreover, EMD does not use any fixed or predefined function in the decomposition stage. However, the extracted components do not necessarily correspond to any meaningful data.

Singular spectrum analysis (SSA) is another powerful method for analysing real-valued time series [14]. It is becoming a popular method in various areas such as mathematics, economics, and biomedical engineering. Basically, SSA decomposes a time series into a number of interpretable components with different subspaces, such as trend and noise, and can be applied to any time series with complex structure [15]. For instance, for decades SSA has been used for both trend detection and prediction in financial time series [14], [15].

Recently, SSA has been employed in biomedical signal processing applications such as separation of EMG and ECG [16] and restoring lung sound from heart sound [17]. A supervised SSA has been attempted in [18] to detect spikes from noisy signals. This method however does not exploit the narrow band property of cyclic data. A supervised approach can incorporate the properties of the desired signal component into the SSA decomposition and reconstruction algorithm. For narrowband signals, the overall strategy is depicted in Fig. 1. In EEG, the brain rhythms manifest themselves as narrow frequency band com-

Manuscript received December 18, 2013; revised May 21, 2014; accepted May 24, 2014. Date of publication June 13, 2014; date of current version January 06, 2015.

S. Kouchaki and S. Sanei are with Faculty of Engineering and Physical Sciences, University of Surrey, GU2 7YW Guildford, U.K.

E. L. Arbon and D. Dijk are with Surrey Sleep Research Centre, Faculty of Health and Medical Sciences, University of Surrey, GU2 7XH Guildford, U.K.

Digital Object Identifier 10.1109/TNSRE.2014.2329557

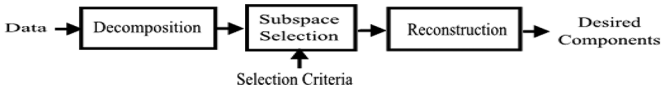

Fig. 1. Supervised singular spectrum analysis where specific criteria should be used to achieve a good result.

ponents. As a good example, sleep is a dynamic process consisting of different stages with different neural activity levels. Each stage is characterized by a distinct set of physiological and neurological features and dominant frequency band. Therefore, sleep can be considered as one good application of this method. As another application, detection of event-related desynchronization and synchronization (ERD/ERS) is important in BCI to mark the onset of body movement. The variations in alpha and beta can be accurately measured following application of SSA.

However, the traditional SSA has some drawbacks. First, it cannot work in nonstationary environments as it works with a covariance matrix of data. Many real signals particularly physiological signals such as EEG are nonstationary. Sleep EEG has a nonstationary structure since its statistical properties in both time and frequency change during the stages of sleep. Furthermore, in SSA selecting the desired subgroup of eigenvalues is still an open problem. As the grouping procedure plays an important role in the reconstruction stage, some techniques should be developed for proper subspace grouping. Hence, the purpose of this paper is to extend the SSA algorithm and theory developed in [14] to separate the single channel data more effectively by using the signal properties. In addition, the data nonstationarity is effectively exploited by jointly factorizing the signal segments. Therefore, in this work, 1-D arrays are converted to 3-D arrays whereby the decomposition stage of the SSA problem is converted to tensor factorization problem. Tensor representation is a way to algebraically describe the datasets by preserving their multi-way structures.

In addition, we show that the desired eigenvalues in the SSA reconstruction stage can be selected automatically and adaptively by exploiting the frequency diversity of the sources.

The remainder of the paper is structured as follows. In Section II, SSA and other related techniques are explained and the proposed method is introduced. In Section III, the results of applying the proposed method to both real and synthetic data are shown. Finally, Section IV concludes the paper.

II. Existing Single Channel Source Separation Methods

A. Single Channel ICA

This algorithm is an extension of ICA to single channel data [11]. In this method, the single channel signal $\mathbf{x}$ is broken up into a sequence of contiguous blocks to be treated as a sequence of vector observations, $\mathbf{x}(k) = [x(k\tau), \ldots, x(k\tau + p - 1)]^T$, where p is the length of each vector obtained by overlapping windowing of the original signal with length $k\tau + p - 1$, τ indicating a time delay, and k is the block index. Then, matrix $\mathbf{X}$ is formed from these data segments. Actually, the performance of this algorithm significantly depends on these parameters. Next, in this method, fastICA algorithm [19] is applied to this matrix to derive the mixing and unmixing matrices as $\mathbf{M}$ and $\mathbf{W}$. Extracting the particular source of interest can be achieved by filtering the signal with the corresponding row of the unmixing matrix $\mathbf{W}$. The filtering is obtained by replacing all the rows in $\mathbf{W}$ with zero except for the desired row, and multiplying it by the signal.

FastICA is used to estimate the ummixing matrix $\mathbf{W}$ as an inverse or the pseudo inverse of $\mathbf{H}$. Therefore, the multiplication of the extracted source $s_i(t)$ using the ith column of the matrix $\mathbf{H}$ produces the ith source waveform. The other sources can then be estimated in a deflation manner.

B. EEMD-ICA

In this method, EMD decomposes the time series into intrinsic subsignals with well-defined instantaneous frequencies called intrinsic mode functions (IMFs). Therefore, each time series $\mathbf{x}$ can be represented as a sum of its IMFs, $\mathbf{x} = \sum_{i=1}^{z} \mathbf{f}_i + \mathbf{e}_z$ where z is the number of IMFs, $\mathbf{f}_i$ is the ith IMF, and $\mathbf{e}_z$ is the residual obtained for zth IMF [20]. The IMFs are ordered in terms of their frequencies. In order to add more robustness against noise, ensemble EMD (EEMD) was introduced in [21]. The algorithm extracts the IMF set for an ensemble of trials obtained from the noisy version of the signal with independent, identically distributed white noise of the same standard deviation. In this way, the noisy components are expected to be cancelled out. However, adjusting the EEMD parameter to cancelling the noise out is not very easy and depends on the noise level and frequency content of the signal.

Then, FastICA is applied to the whole set of IMFs and the independent components (ICs) and the mixing and demixing matrices, $\mathbf{H}$ and $\mathbf{W}$ are estimated. Then, the signal is reconstructed by selecting the ICs of interest and projecting them into the space of mixture signals using the estimated mixing matrix to derive a new IMF set, in which only the ICs of interest are present. Finally, by adding the newly derived IMF set, the desired source can be recovered from the original signal [13].

C. Singular Spectrum Analysis

SSA is a subspace decomposition algorithm which consists of two stages: decomposition and reconstruction. The first step involves embedding followed by singular value decomposition (SVD). In the embedding stage a vector $\mathbf{x}$ of length n is mapped to an $l_1 \times l$ matrix $\mathbf{Y}$

$$\begin{aligned}\mathbf{Y} &= [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_l], \\ &= \begin{pmatrix} x_1 & x_2 & \cdots & x_l \\ x_2 & x_3 & \cdots & x_{l+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{l_1} & x_{l_1+1} & \cdots & x_{l+l_1-1} \end{pmatrix}\end{aligned} \tag{1}$$

where $\mathbf{x}_k = [x_k, x_{k+1}, \ldots, x_{k+l-1}]^T \epsilon \Re^l, l_1 = n - l + 1, l$ is the window length, and $(\cdot)^T$ stands for transpose. l should be large enough to contain the information regarding the data variation.

Next, the SVD of this matrix is computed as

$$\mathbf{Y} = \sum_{i=1}^{d} \mathbf{Y}_i = \sum_{i=1}^{d} \sqrt{\lambda_i} \mathbf{u}_i \mathbf{v}_i^T \tag{2}$$

where λ_i is the ith eigenvalue of covariance matrix $\mathbf{Y}\mathbf{Y}^T$, $\mathbf{u}_i$ is the corresponding eigenvector, d is the total number of eigenvalues, and $\mathbf{v}_i = \mathbf{Y}^T\mathbf{u}_i/\sqrt{\lambda_i}$.

In the next stage, first, the elementary matrices from the previous stage are grouped into several submatrices

$$\mathbf{Y} = \sum_{q=1}^{Q} \hat{\mathbf{Y}}_q \tag{3}$$

where Q determines the total number of groups, index q refers to qth subgroup of eigenvalues, and $\hat{\mathbf{Y}}_q$ indicates the sum of $\mathbf{Y}_i$ within group q. Finally, the resulting matrix is transformed into the form of a Hankel matrix that can be converted to a time series. The Hankelization operator $\mathcal{H}$ for an $i \times j$ matrix $\mathbf{Y}$ is obtained as

$$\begin{aligned}
\mathcal{H}\mathbf{Y} &= \begin{pmatrix} \tilde{y}_1 & \tilde{y}_2 & \cdots & \tilde{y}_i \\ \tilde{y}_2 & \tilde{y}_3 & \cdots & \tilde{y}_{i+1} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{y}_j & \tilde{y}_{j+1} & \cdots & \tilde{y}_{i+j-1} \end{pmatrix} \\
\tilde{y}_k &= 1/num(D_k) \sum_{\hat{i},\hat{j}\epsilon D_k} y_i, j \\
D_k &= \left\{ (\hat{i},\hat{j}) : 1 \le \hat{i} \le i, 1 \le \hat{j} \le j, \hat{i}+\hat{j} = k+1 \right\}
\end{aligned} \tag{4}$$

where $num(D_k)$ refers to the number of combinations of $(\hat{i},\hat{j})$ which makes $\hat{i}+\hat{j} = k+1$.

III. Tensor-Based SSA

Application of SSA to real data does not exploit the inherent nonstationarity and therefore may fail in actual data decomposition. Tensor-based SSA (TSSA) is a robust solution to this problem. Just as in SSA, the first stage of TSSA includes an embedding operation followed by a tensor decomposition method instead of SVD. In the embedding stage a 1-D time series $\mathbf{x}$ with length n is mapped into tensor $\underline{\mathbf{X}}$.

To do that, first $\mathbf{x}$ is segmented using a nonoverlapping window of size l and a $\lfloor n/l \rfloor \times l$ matrix $\hat{\mathbf{X}}$ is obtained from $\mathbf{x}$

$$\begin{aligned}
\hat{\mathbf{X}} &= \begin{pmatrix} x_1 & x_2 & \cdots & x_l \\ x_{l+1} & x_{l+2} & \cdots & x_{2l} \\ \vdots & \vdots & \ddots & \vdots \\ x_{(I-1)l} & x_{(I-1)l+1} & \cdots & x_{Il} \end{pmatrix} \\
I &= \lfloor n/l \rfloor.
\end{aligned} \tag{5}$$

Then, this matrix is converted to tensor $\underline{\mathbf{X}}$ as demonstrated in Fig. 2 by considering each slab of the tensor as a windowed version of $\hat{\mathbf{X}}$. In this work, the segmentation is performed in one direction, but it can be extended to move the window in both directions. Our way of converting a matrix to tensor can be explained by the following equation:

$$\begin{aligned}
\underline{\mathbf{X}}_{i::} &= \hat{\mathbf{X}}\,(:, (j-1)o+1 : (j-1)o+l_1) \\
j &= 1, 2, \ldots, J, J = \lfloor (l-l_1))/o+1 \rfloor \\
i &= 1, 2, \ldots, I
\end{aligned} \tag{6}$$

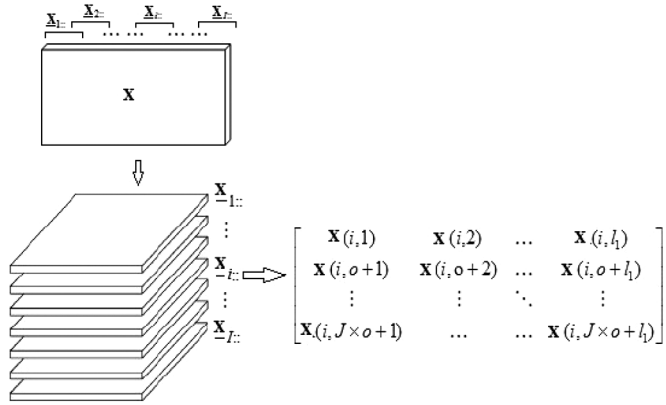

Fig. 2. $I \times K$ matrix $\mathbf{X}$ is converted to tensor $\underline{\mathbf{X}}$ where J is the number of segments obtained by window size l_1 and overlapping interval $l_1 - o$.

where $l_1 - o$ is the overlapping size between the successive windows and l_1 is the window size.

At this stage, we have a 3-D tensor to be decomposed. Parallel factor analysis (PARAFAC) as a canonical decomposition algorithm is used here. It can be considered as a generalization of bilinear PCA [22], [23]. The fundamental expression of the PARAFAC technique is given as [24], [25]

$$x_{ijk} = \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr} + e_{ijk} \tag{7}$$

where x_{ijk} is the (i,j,k)th element in the three-way data set, R is the number of common components or model order, a_{ir}, b_{jr}, and c_{kr} are the elements in $\mathbf{A}, \mathbf{B}$, and $\mathbf{C}$, respectively, and e_{ijk} is the residual term.

As a popular option, alternating least squares (ALS) optimization method can be used to fit the PARAFAC model [26]. In this convex iterative approach, $\mathbf{B}$ and $\mathbf{C}$ are fixed to solve for $\mathbf{A}$; then, $\mathbf{A}$ and $\mathbf{C}$ are fixed to solve for $\mathbf{B}$, and next $\mathbf{A}$ and $\mathbf{B}$ are fixed to solve for $\mathbf{C}$ in an alternating fashion until reaching some convergence. One important characteristic of PARAFAC is that it is convex and performs well in certain underdetermined cases where the number of sources is comparable with tensor rank [27]. Moreover, by segmenting and decomposing the signal in this way, the length of signal being processed at one time can increase.

After this stage, m disjoint subsets of indexes I_k are specified. Selecting a proper subgroup of components has an important impact on the final result. Here, an adaptive grouping is proposed as explained in the next part. If I_ks for $k = 1, \ldots, m$ indicate the group labels, we have

$$\begin{aligned}
&I_1 \cup I_2 \cup \ldots \cup I_m = \{1, 2, \ldots, R\} \\
&\underline{\mathbf{X}} = \sum_{i=1}^{m} \underline{\mathbf{X}}_{I_i}, \\
&\underline{\mathbf{X}}_K = \sum_{k\epsilon K} \mathbf{A}_k \circ \mathbf{B}_k \circ \mathbf{C}_k
\end{aligned} \tag{8}$$

where $\circ$ denotes outer product and $\underline{\mathbf{X}}_K$ indicates the tensor obtained from the I_kth subgroup. Finally, a reconstruction of the

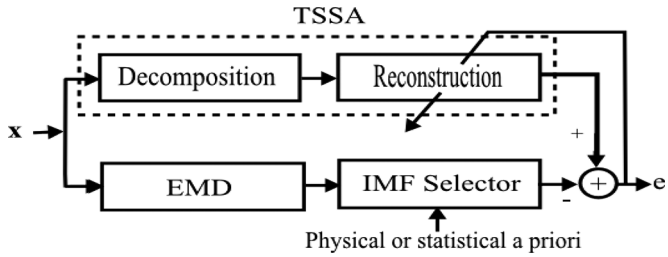

Fig. 3. Block diagram of the single channel source separation system using an adaptive procedure for selecting the desired subspace. This is carried out by tuning a set of weights governed by the EMD process.

original signal can be obtained by Hankelization which is performed by converting the signal segments to a Hankel matrix

$$\hat{\underline{\mathbf{X}}}_{I_i::} = \mathcal{H}\underline{\mathbf{X}}_{I_i::}. \tag{9}$$

Then, the data is reconstructed by one-to-one correspondence.

IV. TSSA-EMD

To employ SSA or TSSA for source separation, the corresponding subspace of the desired signal should be identified. The grouping process is the main problem of the SSA algorithm and it has a significant impact on the final results. The groups are often characterized by their statistical or physical properties. When applied to brain signals, the brain rhythms are well characterized by their frequency properties. Therefore, in this paper, EMD is used to supervise the TSSA-based subspace decomposition. EMD provides a good way to identify the number of frequency components within each subspace. So, EMD is used to select the subgroup of the desired signal as can be seen in Fig. 3. In other words, a number of IMFs falling within a particular frequency band are selected according to some pre-set criteria. For instance, the components that have maximum power in the desired frequency bands can be selected and used to more accurately group the eigentriples. Then, by assuming the Hankel matrix of this signal as $\underline{\mathbf{F}} = \{f_{ijk}\}$, the correct eigenvalue group is selected by minimizing the following objective function:

$$J(g_{ijk}) = \|f_{ijk} - \sum_{r=1}^{R} g_{ijk}a_{ir}b_{jr}c_{kr}\|^2 \tag{10}$$

where $\|\cdot\|$ denotes Frobenius norm, R is the number of common components, a_{ir}, b_{jr} and c_{kr} are the tensor factors, and $\underline{\mathbf{G}} = \{g_{ijk}\}$ is a superdiagonal tensor of adaptive weights. As the objective function is convex, it can be minimized using different methods.

Here, a subgradient method which is an iterative method for solving convex optimization problems is employed to obtain the optimal value of g_{ijk}

$$g_{ijk}^{\text{new}} = g_{ijk}^{\text{old}} - \mu\nabla_{g_{ijk}^{\text{old}}} J \tag{11}$$

where

$$\nabla_{g_{ijk}^{\text{old}}} J = -2\left(f_{ijk} - \sum_{r=1}^{R} g_{ijk}^{\text{old}}a_{ir}b_{jr}c_{kr}\right)\sum_{r=1}^{R} a_{ir}b_{jr}c_{kr} \tag{12}$$

where μ is the step size set empirically. Other optimization methods can also be used with negligible change in the performance.

V. Experimental Results and Discussion

To evaluate the performance of the proposed method in biomedical single channel signal processing and compare its performance with the other available methods, as outlined in previous sections, a number of experiments were performed using both simulated and real data. The first section explains the parameter setting and evaluation criterion; then, different methods are tested on synthetic data and the results are discussed in the next part; finally, the second section shows the results for real sleep EEG and the comparison with two conventional methods for sleep processing.

A. Simulations

Each of the generated signals is a mixture of two signals

$$x(t) = s(t) + \beta e(t) \tag{13}$$

where $s(t)$ indicates the signal to be extracted, $e(t)$ is the unwanted signal, which is considered to be normalized noise, β indicates the noise level and $x(t)$ is the input to the algorithms being examined. Here, in the first part, a noisy signal is generated where the noise variance changes with time in order to add nonstationarity to the data. In other words, we generated a narrowband signal and mixed it with nonstationary noise. Also, several signals which are mixtures of this signal with different noise levels are generated. Signal-to-noise ratio (SNR) in terms of root mean squared (rms) is used as a measure for noise level which is adjusted by changing β

$$\text{SNR} = \text{rms}(s(t))/\beta^2. \tag{14}$$

Sleep data can be considered as a real example for these signals. During sleep, the brain goes through several psychophysiological states characterized by a different frequency band and the corresponding signals corrupted by noise.

Then, the simulation performance is expressed in terms of rms as follows:

$$\text{RMSE} = \frac{\text{rms}(s(t) - \hat{s}(t))}{\text{rms}(s(t))} \tag{15}$$

where $\hat{s}(t)$ indicates the estimated source and RMSE is the root mean squared error.

1) Parameter Settings for Different Methods: In all the simulations, the number of independent components was set to 5. K and τ were set, respectively, to 10 and 1 in SCICA. The number of ensembles in EEMD was set as 2 and between 9 to 12 IMFs were usually produced using EMD. These parameters are selected according to [13]. The SSA window length l was set to 200 in both experiments. In addition, l and l_1 were considered respectively as 200 and 50 in TSSA. The window length is related to 1/Bandwidth of the data and practically we make sure that it is not less than 2 or 3 times this value. On the other hand we will avoid large values in order to keep the computational complexity low.

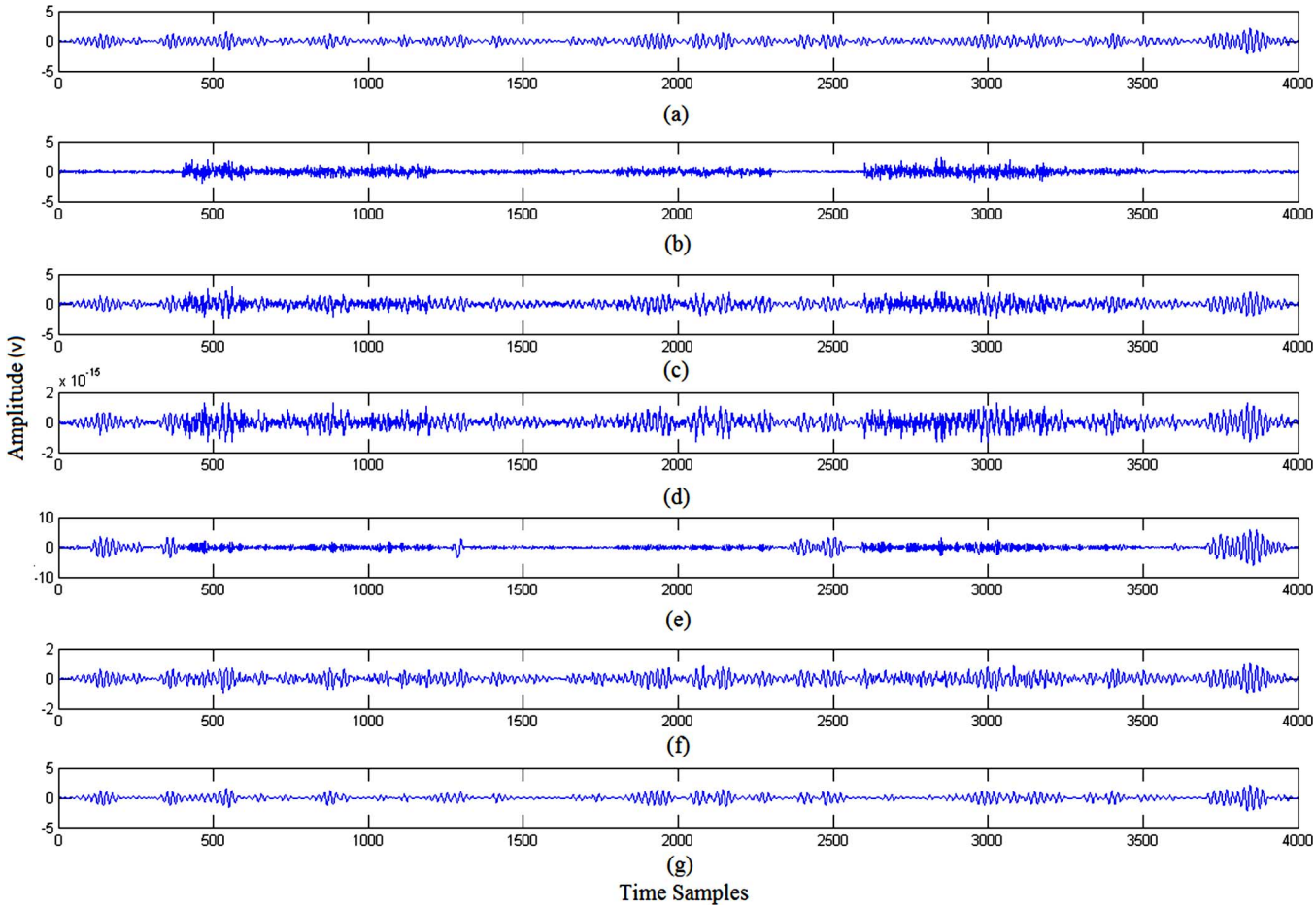

Fig. 4. Results of separation (denoising) of the narrowband signal in (a) from nonstationary noise in (b) by having noisy mixture with $\mathrm{SNR} \simeq 0$ dB in (c) using (d) SCICA, (e) EEMD-ICA, (f) SSA, and (g) TSSA-EMD.

TABLE I
COMPARISON BETWEEN RMSE(%) VALUES OF THE PROPOSED METHOD (TSSA), SSA, SCICA, AND EEMD-ICA OBTAINED BY (14) FOR THE SIMULATED DATA

Method / RMSE	Proposed	SCICA	EEMD-ICA	SSA
Narrowband signal	8.40	87.13	56.83	24.3

2) Narrowband Simulated Data in the Presence of Nonstationary Noise: An $\mathrm{N} = 4000$ sample narrowband signal was generated. Nonstationary Gaussian noise was added to this signal whereby its variance changes with time. In this way, the effect of nonstationarity when different methods are used can be compared. This is useful since many real signals in nature are nonstationary. Then, TSSA-EMD, SSA, EEMD-ICA, and SCICA were applied to extract the original signal. The original signal and the results can be seen in Fig. 4 and the corresponding RMSEs are shown in Table I.

Furthermore, Fig. 5 demonstrates the system performance versus the changes in noise level.

Since many signals in nature have a cyclo-stationary or oscillatory property, this application of the proposed method is very significant. In Section II-C, the importance of the proposed method is further studied and verified by applying it to sleep EEG data.

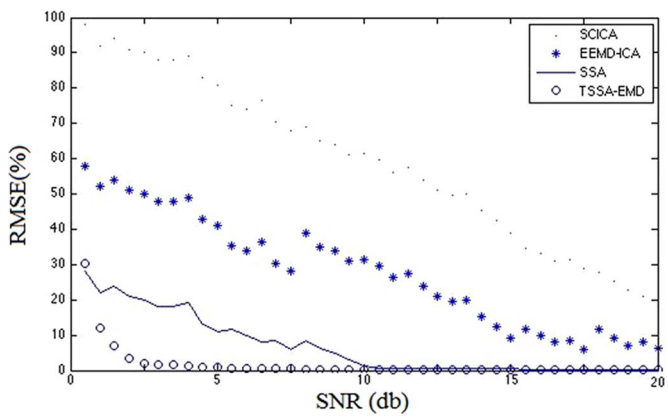

Fig. 5. Effect of noise level for the simulated narrowband data for SCICA, EEMD-ICA, SSA, and TSSA-EMD.

B. Simulations Discussion

Compared to other algorithms in our simulations, TSSA performs best for nonstationary and narrowband data. The EEMD-ICA algorithm can perform quite well for narrowband sources. However, it has several drawbacks. First, in order to achieve the best performance especially in the case of low SNR, its parameters need to be adjusted properly. If there isn't any prior knowledge of the noise power, either the regular EMD

technique might be used or EEMD should be run several times, until achieving the best separation [13]. Furthermore, it uses ICA and therefore, it has a problem with nonstationary data. Using other ICA techniques may solve this problem but not the problem of the first stage of this algorithm.

The SCICA algorithm performs worse than other algorithms used in this study. This approach has two major limitations: 1) it assumes stationarity of the data and 2) it cannot separate sources with overlapping spectra. In both simulations, SCICA did not perform well due to significant spectral overlapping and also nonstationarity of the data.

SSA cannot work in nonstationary environments either. In other words, in SSA, the data is initially divided into stationary segments before the decomposition stage. Therefore, the length of the signal segments is limited by the stationarity requirement while the proposed method can perform in such environments. Moreover, multi-way methods such as PARAFAC are less sensitive to noise [28]. Furthermore, unlike in PCA, there is no rotation problem in PARAFAC and pure spectra can be recovered [28]. Thus, as seen in the results, the proposed method performs better than other methods used for comparison here.

C. Application to Sleep EEG

Sleep is a highly complex brain state and an indicator of the changes in the brain function such as those occurring in many psychiatric and neurological conditions. Sleep is characterized by a reduction in body movements, reduced responsiveness to external stimuli, and changes in metabolic rate [29]. Hence, it is a state of unconsciousness from which a person can be aroused. Brain states during sleep and wakefulness are identified through interactions of activating and inhibiting systems within the brain. There are two distinct states with different levels of neuronal activity: nonrapid eye movement (NREM) and REM sleep. Each stage has a distinct set of physiological and neurological features and also dominant frequency band [29]. Most prominently, the alpha rhythm (8–13 Hz) is attenuated and delta (up to 4 Hz) waves evolve as the NREM sleep deepens. Other features of NREM sleep are sleep spindles (12–14 Hz) and K complexes. It has been assumed that K complexes are due to continuous sensory activation and it can be considered as a building block of slow wave (SW) sleep which has more power than usual delta waves [30]. Since visual sleep scoring is a time consuming process, automatic sleep staging methods hold promise in diagnosing alterations in the sleep EEG more efficiently.

Thirty-six healthy men and women each participated in two laboratory sessions, one involving a sleep extension protocol and the other a sleep restriction protocol. During each session polysomnography (PSG) measures were recorded at a sampling rate of 256 Hz for a baseline (BL) night (8 h), seven condition nights (sleep extension (ES), 10 h; sleep restriction (SR), 6 h) and a recovery night (12 h) following a period of total sleep deprivation. In this paper, the proposed method and also two commonly used methods for EEG sleep processing, wavelet and power spectra analysis using fast Fourier transform (FFT), were applied to sleep data to extract the components in different frequency bands (alpha and delta). FFT is used to calculate the signal spectrum. This refers to application of a bandpass filter. Here, this is performed by multiplication of the signal by the filter impulse response in Fourier domain. Moreover, it is shown that Morlet wavelet can be used for sleep signal analysis for example for detection of sleep spindles and abnormalities [31], [32]. In this work, Morlet wavelet is used for the signal decomposition. Delta rhythm is a slow brain wave which tends to have its highest amplitude during deep sleep in adults and is usually prominent frontally. Alpha can be seen in the posterior regions of the head on both sides and emerges with closing of the eyes and relaxation. Therefore, only one channel with more variations in alpha and another one with more variations in delta were chosen and after applying the following method the average power was determined for each 2 s time segment. Thus, each point i in Figs. 6 and 7 correspond to ith 2 s frame of the signal starting at $i \times 2 \times 256$ sample in the original signal. The results of these methods for one subject in BL night are depicted in Fig. 6. In addition, the result of applying the proposed method to BL, SE, and SR nights is seen in Fig. 7. The result for other available signals follows the same structure for different methods. Table II indicates the start of the second and SW stages with respect to time segments according to a precise hand scoring and the proposed method by a simple thresholding on the resultant values.

D. Results and Discussion

EEG data often contains many transient events and movement related sources and artefacts. Additionally, sleep signals are nonstationary and corrupted by noise. Tensor factorization, when applied to segmented signals, can be considered as one possible solution to restoration of such data and separating of the desired components. Therefore, as explained before, TSSA can act much better in such applications. The results clearly verify this claim.

As indicated in Fig. 6, the proposed method can determine the transitions between the stages of sleep by more accurately evaluating the alpha and delta (SW) brain activity variations. This significant achievement complies with an accurate manual scoring of the sleep data by clinical experts which can be seen in Table II. Moreover, the agreement between the scorers was determined with intraclass correlation coefficients (ICC) which were computed using a two-way analysis of variance [33]. The higher numbers in ICC represent better agreement between raters. We obtained $\mathrm{ICC} = 0.99$ which shows almost complete agreement.

The proposed method performs well for detecting the stage changes. Wavelet transform performs satisfactory for detecting the start of the first stage. However, it cannot show the deep sleep stage very well according to Fig. 6(b). Filtering using FFT also cannot reveal the necessary information about the sleep stages [Fig. 6(c)]. Therefore, based on the illustrated results, the proposed method performs better than the conventional methods i.e., WT and FFT, for sleep processing. Another important result of our method is extraction of the K complexes

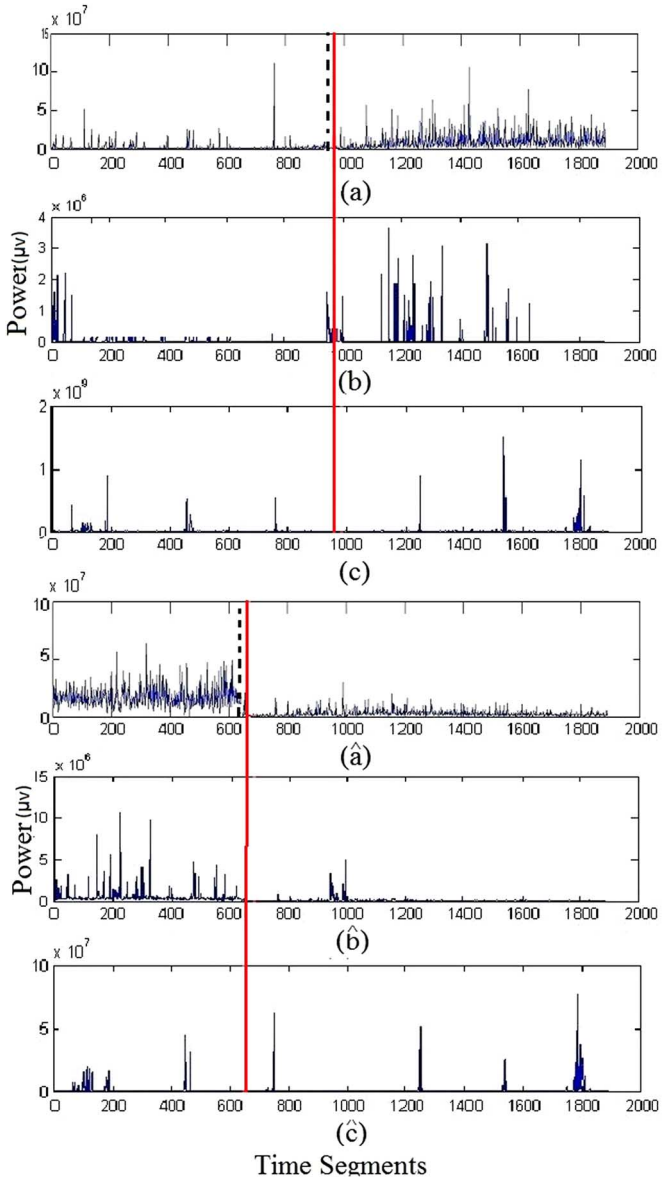

Fig. 6. Changes in alpha and delta band power during one cycle of sleep for baseline night with respect to time segments obtained using respectively TSSA-EMD, wavelet, and FFT; (a)–(c) Changes in delta power and $(\hat{a})$, $(\hat{b})$, and $(\hat{c})$ show the power changes for alpha band. Manual scoring is indicated by a bold solid line and those estimated by dotted lines.

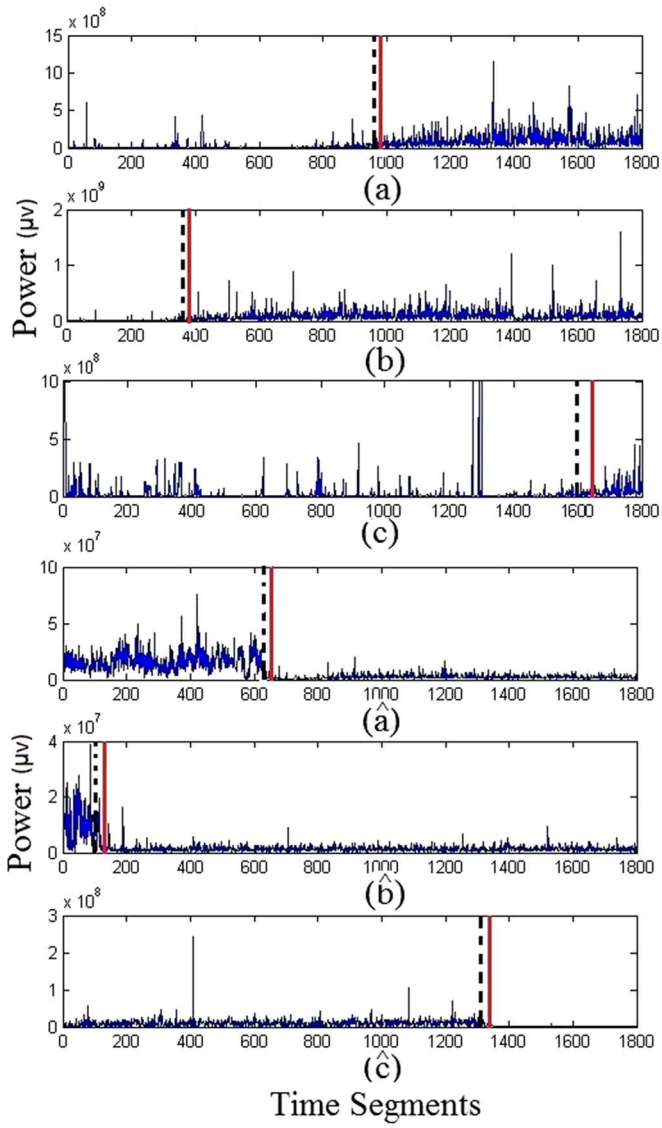

Fig. 7. Changes in alpha and delta band power during one cycle of sleep for BL, SR, and SE with respect to time segments obtained using respectively TSSA-EMD; (a) and $(\hat{a})$ BL night, (b) and $(\hat{b})$ SR, and (c) and $(\hat{c})$ SE. (a)–(c) Changes in delta power and $(\hat{a})$, $(\hat{b})$, and $(\hat{c})$ show the power changes for alpha band. Manual scoring is indicated by a bold solid line and those estimated by dotted lines.

from the sleep signal. K complexes are seen in EEG sleep data after finishing the first stage of sleep and have the same frequency band as delta but with significantly higher amplitude. Therefore, as can be seen in Fig. 6(a) the sudden increases can show the locations of these sudden changes in sleep. Fig. 7 also confirms the results of some previous experiments in which a fast transition from alpha to slow waves is shown in SR condition and the reverse in SE condition.

TABLE II
START OF TIME SEGMENT OF SECOND AND SW STAGES FOR BASELINE NIGHT, SE, AND SR NIGHTS FOR EXAMINED SUBJECT OBTAINED BY MANUAL SCORING AND THE PROPOSED METHOD

Condition \ Scoring	Manual		Proposed Automatic	
	Second	SW	Second	SW
baseline	660	990	638	980
SR	150	390	120	380
SE	1350	1650	1321	1600

VI. CONCLUSION

Decomposition of single channel time series into their statistically well-defined components is highly desired in many applications. For EEG signals, evaluation of narrow band components such as alpha and delta in sleep study and alpha and beta in BCI design is often required. Here, traditional SSA has been extended to tensor based SSA at the decomposition stage of the algorithm. Furthermore, an adaptive supervised approach using EMD has been introduced and implemented in the reconstruction stage of TSSA to accurately select the desired subspace when the frequency band of the signal of interest is known a

priori. In this way, the desired subspace could be chosen automatically. The results of applying the method to synthetic and real EEG data show that this method can produce much better results. In addition, the proposed method paves the way for better understanding and analysis of sleep EEG data, as discussed and examined in this paper. The work can also be extended to identification of underdetermined systems for separation of the sources from a limited number of measurements.

REFERENCES

[1] L. C. Parra and P. Sajda, "Blind source separation via generalized eigenvalue decomposition," *J. Mach. Learn. Res.*, vol. 4, pp. 1261–1269, 2003.

[2] V. Calhoun, J. Liu, and T. Adali, "A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data," *Neuroimage*, vol. 45, pp. 163–172, 2009.

[3] M. De Vos, A. Vergult, L. De Lathauwer, W. De Clercq, S. Van Huffel, P. Dupont, A. Palmini, and W. Van Paesschen, "Canonical decomposition of ictal scalp EEG reliably detects the seizure onset zone," *NeuroImage*, vol. 37, pp. 844–854, 2007.

[4] S. Rickard, "The DUET blind source separation algorithm," in *Blind Speech Separation*. Dordrecht, The Netherlands: Springer, 2007, pp. 217–241.

[5] L. Yuanqing, A. Cichocki, and S. Amari, "Sparse component analysis for blind source separation with less sensors than sources," in *Proc. ICA2003*, 2003, pp. 89–94.

[6] P. Zhou and T. Kuiken, "Eliminating cardiac contamination from myoelectric control signals developed by targeted muscle reinnervation," *Physiol. Meas.*, vol. 27, pp. 1311–1327, 2006.

[7] J. Taelman, A. Spaepen, and S. Van Huffel, "Wavelet-independent component analysis to remove electrocardiography contamination in surface electromyography," in *Proc. Eng. Med. Biol. Soc.*, 2007, pp. 682–685.

[8] M. Lewicki, "A review of methods for spike sorting: The detection and classification of neural action potentials," *Comput. Neural Syst.*, vol. 9, pp. R53–78, 1998.

[9] E. Niedermeyer, "Alpha rhythms as physiological and abnormal phenomena," *Int. J. Psychophysiol.*, vol. 26, pp. 31–49, 1997.

[10] R. E. Mistlberger, B. M. Bergmann, and A. Rechtschaffen, "Relationships among wake episode lengths, contiguous sleep episode lengths, and electroencephalographic delta waves in rats with suprachiasmatic nuclei lesions," *Sleep*, vol. 10, pp. 12–24, 1987.

[11] M. E. Davies and C. J. James, "Source separation using single channel ICA," *Signal Process.*, vol. 87, pp. 1819–1832, 2007.

[12] J. Lin and A. Zhang, "Fault feature separation using wavelet-ICA filter," *NDT & E Int.*, vol. 38, pp. 421–427, 2005.

[13] B. Mijovic, M. De Vos, I. Gligorijevic, J. Taelman, and S. Van Huffel, "Source separation from single-channel recordings by combining empirical-mode decomposition and independent component analysis," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 9, pp. 2188–2196, Sep. 2010.

[14] N. Golyandina, V. Nekrutkin, and A. Zhigljavsky, *Analysis of Time Series Structure: SSA and Related Techniques*. Boca Raton, FL: Chapman & Hall/CRC, 2001.

[15] T. H. Alexandrov and N. Golyandina, "The automatic extraction of time series trend and periodical components with the help of the Caterpillar-SSA approach," *Exponenta Pro 3–4*, pp. 54–61, 2004.

[16] S. Sanei, T. K. M. Lee, and V. Abolghasemi, "A new adaptive line enhancer based on singular spectrum analysis," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 2, pp. 428–432, Feb. 2012.

[17] F. Ghaderi, H. R. Mohseni, and S. Sanei, "Localizing heart sounds in respiratory signals using singular spectrum analysis," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 12, pp. 3360–3367, Dec. 2011.

[18] S. Sanei, M. Ghodsi, and H. Hassani, "A constrained singular spectrum analysis approach to murmur detection from heart sounds," *J. Med. Eng. Phys.*, vol. 33, pp. 362–367, 2010.

[19] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 626–634, May 1999.

[20] N. E. Huang, Z. Shen, S. R. Long, M. L. Wu, H. H. Shih, Q. Zheng, N. C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and Hilbert spectrum for nonlinear and nonstationary time series analysis," in *Proc. R. Soc. London Ser. A*, 1998, vol. 454, pp. 903–995.

[21] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: A noise-assisted data analysis method," *Adv. Adaptive Data Anal.*, vol. 1, pp. 1–41, 2009.

[22] R. A. Harshman, "Foundations of the PARAFAC procedure: Model and conditions for an explanatory multi-mode factor analysis," *UCLA Working Papers Phonetics*, vol. 16, pp. 1–84, 1970.

[23] J. D. Carroll and J. Chang, "Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition," *Psychometrika*, vol. 35, pp. 283–319, 1970.

[24] R. A. Harshman and S. A. Berenbaum, "Basic concepts underlying the PARAFAC-CANDECOMP three-way factor analysis model and its application to longitudinal data," in *Present and Past in Middle Life*. New York: Academic, 1981, pp. 435–459, .

[25] D. S. Burdick, "An introduction to tensor products with applications to multiway data analysis," *J. Chemom. Intel. Lab. Sys.*, vol. 28, pp. 229–237, 1995.

[26] R. Sands and F. W. Young, "Component models for three-way data: An alternating least squares algorithm with optimal scaling features," *Psychometrika*, vol. 45, pp. 39–67, 1980.

[27] L. De Lathauwer and J. Castaing, "Blind identification of underdetermined mixtures by simultaneous matrix diagonalization," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1096–1105, Mar. 2008.

[28] R. Bro, "PARAFAC. Tutorial and Applications," *Chemom. Intell. Lab. Syst., INCINC96*, vol. 38, pp. 149–171, 1997.

[29] S. Sanei, *Adaptive Processing of Brain Signals*. New York: Wiley, 2013.

[30] P. Halasz, I. Pal, and P. Rajna, "K-complex formation of the EEG in sleep: A survey and new examinations," *Physiol. Acad. Sci. Hung.*, vol. 65, pp. 3–35, 1985.

[31] N. Schaltenbrand, L. R. Lengelle, and M. Jean-Paul, "Neural network model: Application to automatic analysis of human sleep," *Comput. Biomed. Res.*, vol. 26, no. 2, pp. 157–171, 1993.

[32] E. Sitnikova, E. H. Alexander, A. K. Alexey, and V. L. Gilles, "Sleep spindles and spikewave discharges in EEG: Their generic features, similarities and distinctions disclosed with Fourier transform and continuous wavelet analysis," *J. Neurosci. Methods*, vol. 180, no. 2, pp. 304–316, 2009.

[33] P. E. Shrout and L. J. Fleiss, "Interclass correlations: Uses in assessing rater reliability," *Psychol. Bull.*, vol. 86, pp. 420–428, 1979.

Samaneh Kouchaki (S'14) received the B.Sc. and M.Sc. degrees in computer engineering from Shiraz University of Technology, Shiraz, Iran, in 2009 and 2011, respectively. She is currently working toward the Ph.D. degree in the Faculty of Engineering and Physical Sciences, University of Surrey, Guildford, U.K.

Her main research interests include blind source separation and biomedical signal processing.

Saeid Sanei (SM'05) received the Ph.D. degree from Imperial College London, London, U.K., in 1991.

Since then he has been a member of academic staff in Iran, Singapore, and the U.K. Currently, he is Deputy Head of Computing Department, Faculty of Engineering and Physical Sciences, University of Surrey, Guildford, U.K. He is the author of three monograms in biomedical signal processing and approximately 300 peer reviewed papers. His research interests include adaptive and nonlinear signal processing, machine learning, and multi-way signal processing.

Emma Arbon received the Ph.D. degree in sleep–wake regulation from the University of Surrey, Guildford, U.K., in 2013. She previously read Biomedical Sciences as an undergraduate at the University of Surrey, before undertaking the M.S. degree in toxicology at the same institution.

Her research interests are primarily concerned with quantitative analysis of the EEG in human sleep and include pharmacological manipulation of sleep, sleep deprivation, as well as individual differences in sleep–wake regulation.

Derk-Jan Dijk is Professor of Sleep and Physiology and Director of the Surrey Sleep Research Centre and Director of Sleep Wake Research of the Surrey Clinical Research Centre. He is also the Associate Dean in the Faculty of Health and Medical Science. He has more than 30 years of experience in clinical sleep research. He has published more than 160 research and review papers in the area of sleep and circadian rhythms. He also serves as consultant to the pharmaceutical industry.

Dr. Dijk is Vice-President of the World Sleep Federation.