Spatiotemporal Sparse Bayesian Learning With Applications to Compressed Sensing of Multichannel Physiological Signals

Zhilin Zhang, *Member, IEEE*, Tzyy-Ping Jung, *Senior Member, IEEE*, Scott Makeig, Zhouyue Pi, *Senior Member, IEEE*, and Bhaskar D. Rao, *Fellow, IEEE*

Abstract—Energy consumption is an important issue in continuous wireless telemonitoring of physiological signals. Compressed sensing (CS) is a promising framework to address it, due to its energy-efficient data compression procedure. However, most CS algorithms have difficulty in data recovery due to nonsparsity characteristic of many physiological signals. Block sparse Bayesian learning (BSBL) is an effective approach to recover such signals with satisfactory recovery quality. However, it is time-consuming in recovering multichannel signals, since its computational load almost linearly increases with the number of channels. This work proposes a spatiotemporal sparse Bayesian learning algorithm to recover multichannel signals simultaneously. It not only exploits temporal correlation within each channel signal, but also exploits inter-channel correlation among different channel signals. Furthermore, its computational load is not significantly affected by the number of channels. The proposed algorithm was applied to brain computer interface (BCI) and EEG-based driver's drowsiness estimation. Results showed that the algorithm had both better recovery performance and much higher speed than BSBL. Particularly, the proposed algorithm ensured that the BCI classification and the drowsiness estimation had little degradation even when data were compressed by 80%, making it very suitable for continuous wireless telemonitoring of multichannel signals.

Index Terms—Brain–computer interface (BCI), compressed sensing (CS), electroencephalography (EEG), sparse Bayesian learning (SBL), spatiotemporal correlation, telemonitoring, wireless body-area network (WBAN).

I. INTRODUCTION

COMPRESSED sensing (CS) [1] has been drawing increasing attention in the wireless telemonitoring of physiological signals as an emerging data compression

*Z. Zhang was with the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093 USA. He is currently with the Emerging Technology Lab, Samsung Research America-Dallas, Richardson, TX 75082 USA (e-mail: zhilinzhang@ieee.org).

T.-P. Jung and S. Makeig are with the Swartz Center for Computational Neuroscience and the Center for Advanced Neurological Engineering, University of California at San Diego, La Jolla, CA 92093 USA.

Z. Pi is with the Emerging Technology Lab, Samsung Research America-Dallas, Richardson, TX 75082 USA.

B. D. Rao is with the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNSRE.2014.2319334

methodology[1] [2]–[7]. It has been shown that CS, compared to traditional data compression methodologies, consumes much less energy and power [8], saves lots of on-chip computational resources [9], and is robust to packet loss during wireless transmission [10]. Thus it is attractive to wireless body-area networks for ambulatory monitoring.

A. CS Models

The basic CS framework [1], also called the single measurement vector (SMV) model, can be expressed as

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{v} \tag{1}$$

where, in the context of data compression, $\mathbf{x} \in \mathbb{R}^{M \times 1}$ is a single-channel signal, $\mathbf{\Phi} \in \mathbb{R}^{N \times M}(N < M)$ is a user-designed measurement matrix, $\mathbf{v} \in \mathbb{R}^{N \times 1}$ is sensor noise, and $\mathbf{y} \in \mathbb{R}^{N \times 1}$ is the compressed signal. This compression task is performed in sensors of a wireless body-area network. Then, the compressed signal $\mathbf{y}$, through Bluetooth and Internet, is sent to a remote terminal. At the terminal, the original signal is recovered by a CS algorithm using the shared measurement matrix $\mathbf{\Phi}$, namely[2]

$$\widehat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2^2 + \lambda g(\mathbf{x}) \tag{2}$$

where λ is a regularization parameter, and $g(\mathbf{x})$ is a penalty function of $\mathbf{x}$. The most popular penalty may be the ℓ_1-minimization based penalty, namely $g(\mathbf{x}) = \|\mathbf{x}\|_1$. This method is called *signal recovery in the original domain*.

When the original signal $\mathbf{x}$ is sufficiently sparse (i.e., only a few entries of $\mathbf{x}$ are nonzero), many CS algorithm can exactly recover $\mathbf{x}$ from $\mathbf{y}$ in the absence of noise $\mathbf{v}$ or with high quality in the presence of noise[3]. If $\mathbf{x}$ is not sparse, one can seek a dictionary matrix $\mathbf{D}$ such that $\mathbf{x}$ can be sparsely represented under the dictionary matrix, i.e., $\mathbf{x} = \mathbf{D}\mathbf{z}$, where the representation coefficients $\mathbf{z}$ are sparse. The dictionary matrix

---

[1]The CS technique can be used for data compression and signal sampling [1]. In this paper the use of CS for data compression/decompression is considered. But note the proposed algorithm can be also used as a signal recovery method in CS-based sampling.

[2]There are other mathematical expressions, which are equivalent given suitable values for regularization parameters

[3]Admittedly, when $\mathbf{x}$ is sparse, it is trivial to use CS for data compression, because one can just send nonzero entries (and associated locations) of $\mathbf{x}$ to a remote terminal and then recover it over there. When $\mathbf{x}$ is nonsparse, directly using the recovery method (2) results in failure for existing CS algorithms. Thus, the recovery method (2) is rarely used by CS algorithms. But the block sparse Bayesian learning can adopt this method (2) to recover a nonsparse $\mathbf{x}$ with correlated entries (with very small errors)[6].

$\mathbf{D} \in \mathbb{R}^{M \times m} (M \leq m)$ can be formed from orthonormal bases of known transforms such as discrete wavelet transform or discrete Cosine transform (DCT), or can be learned from data using dictionary learning [11]. Then a CS algorithm recovers the original signal according to

$$\widehat{\mathbf{x}} = \mathbf{D}\widehat{\mathbf{z}} \quad \text{with} \quad \widehat{\mathbf{z}} = \arg \min_{\mathbf{z}} \|\mathbf{y} - \mathbf{\Omega}\mathbf{z}\|_2^2 + \lambda g(\mathbf{z}) \qquad (3)$$

where $\mathbf{\Omega} \triangleq \mathbf{\Phi}\mathbf{D}$. The method is called *signal recovery in a transformed domain*.

The basic CS framework has been widely studied for data compression/decompression of biosignals [2], [5]–[7], [12]–[14]. For example, Aviyente [2] studied the use of Gabor dictionary matrix for EEG. Later Abdulghani *et al.* [12] further investigated various kinds of dictionary matrices. Instead of using the popular ℓ_1-minimization based penalty, other more effective penalties were proposed, such as the block-sparsity with intra-block correlation [7], [15], the analysis prior formulation [13], and the sparsity on second-order difference [14]. Chen *et al.* [5] proposed an energy-efficient digital implementation of CS architecture for data compression in wireless sensors. Using a field programmable gate array (FPGA) platform, Liu *et al.* [9] showed that CS, when compared to a low-power wavelet compression procedure, can largely save energy, power, and other on-chip computational resources.

In addition to the SMV model (1), another widely studied CS model is the multiple measurement vector (MMV) model [16], an extension of the SMV model. It can be expressed as follows:

$$\mathbf{Y} = \mathbf{\Phi}\mathbf{X} + \mathbf{V} \qquad (4)$$

where $\mathbf{Y} \in \mathbb{R}^{N \times L}$, $\mathbf{X} \in \mathbb{R}^{M \times L}$ and $\mathbf{V} \in \mathbb{R}^{N \times L}$ are matrices. A key assumption in the MMV model is that $\mathbf{X}$ is row sparse, namely only a few rows of $\mathbf{X}$ are nonzero rows. Similar to (2) and (3), the estimate of $\mathbf{X}$ is given by

$$\widehat{\mathbf{X}} = \arg \min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{\Phi}\mathbf{X}\|_2^2 + \lambda f(\mathbf{X}) \qquad (5)$$

or given by

$$\widehat{\mathbf{X}} = \mathbf{D}\widehat{\mathbf{Z}} \quad \text{with} \quad \widehat{\mathbf{Z}} = \arg \min_{\mathbf{Z}} \|\mathbf{Y} - \mathbf{\Omega}\mathbf{Z}\|_2^2 + \lambda f(\mathbf{Z}) \quad (6)$$

where $\mathbf{\Omega} \triangleq \mathbf{\Phi}\mathbf{D}$, and $\mathbf{D}$ is a dictionary matrix. $f(\mathbf{X})$ is a penalty encouraging row-sparsity of $\mathbf{X}$. One popular penalty is the ℓ_1/ℓ_2-minimization based penalty, namely $f(\mathbf{X}) = \sum_{i=1}^{M} \|\mathbf{X}_{i\cdot}\|_2$. In (6) it is assumed that $\mathbf{Z}$ is row-sparse.

Compared to recovering $\mathbf{X}$ column by column, i.e., treating (5) [or (6)] as L individual subproblems, the joint recovery as in (5) [or (6)] can greatly improve the recovery quality of $\mathbf{X}$ [16], [17]. Aviyente [2] explored this model to jointly recover multichannel EEG signals. Polania *et al.* [18] explored this model to jointly recover multichannel ECG signals. However, the benefit of the MMV model is largely compromised if columns of $\mathbf{X}$ exhibit inter-vector correlation; the benefit even almost disappears when the inter-vector correlation is very high [19].

Recently, by proposing the T-MSBL algorithm [19], we showed that suitably exploiting the inter-vector correlation can greatly alleviate its negative effect. Particularly, in noiseless environments, under mild conditions the negative effect disappears no matter how large the inter-vector correlation is (as long as the correlation is not ±1). This algorithm motivated the development of the spatiotemporal algorithm presented in this paper.

B. Challenges in the Use of CS for Wireless Telemonitoring

It is worth pointing out that most CS algorithms may not be used for energy-efficient wireless telemonitoring especially ambulatory monitoring, due to several challenges [20]–[22].

One challenge comes from the strict energy constraint. A wireless telemonitoring system is generally battery-operated. This situation with other constraints (e.g., wearability and device cost) requires that the compression procedure should be as simple as possible. In other words, the preprocessing such as filtering, peak detection, and dynamical thresholding, is not favored, since they increase circuitry complexity and cost extra energy. In fact, the data compression stage should be very simple. Lots of evidence have shown that the energy-saving advantage of CS over conventional data compression methods might be true only when the measurement matrix $\mathbf{\Phi}$ was a *sparse binary matrix*; when $\mathbf{\Phi}$ was a random Gaussian matrix or other kinds of matrices, the advantage disappeared.

Another challenge comes from strong artifacts caused by human movement during data recording. The goal of wireless telemonitoring is to allow people to move freely. Thus, the collected physiological signals are inevitably contaminated by strong artifacts caused by muscle movement and electrode motion. As a result, even a sparse signal can become nonsparse in the time domain and also nonsparse in transformed domains [20]. The nonsparsity seriously degrades CS algorithms' performance, resulting in their failure [6]. Therefore, CS algorithms generally need to remove artifacts before compression. But this greatly increases circuitry complexity, and conflicts with the energy constraint. The conflict is more sharp in some scenarios such as ambulatory telemonitoring.

Very recently, we proposed using the block sparse Bayesian learning (BSBL) framework [15] for CS of nonsparse physiological signals, and achieved success in telemonitoring of fetal ECG [6] and single-channel EEG [7]. The significant innovation in those works is that, instead of using the mentioned preprocessing methods or seeking optimal dictionary matrices, we proposed a completely different approach: namely recovering nonsparse signals directly without resorting to optimal dictionary matrices or preprocessing methods. The key element in BSBL is exploitation of correlation structures of a signal.

However, BSBL is designed for recovering single-channel signals. When recovering multichannel signals, BSBL has to recover the signals channel by channel, which is time-consuming and thus not suitable for real-time telemonitoring of multichannel signals. Besides, for many multichannel physiological signals, there is strong correlation among signals of different channels. Exploiting the inter-channel correlation is necessary and very beneficial. Unfortunately, BSBL ignores it.

C. Summary of the Work

The work introduces a spatiotemporal sparse model to the field of CS. This model is an extension of the classic multivariate

Bayesian variable selection model [23], and was recently used in overdetermined multivariate regression models to identify predictors by exploiting nonlinear relationships between predictors and responses [24]. However, this model has not been studied in CS.

Based on this model, we derive an expectation-maximization based spatiotemporal sparse Bayesian learning algorithm, and apply it to CS of multichannel signals. This algorithm has several advantages.

- It can efficiently exploit temporal correlation of each channel signal and inter-channel correlation among different channel signals to improve recovery performance. As we will see later, exploiting the inter-channel correlation is very important in CS of multichannel signals.
- It has the ability to recover nonsparse correlated signals, and signals with less-sparse representation coefficients, a desired ability for wireless telemonitoring of physiological signals.
- Compared to BSBL, it not only has better recovery performance, but also has higher speed. Its computational load is not significantly affected by the number of channels, an obvious advantage over BSBL. Thus it is very suitable for CS of multichannel signals.
- Different from most CS algorithms, which require preprocessing before compressing raw data, the proposed algorithm does not require any preprocessing. Its compression procedure can be implemented by very simple circuits, thus costing ultra-low energy consumption. This is highly desired for long-term wireless telemonitoring of physiological signals.

In experiments on steady-state visual evoked potential (SSVEP) based BCI and EEG-based driver's drowsiness estimation, the proposed algorithm ensured that the BCI classification and the drowsiness estimation on recovered data were almost the same as on original data, even when the original data were compressed by more than 80%.

Some preliminary results were published in [20]. The MATLAB code of the proposed algorithm can be downloaded at https://sites.google.com/site/researchbyzhang/stsbl.

D. Organization and Notations

The rest of the paper is organized as follows. Section II describes the spatiotemporal sparse model. Section III derives a spatiotemporal sparse Bayesian learning algorithm using the expectation-maximization method. Section IV discusses some specific settings when applying the algorithm for CS of multichannel physiological signals. Section V presents experimental results on BCI and EEG-based driver's drowsiness estimation. Discussion and conclusion are given in the last two sections.

We introduce the notations used in this paper.

- Bold symbols are reserved for vectors and matrices. Particularly, $\mathbf{I}_L$ denotes the identity matrix with size $L \times L$. When the dimension is evident from the context, for simplicity, we just use $\mathbf{I}$.
- $\|\mathbf{x}\|_1, \|\mathbf{x}\|_2, \|\mathbf{A}\|_{\mathcal{F}}$ denote the ℓ_1 norm of the vector $\mathbf{x}$, the ℓ_2 norm of $\mathbf{x}$, and the Frobenius norm of the matrix $\mathbf{A}$, respectively.

- $\mathrm{diag}\{a_1, \cdots, a_m\}$ denotes a diagonal matrix with principal diagonal elements being $a_1, \cdots, a_m$ in turn; if $\mathbf{A}_1, \cdots, \mathbf{A}_m$ are square matrices, then $\mathrm{diag}\{\mathbf{A}_1, \cdots, \mathbf{A}_m\}$ denotes a block diagonal matrix with principal diagonal blocks being $\mathbf{A}_1, \cdots, \mathbf{A}_m$ in turn.
- $\mathbf{A} \otimes \mathbf{B}$ represents the Kronecker product of the two matrices $\mathbf{A}$ and $\mathbf{B}$. $\mathrm{vec}(\mathbf{A})$ denotes the vectorization of the matrix $\mathbf{A}$ formed by stacking its columns into a single column vector. $\mathrm{Tr}(\mathbf{A})$ denotes the trace of $\mathbf{A}$. $\mathbf{A}^T$ denotes the transpose of $\mathbf{A}$.
- For a matrix $\mathbf{A}$, $\mathbf{A}_{i\cdot}$ denotes the ith row, and $\mathbf{A}_{\cdot i}$ denotes the ith column. $\mathbf{A}_{[i],j}$ denotes the ith block in the jth column. $\mathbf{A}_{i,[j]}$ denotes the jth block in the ith row. When assuming all columns of $\mathbf{A}$ have the same block partition, $\mathbf{A}_{[i]\cdot}$ denotes the ith block of all columns of $\mathbf{A}$.

II. SPATIOTEMPORAL SPARSE MODEL

To enhance the readability of the paper, we first describe the spatiotemporal sparse model in this section, and delay the description of the proposed algorithm to the next section.

The spatiotemporal sparse model is described as follows:

$$\mathbf{Y} = \mathbf{\Phi}\mathbf{X} + \mathbf{V} \tag{7}$$

where $\mathbf{Y} \in \mathbb{R}^{N \times L}$, $\mathbf{\Phi} \in \mathbb{R}^{N \times M}$[4], and $\mathbf{X} \in \mathbb{R}^{M \times L}$. The matrices $\mathbf{Y}$ and $\mathbf{\Phi}$ are known. The goal is to estimate $\mathbf{X}$. In the context of data compression, the lth column of $\mathbf{X}$, denoted by $\mathbf{X}_{\cdot l}$, is a segment of an original physiological signal in the lth channel, and the lth column of $\mathbf{Y}$ is the corresponding compressed segment.

The matrix $\mathbf{X}$ is assumed to have the following block structure:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_{[1]\cdot} \\ \mathbf{X}_{[2]\cdot} \\ \vdots \\ \mathbf{X}_{[g]\cdot} \end{bmatrix} \tag{8}$$

where $\mathbf{X}_{[i]\cdot} \in \mathbb{R}^{d_i \times L}$ is the ith block of $\mathbf{X}$, and $\sum_{i=1}^{g} d_i = M$. For convenience, $\{d_1, \cdots, d_g\}$ is called the block partition. Among the g blocks, only a few are nonzero. The key assumption is that each block $\mathbf{X}_{[i]\cdot} (\forall i)$ is assumed to have spatiotemporal correlation. In other words, entries in the same column of $\mathbf{X}_{[i]\cdot}$ are correlated[5], and entries in the same row of $\mathbf{X}_{[i]\cdot}$ are also correlated[6].

The ith block $\mathbf{X}_{[i]\cdot}$ is assumed to have the parameterized Gaussian distribution $p(\mathrm{vec}(\mathbf{X}_{[i]\cdot}^T); \gamma_i, \mathbf{B}, \mathbf{A}_i) = \mathcal{N}(\mathbf{0}, (\gamma_i \mathbf{A}_i) \otimes \mathbf{B})$. Here $\mathbf{B} \in \mathbb{R}^{L \times L}$ is an unknown positive definite matrix capturing the correlation structure in each row of $\mathbf{X}_{[i]\cdot}$. The matrix $\mathbf{A}_i \in \mathbb{R}^{d_i \times d_i}$ is an unknown positive definite matrix capturing the correlation structure in each column of $\mathbf{X}_{[i]\cdot}$. The unknown parameter γ_i is a nonnegative scalar, determining whether the ith block is a zero block or not.

---

[4]The model and the developed algorithm does not require $N < M$ or $N \geq M$. Thus they can be used for many other applications.

[5]In our data compression formulation, the correlation is a kind of temporal correlation of a channel signal.

[6]In our data compression formulation, the correlation is called inter-channel correlation, and is also called spatial correlation.

Assuming the blocks $\{\mathbf{X}_{[i]}\}_{i=1}^{g}$ are mutually independent, the distribution of the matrix $\mathbf{X}$ is

$$p\left(\mathrm{vec}(\mathbf{X}^{T}); \mathbf{B}, \{\gamma_i, \mathbf{A}_i\}_i\right) = \mathcal{N}(\mathbf{0}, \mathbf{\Pi} \otimes \mathbf{B}) \qquad (9)$$

where $\mathbf{\Pi}$ is a block diagonal matrix defined by

$$\mathbf{\Pi} \triangleq \begin{bmatrix} \gamma_1 \mathbf{A}_1 & & & \\ & \gamma_2 \mathbf{A}_2 & & \\ & & \ddots & \\ & & & \gamma_g \mathbf{A}_g \end{bmatrix}. \qquad (10)$$

Besides, each row of the noise matrix $\mathbf{V}$ is assumed to have the distribution $p(\mathbf{V}_{i\cdot}; \lambda, \mathbf{B}) = \mathcal{N}(\mathbf{0}, \lambda \mathbf{B})$, where λ is an unknown scalar. Assuming the rows are mutually independent, we have

$$p\left(\mathrm{vec}(\mathbf{V}^{T}); \lambda, \mathbf{B}\right) = \mathcal{N}(\mathbf{0}, \lambda \mathbf{I} \otimes \mathbf{B}). \qquad (11)$$

Remark 1: Note that $\mathbf{X}$ and $\mathbf{V}$ share the common matrix $\mathbf{B}$ for modeling the correlation structure of each row. This is a traditional setting in Bayesian variable selection models [23], which facilitates the use of conjugate priors for multivariate linear regression. Besides, since in our applications the sensor noise $\mathbf{V}$ can be ignored, the covariance model of $\mathbf{V}$ is not important. It only facilitates the development of our algorithm.

Remark 2: The proposed STSBL model is an extension of the model used by BSBL [15]. Setting $L = 1$, the STSBL model reduces to the latter. In other words, the STSBL model can be viewed as a set of multiple BSBL models with their solution vectors *mutually correlated*. In Section VI-C we will see the necessity of modeling the mutual correlation.

Remark 3: The proposed STSBL model is also closely related to the T-MSBL model [19][7]. When $d_i = 1 (\forall i)$, STSBL reduces to the latter. Note that T-MSBL only exploits correlation among entries of the same row in $\mathbf{X}$, while STSBL also exploits correlation among entries of the same column in $\mathbf{X}$. In the context of data compression, T-MSBL only exploits the inter-channel correlation, while STSBL exploits both the inter-channel correlation and the temporal correlation within each channel signal.

The relationships revealed in Remark 2 and Remark 3 inspire us to derive an efficient algorithm, as shown below.

III. SPATIOTEMPORAL SBL ALGORITHM

Due to the coupling between $\mathbf{A}_i(\forall i)$ and $\mathbf{B}$, directly estimating parameters in the model (7) can result in an algorithm with heavy computational load. However, the observations in Remark 2 and Remark 3 imply that we can use $\mathbf{B}$ as a spatially whitening matrix, transforming the original model (7) to a spatially whitened model, and use $\mathbf{A}_i(\forall i)$ to transform the original model to a temporally whitened model[8]. Thus, we propose an *alternating-learning approach*, where the parameters $\{\gamma_i, \mathbf{A}_i\}_{i=1}^{g}$ and λ are estimated from the spatially whitened

[7]Due to the difference in problem formulation, the temporal correlation studied in [19] is the inter-channel correlation in this work.

[8]In fact, the block partition is still present. But for convenience, we call the equivalent model a "temporally whitened" model.

model, and the parameter $\mathbf{B}$ is estimated from the temporally whitened model. The resulting algorithm alternates the estimation between the two models until convergence. The alternating-learning approach largely simplifies the algorithm development.

A. Learning in the Spatially Whitened Model

To facilitate algorithm development, we assume $\mathbf{B}$ is known. Letting $\widetilde{\mathbf{Y}} \triangleq \mathbf{Y}\mathbf{B}^{-(1/2)}$, $\widetilde{\mathbf{X}} \triangleq \mathbf{X}\mathbf{B}^{-(1/2)}$, and $\widetilde{\mathbf{V}} \triangleq \mathbf{V}\mathbf{B}^{-(1/2)}$, the original STSBL model (7) becomes

$$\widetilde{\mathbf{Y}} = \mathbf{\Phi}\widetilde{\mathbf{X}} + \widetilde{\mathbf{V}} \qquad (12)$$

where the columns of $\widetilde{\mathbf{X}}$ are independent, and so does $\widetilde{\mathbf{V}}$. Thus, the original STSBL model is now spatially whitened, and the algorithm development becomes easier.

First, we have priors for $p(\widetilde{\mathbf{X}}; \mathbf{\Pi})$ and $p(\widetilde{\mathbf{V}}; \lambda)$ as follows:

$$p(\widetilde{\mathbf{X}}; \mathbf{\Pi}) = \prod_{i=1}^{L} p(\widetilde{\mathbf{X}}_{\cdot i}) \sim \prod_i \mathcal{N}(\mathbf{0}, \mathbf{\Pi}) \qquad (13)$$

$$p(\widetilde{\mathbf{V}}; \lambda) = \prod_{i=1}^{L} p(\widetilde{\mathbf{V}}_{\cdot i}) \sim \prod_i \mathcal{N}(\mathbf{0}, \lambda \mathbf{I}). \qquad (14)$$

Then we have the likelihood:

$$p(\widetilde{\mathbf{Y}}_{\cdot i} | \widetilde{\mathbf{X}}_{\cdot i}; \lambda) = \mathcal{N}(\mathbf{\Phi}\widetilde{\mathbf{X}}_{\cdot i}, \lambda \mathbf{I}) \quad \forall i. \qquad (15)$$

Thus, we obtain the posterior

$$p(\widetilde{\mathbf{X}}_{\cdot i} | \widetilde{\mathbf{Y}}_{\cdot i}; \lambda, \mathbf{\Pi}) = \mathcal{N}(\boldsymbol{\mu}_{\cdot i}, \boldsymbol{\Sigma}) \quad \forall i \qquad (16)$$

with the mean $\boldsymbol{\mu}_{\cdot i}$ and the covariance matrix $\boldsymbol{\Sigma}$ given by

$$\boldsymbol{\mu}_{\cdot i} = \mathbf{\Pi}\mathbf{\Phi}^{T}(\lambda \mathbf{I} + \mathbf{\Phi}\mathbf{\Pi}\mathbf{\Phi}^{T})^{-1}\widetilde{\mathbf{Y}}_{\cdot i} \quad \forall i \qquad (17)$$

$$\boldsymbol{\Sigma} = \left(\mathbf{\Pi}^{-1} + \frac{1}{\lambda}\mathbf{\Phi}^{T}\mathbf{\Phi}\right)^{-1} \qquad (18)$$

$$= \mathbf{\Pi} - \mathbf{\Pi}\mathbf{\Phi}^{T}(\lambda \mathbf{I} + \mathbf{\Phi}\mathbf{\Pi}\mathbf{\Phi}^{T})^{-1}\mathbf{\Phi}\mathbf{\Pi}. \qquad (19)$$

Once the parameters $\mathbf{\Pi}$ and λ are estimated, the maximum-a-posteriori (MAP) estimate of $\widetilde{\mathbf{X}}$ is directly given by the mean of the posterior, i.e.,

$$\widetilde{\mathbf{X}} \leftarrow \mathbf{\Pi}\mathbf{\Phi}^{T}(\lambda \mathbf{I} + \mathbf{\Phi}\mathbf{\Pi}\mathbf{\Phi}^{T})^{-1}\widetilde{\mathbf{Y}} \qquad (20)$$

and the solution matrix $\mathbf{X}$ in the original STSBL model (7) can be obtained

$$\mathbf{X} \leftarrow \widetilde{\mathbf{X}}\mathbf{B}^{\frac{1}{2}}. \qquad (21)$$

Thus, estimating the parameters $\mathbf{\Pi}$ and λ is crucial to the algorithm. There are many optimization methods which can be used to estimate these parameters, such as bound-optimization methods [15], fast marginal likelihood maximization [25], and variational methods [26]. In this work, we use the expectation maximization (EM) method to estimate them, since we find the

resulting algorithm can provide better recovery performance in our application.

Using the EM method, $\widetilde{\mathbf{X}}$ is treated as a hidden variable. The Q-function for estimating $\{\gamma_i\}_i$ and $\{\mathbf{A}_i\}_i$ is given by

$$
\begin{aligned}
\mathcal{Q}(\mathbf{\Pi}) &\triangleq E_{\widetilde{\mathbf{X}}|\widetilde{\mathbf{Y}};\Theta^{(\text{old})}} \left[\log p\left(\widetilde{\mathbf{X}}; \{\gamma_i\}_i, \{\mathbf{A}_i\}_i \right) \right] \\
&= -\frac{L}{2}\log|\mathbf{\Pi}| - \frac{1}{2}\sum_{i=1}^{L} E_{\widetilde{\mathbf{X}}|\widetilde{\mathbf{Y}};\Theta^{(\text{old})}} \left[\widetilde{\mathbf{X}}_{\cdot i}^T \mathbf{\Pi}^{-1} \widetilde{\mathbf{X}}_{\cdot i} \right] \\
&= -\frac{L}{2}\sum_{i=1}^{g}\log|\gamma_i\mathbf{A}_i| - \frac{1}{2}\sum_{l=1}^{L}\text{Tr}\left[\mathbf{\Pi}^{-1}\left(\mathbf{\Sigma} + \boldsymbol{\mu}_{\cdot l}\boldsymbol{\mu}_{\cdot l}^T \right) \right] \\
&= -\frac{L}{2}\sum_{i=1}^{g} d_i \log\gamma_i - \frac{L}{2}\sum_{i=1}^{g}\log|\mathbf{A}_i| \\
&\quad -\frac{1}{2}\sum_{l=1}^{L}\sum_{j=1}^{g}\frac{1}{\gamma_j}\text{Tr}\left[\mathbf{A}_j^{-1}\left(\mathbf{\Sigma}_{[j]} + \boldsymbol{\mu}_{[j]l}\boldsymbol{\mu}_{[j]l}^T \right) \right] \quad (22)
\end{aligned}
$$

where $\Theta^{(\text{old})}$ denotes all the parameters estimated in the previous iteration, $\mathbf{\Sigma}_{[j]}$ denotes the jth diagonal block in $\mathbf{\Sigma}$ which corresponds to the jth block in $\widetilde{\mathbf{X}}$, and $\boldsymbol{\mu}_{[j]l}$ denotes the jth block in the lth column of $\boldsymbol{\mu}$.

Setting to zero the derivative of (22) with respect to γ_i, we obtain the updating rule for γ_i

$$
\gamma_i \leftarrow \frac{1}{Ld_i}\sum_{l=1}^{L}\text{Tr}\left[\mathbf{A}_i^{-1}\left(\mathbf{\Sigma}_{[i]} + \boldsymbol{\mu}_{[i]l}\boldsymbol{\mu}_{[i]l}^T \right) \right]. \quad (23)
$$

Setting to zero the derivative of (22) with respect to $\mathbf{A}_i$, we obtain the updating rule for $\mathbf{A}_i$

$$
\mathbf{A}_i \leftarrow \frac{1}{L}\sum_{l=1}^{L}\frac{\mathbf{\Sigma}_{[i]} + \boldsymbol{\mu}_{[i]l}\boldsymbol{\mu}_{[i]l}^T}{\gamma_i}. \quad (24)
$$

The estimate will be further regularized as shown later.

To estimate λ, the Q-function is given by

$$
\begin{aligned}
\mathcal{Q}(\lambda) &= E_{\widetilde{\mathbf{X}}|\widetilde{\mathbf{Y}};\Theta^{(\text{old})}} \left[\log p(\widetilde{\mathbf{Y}}|\widetilde{\mathbf{X}}; \lambda) \right] \\
&\propto -\frac{NL}{2}\log\lambda - \frac{1}{2\lambda} E_{\widetilde{\mathbf{X}}|\widetilde{\mathbf{Y}};\Theta^{(\text{old})}} \left[\sum_{l=1}^{L}\|\widetilde{\mathbf{Y}}_{\cdot l} - \mathbf{\Phi}\widetilde{\mathbf{X}}_{\cdot l}\|_2^2 \right] \\
&= -\frac{NL}{2}\log\lambda \\
&\quad -\frac{1}{2\lambda}\sum_{l=1}^{L}\left[\|\widetilde{\mathbf{Y}}_{\cdot l} - \mathbf{\Phi}\boldsymbol{\mu}_{\cdot l}\|_2^2 \right. \\
&\qquad \left. + E_{\widetilde{\mathbf{X}}|\widetilde{\mathbf{Y}};\Theta^{(\text{old})}} \left[\left\|\mathbf{\Phi}(\widetilde{\mathbf{X}}_{\cdot l} - \boldsymbol{\mu}_{\cdot l})\right\|_2^2 \right] \right] \\
&= -\frac{NL}{2}\log\lambda - \frac{1}{2\lambda}\|\widetilde{\mathbf{Y}} - \mathbf{\Phi}\boldsymbol{\mu}\|_{\mathcal{F}}^2 - \frac{L}{2\lambda}\text{Tr}(\mathbf{\Sigma}\mathbf{\Phi}^T\mathbf{\Phi}). \\
&\hspace{8cm} (25)
\end{aligned}
$$

Setting its derivative to zero, we have

$$
\lambda \leftarrow \frac{1}{NL}\|\widetilde{\mathbf{Y}} - \mathbf{\Phi}\boldsymbol{\mu}\|_{\mathcal{F}}^2 + \frac{1}{N}\text{Tr}(\mathbf{\Sigma}\mathbf{\Phi}^T\mathbf{\Phi}). \quad (26)
$$

Similar to the approach adopted in [15], at low signal-to-noise (SNR) situations the above updating rule is modified to

$$
\lambda \leftarrow \frac{1}{NL}\|\widetilde{\mathbf{Y}} - \mathbf{\Phi}\boldsymbol{\mu}\|_{\mathcal{F}}^2 + \frac{1}{N}\sum_{i=1}^{g}\text{Tr}\left(\mathbf{\Sigma}_{[i]}\mathbf{\Phi}_{\cdot[i]}^T\mathbf{\Phi}_{\cdot[i]} \right) \quad (27)
$$

where $\mathbf{\Phi}_{\cdot[i]}$ denotes the consecutive columns in $\mathbf{\Phi}$ which correspond to the ith block in $\widetilde{\mathbf{X}}$. In noiseless situations one can simply set $\lambda = 10^{-10}$ or other small values, instead of performing the updating rule (26).

In the above development we have assumed that $\mathbf{B}$ is given. This parameter can be estimated in a temporally whitened model discussed below.

B. Learning in the Temporally Whitened Model

To estimate the matrix $\mathbf{B}$, we consider the following equivalent form of the original model (7):

$$
\mathbf{Y} = \overline{\mathbf{\Phi}} \cdot \overline{\mathbf{X}} + \mathbf{V} \quad (28)
$$

where $\overline{\mathbf{\Phi}} \triangleq \mathbf{\Phi}\mathbf{A}^{1/2}$, $\overline{\mathbf{X}} \triangleq \mathbf{A}^{-(1/2)}\mathbf{X}$, and $\mathbf{A}$ is defined as $\mathbf{A} \triangleq \text{diag}\{\mathbf{A}_1, \cdots, \mathbf{A}_g\}$. Note that in this model, $\overline{\mathbf{X}}$ maintains the same block partition as $\mathbf{X}$, but its every block has no temporal correlation due to the temporally whitening effect from $\mathbf{A}_i^{-(1/2)}(\forall i)$. Thus, estimating $\mathbf{B}$ in this model becomes easier.

Following the approach used to derive the T-MSBL algorithm [19] and assuming $\mathbf{X}$, $\{\gamma_i\}_i$ and $\{\mathbf{A}_i\}_i$ have been obtained from the spatially whitened model (12), we have the following updating rule for the matrix $\mathbf{B}$:

$$
\begin{aligned}
\check{\mathbf{B}} &\leftarrow \sum_{i=1}^{g}\gamma_i^{-1}\overline{\mathbf{X}}_{[i]\cdot}^T\overline{\mathbf{X}}_{[i]\cdot} + \lambda^{-1}(\mathbf{Y} - \overline{\mathbf{\Phi}\mathbf{X}})^T(\mathbf{Y} - \overline{\mathbf{\Phi}\mathbf{X}}) \\
&= \sum_{i=1}^{g}\frac{\mathbf{X}_{[i]\cdot}^T\mathbf{A}_i^{-1}\mathbf{X}_{[i]\cdot}}{\gamma_i} + \frac{(\mathbf{Y} - \mathbf{\Phi}\mathbf{X})^T(\mathbf{Y} - \mathbf{\Phi}\mathbf{X})}{\lambda} \quad (29)
\end{aligned}
$$

$$
\mathbf{B} \leftarrow \frac{\check{\mathbf{B}}}{\|\check{\mathbf{B}}\|_{\mathcal{F}}} \quad (30)
$$

where $\overline{\mathbf{X}}_{[i]\cdot} \in \mathbb{R}^{d_i \times L}$ is the ith block in $\overline{\mathbf{X}}$, and $\overline{\mathbf{X}}_{[i]\cdot} \triangleq \mathbf{A}_i^{-(1/2)}\mathbf{X}_{[i]\cdot}$. The second item in (29) is noise-related. When the noise is very small, or does not exist (i.e., $\lambda \to 0$), it is suggested to remove the second item for robustness.

C. Regularization

In the proposed spatiotemporal model the number of unknown parameters is much larger than the number of available data. Thus regularization to the estimated $\mathbf{B}$ and $\{\mathbf{A}_i\}_i$ is very important. Suitable regularization helps to overcome learning difficulties resulting from the very limited data.

As in [19], we can regularize the $\check{\mathbf{B}}$ in (29) by

$$
\check{\mathbf{B}} \leftarrow \sum_{i=1}^{g}\gamma_i^{-1}\mathbf{X}_{[i]\cdot}^T\mathbf{A}_i^{-1}\mathbf{X}_{[i]\cdot} + \eta\mathbf{I} \quad (31)
$$

where η is a positive scalar. This regularization is shown empirically to increase robustness in noisy environments. In noiseless environments, this regularization is not needed.

To regularize the estimates of $\{\mathbf{A}_i\}_i$, we use the strategy in [15], i.e., modeling the correlation matrix of each column in $\mathbf{X}_{[i]}$ as the correlation matrix of an AR(1) process with the common AR coefficient r for all i. The strategy can be summarized as follows.

- Step 1: From each $\mathbf{A}_i$, calculate the quantity r_i by $r_i \leftarrow (m_1^i/m_0^i)(\forall i)$, where m_0^i is the average of entries in the main diagonal of $\mathbf{A}_i$ and m_1^i is the average of entries in the main sub-diagonal of $\mathbf{A}_i$. Note that due to some numerical problems, m_1^i/m_0^i may be out of the feasible range $(-1, 1)$, and thus further constraints may be imposed; for example, $r_i \leftarrow \text{sign}(m_1^i/m_0^i) \min\{|(m_1^i/m_0^i)|, 0.99\}$.
- Step 2: Average: $r \leftarrow (1/g) \sum_{i=1}^{g} r_i$.
- Step 3: Reconstruct the regularized $\mathbf{A}_i(\forall i)$

$$\check{\mathbf{A}}_i \leftarrow \begin{bmatrix} 1 & r & \cdots & r^{d_i-1} \\ \vdots & \vdots & & \vdots \\ r^{d_i-1} & r^{d_i-2} & \cdots & 1 \end{bmatrix}$$

$$\mathbf{A}_i \leftarrow \frac{\check{\mathbf{A}}_i}{\|\check{\mathbf{A}}_i\|_{\mathcal{F}}}.$$

The parameter-averaging strategy has been widely used in artificial neural networks and the machine learning communities to overcome overfitting.

Experiments showed these regularization strategies helped further improve the algorithm's performance. In fact, using the Theorem 1 in [19] it can be proved that in noiseless situations the regularization strategies to $\mathbf{A}_i$ and $\mathbf{B}$ do not affect the global minimum of the cost function of our algorithm, in the sense that the global minimum corresponds to the true sparse solution. This implies that a good regularization strategy can significantly enhance global convergence of our algorithm.

Up to now we have derived the updating rules for $\mathbf{X}$, $\{\mathbf{A}_i\}_i$, $\{\gamma_i\}_i$ and λ in the spatially whitened model and the updating rules for $\mathbf{B}$ in the temporally whitened model. Combining these updating rules we obtain the EM-based spatiotemporal sparse Bayesian learning algorithm, denoted by **STSBL-EM**.

IV. PRACTICAL CONSIDERATIONS WHEN APPLYING STSBL-EM

The proposed STSBL-EM algorithm has wide applications. This section discusses some practical considerations when applying it in practice.

In CS of multichannel physiological signals, if the channel signal $\mathbf{X}_{\cdot l}$ has strong temporal correlation[9] in the time domain, using the original spatiotemporal model (7) can achieve good recovery performance. A typical signal is ECG signals [6].

When each channel signal $\mathbf{X}_{\cdot l}$ does not have strong temporal correlation, exploiting the temporal correlation may not be very beneficial. Thus one can alternatively exploit the sparsity of each channel signal in some transformed domain by using a dictionary matrix in STSBL-EM, as stated in Section I. In particular, one can first apply the algorithm to the following model:

$$\mathbf{Y} = \mathbf{\Omega}\mathbf{Z} + \mathbf{V} \tag{32}$$

to find the solution $\mathbf{Z}$, where $\mathbf{\Omega} \triangleq \mathbf{\Phi}\mathbf{D}$, and $\mathbf{D}$ is a dictionary matrix under which $\mathbf{X}_{\cdot l}(\forall l)$ has sparse representation $\mathbf{Z}_{\cdot l}$. Then one can obtain the original solution by computing $\mathbf{X} = \mathbf{D}\mathbf{Z}$. Note that in this method $\mathbf{Z}_{\cdot l}$ is sparser than $\mathbf{X}_{\cdot l}$, but generally has less correlation than the latter, or the correlation structure in $\mathbf{Z}_{\cdot l}$ is not well captured by STSBL-EM. Hence, this method mainly exploits each channel signal's sparsity in a transformed domain instead of exploiting the channel signal's temporal correlation.[10] This method can yield better results than using the original model (7), if each channel signal has no strong temporal correlation. A typical signal is EEG signals [7].

In the following experiments on EEG signals we will adopt the model (32) with the dictionary matrix $\mathbf{D}$ formed by the orthogonal DCT bases.[11] Due to the "energy compaction" property of DCT, for the lth channel signal $\mathbf{X}_{\cdot l}$, the DCT coefficients with significantly nonzero values are concentrated in the first K entries in $\mathbf{Z}_{\cdot l}$. Note that the first K nonzero entries (with other coefficients with insignificantly nonzero values locating at the $(K + 1)$th entry, the $(K + 2)$th entry, etc.) can be viewed as concatenation of a number of nonzero blocks. In this sense, the value of K does not need to be known *a priori*, and the block partition in STSBL-EM can be set rather arbitrarily. In our experiments we found STSBL-EM showed stable performance when the block partition $d_i(\forall i)$ chose values from a wide range (15–60). (Similar robustness was also observed on BSBL [6].) Thus we simply set $d_i = 16(\forall i)$.

In practice most SBL algorithms implicitly adopt a γ_i-pruning mechanism [15], [19], [29]. The mechanism sets a small γ_i to zero if it is smaller than a threshold, thus speeding up convergence and encouraging solutions to be sparse in the level of entries [29], blocks [15], or rows [19]. However, for raw EEG signals (especially those recorded during ambulatory monitoring) the value of K/M could be very large [20]. Thus the DCT coefficient vectors are not sparse. In this case, better recovery performance can be achieved by setting the γ_i-pruning threshold to a very small value or even zero and allowing algorithms to iterate only a few times [6], [7]. In our experiments we set this threshold to zero, and terminated the algorithm when the iteration number reached 40 or the maximum change in any entry of the estimated $\mathbf{X}$ in two successive iterations was smaller than 10^{-6}. But when used in other applications such as source localization, it may need hundreds of iterations to converge.

In our work the problem of data compression is modeled as a noiseless CS problem (i.e., the sensor noise $\mathbf{V}$ is ignored). Therefore, in our experiments STSBL-EM was performed in the noiseless situation with the parameter λ set to $\lambda = 10^{-10}$. But this does not mean that artifacts and noise in raw physiological signals are ignored. In fact, in our model $\mathbf{X}_{\cdot l}$ is a raw physiological signal contaminated by noise and artifacts.

---

[9]Here 'strong temporal correlation' means that if modeling the signal by an AR(1) process, the absolute value of the AR coefficient is very large.

[10]Note that when using some dictionary matrices such as wavelet dictionaries, one may exploit both sparsity and wavelet tree structures in $\mathbf{Z}_{\cdot l}$, which is more beneficial than merely exploiting the sparsity [27], [28].

[11]One may find other dictionary matrices which can yield better results than the DCT dictionary matrix on EEG signals [12]. But seeking the optimal dictionary matrix is not the focus of this work.

V. APPLICATION

The proposed STSBL-EM was used for CS of multichannel EEG signals in SSVEP-based BCI and EEG-based driver's drowsiness estimation.

To show the superior performance of STSBL-EM, we chose the BSBL-BO algorithm, an MMV-model-based CS algorithm, and an SMV-model-based CS algorithm for comparison. We did not choose many algorithms for comparison, since in [6] it has been shown that ten state-of-the-art CS algorithms were inferior to BSBL-BO. Thus, our focus was the comparison between STSBL-EM and BSBL-BO. The three algorithms are briefly described as follows.

- BSBL-BO [15].[12] To the best of our knowledge, it may be the only algorithm that has the ability to recover both nonsparse physiological signals [6] and the physiological signals with nonsparse representation coefficients [7]. Its block partition was set to $d_1 = d_2 = \cdots = 16$.
- ISL0 [30].[13] It is based on the MMV model. When $\mathbf{Z}$ is less row-sparse, it has robust performance than many MMV-model-based algorithms.
- Basis Pursuit (BP) [31].[14] It is a classic CS algorithm based on the SMV model. Some work [12] claimed that it was more suitable for CS of EEG than other classic CS algorithms. We used the SPGL1 software [32] to implement this algorithm.

All the algorithms recovered signals in the transformed domain. The dictionary matrix $\mathbf{D}$ was the DCT dictionary matrix. For all algorithms, the measurement matrix $\mathbf{\Phi}$ was an $N \times M$ sparse binary matrix of full row-rank, where M was fixed to 256 and N was varied to meet a desired compression ratio (CR). The CR was defined as

$$\text{CR} = \frac{M - N}{M} \times 100. \tag{33}$$

Irrespective of CR values, each column of the measurement matrix $\mathbf{\Phi}$ contained only *two* entries of 1's with random locations, while other entries were zeros.

Mean square error (MSE) is often used for measuring recovery quality. However, it is shown [33] that MSE is not a reasonable measure for natural signals. Thus it is not suitable for EEG, especially raw EEG signals contaminated by strong noise and artifacts. A smaller MSE does not necessarily mean that a desired task (e.g., EEG classification) on the recovered EEG signals can be better accomplished. Therefore, we used a task-oriented performance evaluation method, which was initially suggested in [6], [7].

The main idea of this evaluation method is that a practical task is first performed on an original dataset, and then the same task (using the same algorithm with the same initialization) is performed on the recovered dataset, and finally the results of the two tasks are compared. If the results are the same, this means that the recovered dataset has high fidelity and does not

affect the practical task. If the results are far from each other, this means that the recovered dataset is seriously distorted.

Using this idea, in our BCI experiment we compared the classification rate on original EEG signals to the classification rate on recovered signals. In the experiment on drowsiness estimation, we compared the estimation result using original signals to the estimation result using recovered signals. All the comparisons were repeated with different CR values.

Experiments were carried out on a computer with dual-core 2.8-GHz CPU and 6.0 GiB RAM.

A. SSVEP-Based BCI

In neurology, SSVEP is a response to a visual stimulus modulated at a specific frequency. The response has a fundamental frequency and several harmonics. The fundamental frequency is the same as that of the visual stimulus. This characteristic has been widely used in BCI [34] to classify stimuli with different frequencies, thereby finishing some control tasks. A trend in BCI is to develop wearable wireless systems [35], [36]. In such systems developing energy-efficient data acquisition modules is highly desired.

In this experiment the dataset analyzed in [35] was used.[15] The dataset was recorded from twelve subjects. We chose the recordings of "Subject 1" for illustration, which corresponded to visual stimuli of 9, 10, 11, 12, and 13 Hz. Each stimulus flashed for 4 s. The data sampling rate was 256 Hz. The monitor refresh rate was 75 Hz. As in [35], canonical correlation analysis (CCA) was used as the classifier. The selected channel indexes were 129, 133, 193, 196, 199, 200, 203, and 210 (all in the occipital area). Detailed descriptions on the dataset, the experiment equipment, and the recording procedure can be found in [35].

The signals were compressed and then recovered by STSBL-EM, BSBL-BO, ISL0, and BP, respectively. CR ranged from 50 to 90. The recovered signals were bandpass-filtered between 8–35 Hz. Each 8-channel epoch which corresponded to a visual stimulus was classified by CCA. The classification rate was calculated by averaging over all classification results on the whole recovered signals. The same bandpass filtering and classification were performed on the original signals.

The classification rates of all algorithms are given in Table I. Note that the classification rate on the original signals was 1.00. We can see that when $\text{CR} \leq 60$, the classification rate on the recovered signals by STSBL-EM was also 1.00. Even if $\text{CR} = 80$, the classification rate was very close to 1.00. These results imply that when the signals were compressed by 80%, the recovered signals by our algorithm were still of good quality. In contrast, all the compared algorithms did not recover the signals with satisfactory quality even with $\text{CR} = 60$.

To visually examine the data recovery quality, we randomly chose a time slot which corresponded to a visual stimulus of 10 Hz (duration was 4 s). Then we picked signals during this time slot in each channel from the original signals, and averaged their power spectrum densities (PSD's), shown in Fig. 1(a). We can clearly see the fundamental frequency (10 Hz) and the harmonic frequency (20 Hz). Similarly, we calculated the averaged PSD from the recovered signals by STSBL-EM when $\text{CR} = 80$,

---

[12]The MATLAB code was downloaded at https://sites.google.com/site/researchbyzhang/bsbl.

[13]The MATLAB code was provided by the first author of [30] via private communication.

[14]The MATLAB code was downloaded at http://www.cs.ubc.ca/mpf/spgl1/.

[15]The dataset was downloaded at ftp://sccn.ucsd.edu/pub/SSVEP.

TABLE I
CLASSIFICATION RATES OF ALL ALGORITHMS WHEN CR VARIED FROM
50 TO 90. CLASSIFICATION RATE ON THE ORIGINAL SIGNALS WAS **1.00**

CR	50	60	70	80	85	90
STSBL-EM	**1.00**	**1.00**	**0.984**	**0.984**	**0.976**	**0.672**
BSBL-BO	0.992	0.976	0.952	0.864	0.824	0.576
ISL0	0.888	0.840	0.800	0.704	0.536	0.488
Basis Pursuit	0.944	0.920	0.856	0.728	0.600	0.528
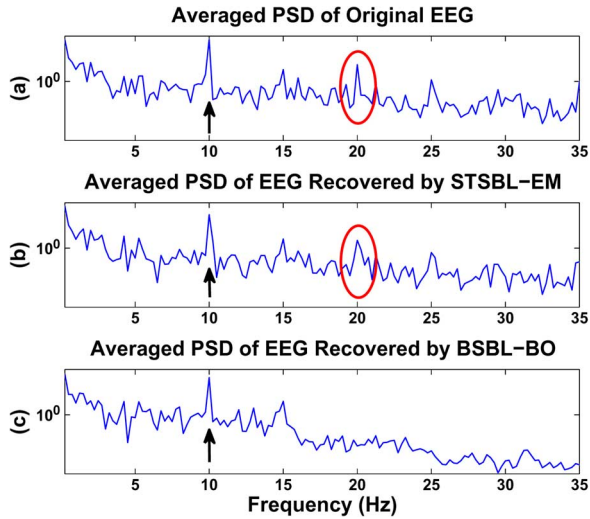
Fig. 1. (a) Averaged PSD of signals from the original signals, which corresponded to a visual stimulus of 10 Hz. (b) Averaged PSD of signals from the signals recovered by STSBL-EM when $CR = 80$. (c) Averaged PSD of signals from the signals recovered by BSBL-BO when $CR = 80$. Arrows indicate the fundamental frequency (10 Hz). Circles indicate the harmonic frequency (20 Hz).
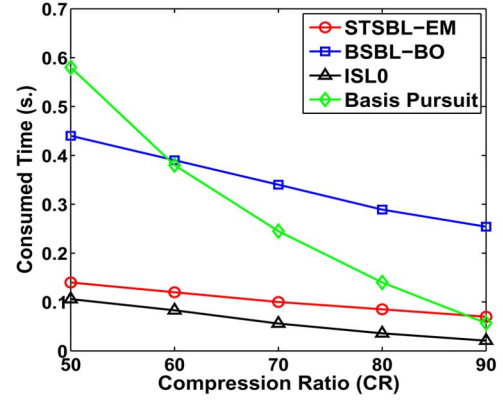

Fig. 2. Comparison of consumed time in recovery of 8-channel signals of 1 s duration at different CR. Only when the consumed time of an algorithm is far less than 1 s, can it be used for real-time (or near real-time) systems.

shown in Fig. 1(b), and the averaged PSD from the recovered signals by BSBL-BO when $CR = 80$, shown in Fig. 1(c). We can see both the fundamental frequency and the harmonic frequency in Fig. 1(b). But we do not see the harmonic frequency in Fig. 1(c). This explains why the classification rate on the signals recovered by BSBL-BO was lower than the classification rate on the signals recovered by STSBL-EM, since CCA exploited both the fundamental frequency and the harmonic frequency for classification.

Maintaining harmonic frequencies on recovered signals implies subtle waveforms in original signals are recovered. Therefore, the results shown in Fig. 1 further confirms that STSBL-EM has better data recovery quality than BSBL-BO.

Fig. 2 shows the averaged consumed time of each algorithm in recovering 8-channel signals of 1-s duration at different CR values. STSBL-EM was much faster than BSBL-BO. Their speed gap will be more significant in the next application, in which the number of EEG channels was 30.

B. EEG-Based Driver's Drowsiness Estimation

EEG-based driver's drowsiness estimation and prediction is an emerging technology for driving safety [37]–[39] and an important application of EEG. Such systems are powered by batteries and are generally embedded in a wearable device such as

an ordinary hat. Thus, it is highly desired to develop wireless EEG systems with low energy consumption [38]. In the following we will show that the proposed algorithm can be used in this application for energy efficient data transmission.

A set of EEG signals used in [38] were used in this experiment. The data were recorded from a subject using a 30-channel EEG system, when the subject was driving with some degree of drowsiness in a realistic kinesthetic virtual-reality driving simulator. The sampling rate was 250 Hz. During the driving, the deviation between the center of the vehicle and the center of the cruising lane was recorded, which was viewed as a driving error. The driving error is known to be a good indicator to drowsiness level [38], [39]. Details on the recording system, the recording procedure, and the virtual-reality driving simulator are given in [38].

Many methods were proposed to estimate the drowsiness level from recorded EEG signals. One method is given in [38], [39].

- Use lowpass filter with a cut-off frequency of 50 Hz to remove power line noise and other high-frequency noise from raw EEG signals.
- Perform online independent component analysis (ICA) [40] on the signals, and select an independent component (IC) for further analysis.
- Calculate log PSD of the selected IC at a frequency every 2 s. The time-varying subband log PSD is then used as the drowsiness estimate.[16]

To evaluate the quality of the drowsiness estimate, the Pearson correlation between the driving error (an indicator to the drowsiness level) and the time-varying subband log PSD of the selected IC is often evaluated. High Pearson correlation indicates a good drowsiness estimate. Details of the method can be found in [39].

Since our goal is to show that the proposed algorithm can be used in this application, we need to investigate whether the drowsiness estimation accuracy is degraded when using the recovered signals. Thus, we compared the drowsiness estimate

---

[16]For more robust estimation, one can seek an optimal mapping from the log PSD to the driving error using a training set. Since our goal in this experiment was to show the data recovery quality of the proposed algorithm, we just simply treated the time-varying log PSD as the drowsiness estimate.
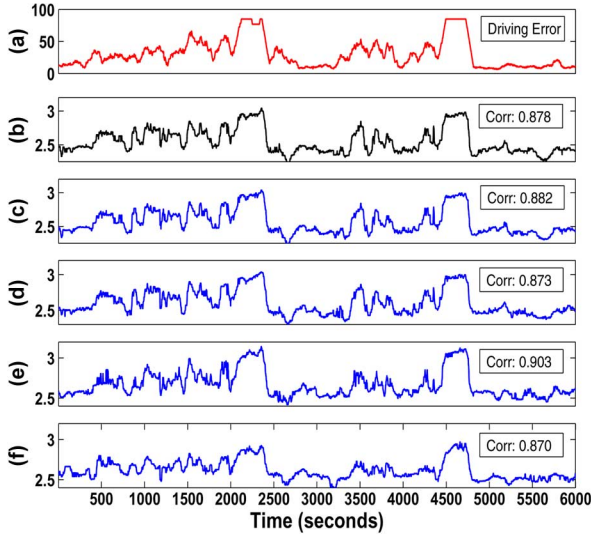
Fig. 3. Comparison of the driving error, the log PSD of IC_0 at $f = 5$ Hz, and the log PSD of IC_{rec} at $f = 5$ Hz and at different CR values. IC_{rec} was obtained from recovered signals by STSBL-EM which had the highest correlation with IC_0. (a) Driving error. (b) Log PSD of IC_0 at $f = 5$ Hz. (c)–(f) Log PSD of IC_{rec} at $f = 5$ Hz when $CR = 50, 60, 70$, and 80, respectively. Their Pearson correlations with the driving error are shown in each subplot.

from the recovered signals to the one from the original signals. Particularly, we adopted the following procedure.

1) Repeat the above drowsiness estimation using the original signals by selecting an IC (denoted by IC_0) and a frequency f. Evaluate the Pearson correlation between the driving error and the time-varying log PSD of IC_0 at the frequency f. Denote the correlation by r_0.

2) Perform the same ICA decomposition on the recovered signals, and choose the IC which has the highest Pearson correlation with IC_0. Denote the IC by IC_{rec}.

3) Calculate the time-varying log PSD of IC_{rec} at the frequency f.

4) Evaluate the Pearson correlation between the driving error and the time-varying log PSD calculated in the above step. Denote the Pearson correlation by r_{rec}.

5) Compare r_{rec} to r_0.

In our experiment, IC_0 was the IC whose log PSD at $f = 5$ (Hz) had the highest correlation with the driving error.

Fig. 3 shows the driving error signal, the time-varying log PSD of IC_0 at $f = 5$ Hz, and the time-varying log PSD of IC_{rec} at $f = 5$ Hz at different CR values. IC_{rec} was obtained from recovered signals by STSBL-EM. The r_0 and the r_{rec} at different CR values are also given in corresponding subplots. Clearly, when CR was no more than 80, the drowsiness estimate from the recovered signals by STSBL-EM was almost the same as the one from the original signals.

Table II further shows the r_0 and the r_{rec}'s of all algorithms when $f = 4, 5, 6, 7$ (Hz) and CR varied from 50 to 80. We can see when CR was small (e.g., 50–60), all the algorithms recovered the signals well. Their drowsiness estimates were almost the same as the estimate from the original signals. However, when CR increased, only STSBL-EM ensured accurate drowsiness estimation; particularly, the drowsiness estimate was almost not affected even if the raw EEG signals were compressed by 80%.

Fig. 4 shows the averaged consumed time of all algorithms in recovery of the 30-channel signals of 1.024 second duration at different CR values. STSBL-EM was much faster than BSBL-BO, suggesting that it is more suitable for real-time applications especially when the channel number is very large.

It is worth pointing out that the raw EEG signals contained strong artifacts due to muscle movement. However, the proposed algorithm did not require any preprocessing before data compression.

VI. DISCUSSIONS

A. Energy Consumption

We have mentioned that the proposed algorithm compresses data with ultra-low energy consumption. This is due to the use of the simplest measurement matrix and the algorithm's powerful data recovery ability.

The measurement matrix Φ is a very simple sparse binary matrix. Its each column contains only two entries of 1's, while other entries are zeros. Using this matrix has two major benefits.

- Code execution in data compression is largely reduced. Consequently, the energy dissipated in code execution is very low.
- Using this measurement matrix largely simplifies circuit design. Therefore the cost and the size of chips can be reduced.

It is worth noting that such a measurement matrix is not suitable for any CS algorithms. Some algorithms may have seriously degraded performance when using the measurement matrix.

Besides, many CS algorithms require preprocessing on raw data before compression, such as dynamic thresholding, filtering, and seeking specific waveform features. These preprocessing consumes lots of energy.[17] In contrast, our proposed algorithm does not require these preprocessing steps.

On the other hand, our algorithm's powerful recovery ability ensures high recovery performance when the compression ratio is high (e.g., $CR = 80$). Thus, the energy dissipated in wireless transmission can also be largely reduced.

In [9] and [10] the compression procedure of BSBL-BO was analyzed. These works showed that BSBL-BO, compared to conventional data compression procedures, dissipated only about 10%–20% energy, shortened compression time by more than 90%, and largely saved other computational resources. Since the compression procedures of BSBL and STSBL-EM are the same, these analysis results are applicable to STSBL-EM. But it is worth noting that STSBL-EM has more powerful recovery ability than BSBL-BO.

B. Stable Speed Regardless of Channel Numbers

Comparing Fig. 4 with Fig. 2 we find that the consumed time of STSBL-EM was relatively stable, although the channel number in Fig. 4 was almost four times of the channel number in Fig. 2. The reason is that to recover multichannel physiological signals, the algorithmic complexity of STSBL-EM mainly depends on the computation of (19) and (20), which is totally

---

[17]It is highly doubted that if using such preprocessing, CS still has its energy-saving advantages over traditional data compression algorithms.

TABLE II
COMPARISON BETWEEN r_0 CALCULATED FROM THE ORIGINAL SIGNALS AND r_{rec} CALCULATED FROM RECOVERED SIGNALS BY ALL ALGORITHMS AT 4–7 HZ AND DIFFERENT CR VALUES. "-" MEANS THE ICA DECOMPOSITION ON THE RECOVERED SIGNALS BY THE CORRESPONDING ALGORITHM DID NOT YIELD THE DESIRED IC

|	4Hz				5Hz			
CR	50	60	70	80	50	60	70	80
Original Data	0.853				0.878			
STSBL-EM	**0.853**	**0.842**	**0.866**	**0.848**	**0.882**	**0.873**	**0.903**	**0.870**
BSBL-BO	0.853	0.841	0.841	0.793	0.880	0.875	0.874	0.776
ISL0	0.851	0.735	-	-	0.885	0.776	-	-
Basis Pursuit	0.839	0.842	0.824	0.780	0.873	0.854	0.840	0.795
	6Hz				7Hz			
CR	50	60	70	80	50	60	70	80
Original Data	0.881				0.807			
STSBL-EM	**0.879**	**0.870**	**0.896**	**0.867**	**0.809**	**0.771**	**0.849**	**0.808**
BSBL-BO	0.873	0.871	0.882	0.733	0.788	0.801	0.802	0.526
ISL0	0.874	0.783	-	-	0.806	0.654	-	-
Basis Pursuit	0.867	0.842	0.808	0.766	0.779	0.753	0.728	0.584
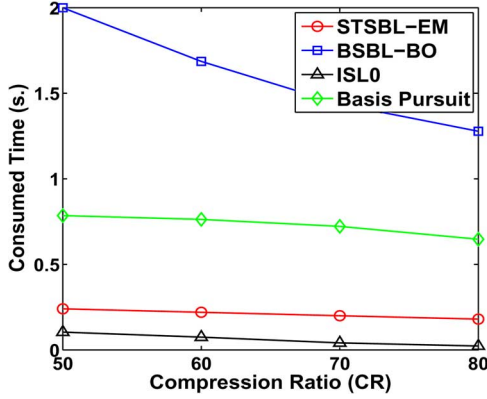
Fig. 4. Averaged consumed time of all algorithms in recovery of the 30-channel signals of 1.024-s duration at different CR values. BSBL-BO was slow, because it had to recover these signals channel by channel.

$\mathcal{O}(M^3 + 2M^2N + MN^2 + MNL + N^3)$. When L is small compared to M[18], the algorithmic complexity is approximately $\mathcal{O}(M^3 + 2M^2N)$, which does not depends on L. Thus the consumed time of STSBL-EM does not change significantly when the channel number dramatically changes. Note that when recovering a single-channel signal, the algorithmic complexity of BSBL-BO is also dominated by $\mathcal{O}(M^3 + 2M^2N)$. But when recovering L-channel signals, its computation load increases L-fold, since it has to recover the signals channel by channel. This explains why the consumed time by BSBL-BO in Fig. 4 was roughly four times the consumed time in Fig. 2.

C. Exploitation of Inter-Channel Correlation

Jointly recovering multichannel biosignals have been studied in a number of works. However, these works were generally based on the MMV model. They only exploited common sparsity profiles among channel signals, but did not exploit the inter-channel correlation. It is shown [19] that if ignoring the

correlation, most MMV-model based CS algorithms will have degraded recovery performance, especially in the presence of high inter-channel correlation. In the two EEG datasets used in our experiments, the inter-channel correlation between $\mathbf{Z}_{\cdot,i}$ and $\mathbf{Z}_{\cdot,j}$ is very high, generally above 0.9 (when $|i-j| \leq 5$). Thus it is not difficult to understand why ISL0 had poor performance in the experiments. In fact, in the two experiments if STSBL-EM was performed without exploiting the inter-channel correlation (i.e., setting $\mathbf{B} = \mathbf{I}$), the BCI classification rate and the drowsiness estimation were very poor, even poorer than those by BSBL-BO.

Therefore, exploiting the inter-channel correlation is necessary in CS of multichannel signals; ignoring it can seriously deteriorate CS algorithms' performance. This also indicates the importance of our work in developing the STSBL-EM algorithm which can exploit the correlation.

VII. CONCLUSION

We proposed a spatiotemporal sparse Bayesian learning algorithm for energy-efficient compressed sensing of multichannel signals. In contrast to existing compressed sensing algorithms, it not only exploits correlation structures within a single channel signal, but also exploits inter-channel correlation. It has much better recovery performance than state-of-the-art algorithms. Its speed is relatively stable even when the channel number significantly changes. Experiments on SSVEP-based BCI and EEG-based driver's drowsiness estimation showed that when using the proposed algorithm, the BCI classification rate and the drowsiness estimation on recovered signals were almost the same as those on original signals, even when the signals were compressed by 80%.

Since the algorithm takes root in Bayesian basis selection, it can be used in many other applications, such as feature selection, source localization, and sparse representation.

---

[18] In a typical scenario of telemonitoring of multichannel physiological signals, L varies from two to dozens, while M varies from 200 to 1000.

REFERENCES

[1] E. Candés and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[2] S. Aviyente, "Compressed sensing framework for EEG compression," in *Proc. IEEE/SP 14th Workshop Stat. Signal Process.*, 2007, pp. 181–184.

[3] E. C. Pinheiro, O. A. Postolache, and P. S. Girao, "Implementation of compressed sensing in telecardiology sensor networks," *Int. J. Telemed. Appl.*, 2010.

[4] A. M. Dixon, E. G. Allstot, D. Gangopadhyay, and D. J. Allstot, "Compressed sensing system considerations for ECG and EMG wireless biosensors," *IEEE Trans. Biomed. Circuits Syst.*, vol. 6, no. 2, pp. 156–166, Apr. 2012.

[5] F. Chen, A. Chandrakasan, and V. Stojanovic, "Design and analysis of a hardware-efficient compressed sensing architecture for data compression in wireless sensors," *IEEE J. Solid-State Circuits*, vol. 47, no. 3, pp. 744–756, Mar. 2012.

[6] Z. Zhang, T.-P. Jung, S. Makeig, and B. D. Rao, "Compressed sensing for energy-efficient wireless telemonitoring of noninvasive fetal ECG via block sparse Bayesian learning," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 2, pp. 300–309, Feb. 2013.

[7] Z. Zhang, T.-P. Jung, S. Makeig, and B. D. Rao, "Compressed sensing of EEG for wireless telemonitoring with low energy consumption and inexpensive hardware," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 1, pp. 221–224, Jan. 2013.

[8] H. Mamaghanian, N. Khaled, D. Atienza, and P. Vandergheynst, "Compressed sensing for real-time energy-efficient ECG compression on wireless body sensor nodes," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 9, pp. 2456–2466, Sep. 2011.

[9] B. Liu, Z. Zhang, G. Xu, H. Fan, and Q. Fu, "Energy efficient telemonitoring of physiological signals via compressed sensing: A fast algorithm and power consumption evaluation," *Biomed. Signal Process. Control*, vol. 11, pp. 80–88, 2014.

[10] S. Fauvel and R. K. Ward, "An energy efficient compressed sensing framework for the compression of electroencephalogram signals," *Sensors*, vol. 14, no. 1, pp. 1474–1496, 2014.

[11] K. Kreutz-Delgado, J. F. Murray, B. D. Rao, K. Engan, T.-W. Lee, and T. J. Sejnowski, "Dictionary learning algorithms for sparse representation," *Neural Computat.*, vol. 15, no. 2, pp. 349–396, 2003.

[12] A. M. Abdulghani, A. J. Casson, and E. Rodriguez-Villegas, "Compressive sensing scalp EEG signals: Implementations and practical performance," *Med. Biol. Eng. Comput.*, vol. 50, no. 11, pp. 1137–1145, 2012.

[13] M. Mohsina and A. Majumdar, "Gabor based analysis prior formulation for EEG signal reconstruction," *Biomed. Signal Process. Control*, vol. 8, no. 6, pp. 951–955, 2013.

[14] J. K. Pant and S. Krishnan, "Compressive sensing of electrocardiogram signals by promoting sparsity on the second-order difference and by using dictionary learning," *IEEE Trans. Biomed. Circuits Syst.*, vol. 8, no. 2, pp. 293–302, Apr. 2014.

[15] Z. Zhang and B. D. Rao, "Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation," *IEEE Trans. Signal Process.*, vol. 61, no. 8, pp. 2009–2015, Aug. 2013.

[16] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477–2488, Jul. 2005.

[17] Y. Eldar and H. Rauhut, "Average case analysis of multichannel sparse recovery using convex relaxation," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 505–519, Jan. 2010.

[18] L. F. Polania, R. E. Carrillo, M. Blanco-Velasco, and K. E. Barner, "Compressed sensing based method for ECG compression," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2011, pp. 761–764.

[19] Z. Zhang and B. D. Rao, "Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 912–926, Sep. 2011.

[20] Z. Zhang, B. D. Rao, and T.-P. Jung, "Compressed sensing for energy-efficient wireless telemonitoring: Challenges and opportunities," in *Asilomar Conf. Signals, Syst., Comput.*, 2013.

[21] A. Milenkovic, C. Otto, and E. Jovanov, "Wireless sensor networks for personal health monitoring: Issues and an implementation," *Comput. Commun.*, vol. 29, no. 13–14, pp. 2521–2533, 2006.

[22] T. Martin, E. Jovanov, and D. Raskovic, "Issues in wearable computing for medical monitoring applications: A case study of a wearable ECG monitoring device," in *Proc. 4th Int. Symp. Wearable Comput.*, 2000, pp. 43–49.

[23] P. Brown, M. Vannucci, and T. Fearn, "Multivariate Bayesian variable selection and prediction," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 60, no. 3, pp. 627–641, 1998.

[24] J. Wan, Z. Zhang, B. Rao, S. Fang, J. Yan, A. Saykin, and L. Shen, "Identifying the neuroanatomical basis of cognitive impairment in Alzheimer's disease by correlation-and nonlinearity-aware sparse Bayesian learning.," *IEEE Trans. Med. Imag.*, to be published.

[25] M. Tipping and A. Faul *et al.*, "Fast marginal likelihood maximisation for sparse Bayesian models," in *Proc. 9th Int. Workshop Artif. Intell. Stat.*, 2003, vol. 1, no. 3.

[26] D. Shutin, T. Buchgraber, S. Kulkarni, and H. Poor, "Fast variational sparse Bayesian learning with automatic relevance determination for superimposed signals," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 6257–6261, Dec. 2011.

[27] L. He and L. Carin, "Exploiting structure in wavelet-based bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3488–3497, Sep. 2009.

[28] C. Chen, Y. Li, and J. Huang, "Forest sparsity for multi-channel compressive sensing," *IEEE Trans. Signal Process.*, vol. 62, no. 11, pp. 2803–2813, Jun. 2014.

[29] M. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, 2001.

[30] M. M. Hyder and K. Mahata, "A robust algorithm for joint-sparse recovery," *IEEE Signal Process. Lett.*, vol. 16, no. 12, pp. 1091–1094, Dec. 2009.

[31] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.

[32] E. Van Den Berg and M. Friedlander, "Probing the Pareto frontier for basis pursuit solutions," *SIAM J. Sci. Comput.*, vol. 31, no. 2, pp. 890–912, 2008.

[33] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98–117, 2009.

[34] Y. Wang, X. Gao, B. Hong, C. Jia, and S. Gao, "Brain-computer interfaces based on visual evoked potentials," *IEEE Eng. Med. Biol. Mag.*, vol. 27, no. 5, pp. 64–71, Sep./Oct. 2008.

[35] Y. M. Chi, Y.-T. Wang, Y. Wang, C. Maier, T.-P. Jung, and G. Cauwenberghs, "Dry and noncontact EEG sensors for mobile brain-computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 2, pp. 228–235, Mar. 2012.

[36] L.-D. Liao, C.-Y. Chen, I.-J. Wang, S.-F. Chen, S.-Y. Li, B.-W. Chen, J.-Y. Chang, and C.-T. Lin, "Gaming control using a wearable and wireless EEG-based brain-computer interface device with novel dry foam-based sensors," *J. Neuroeng. Rehabil.*, vol. 9, no. 1, p. 5, 2012.

[37] T.-P. Jung, S. Makeig, M. Stensmo, and T. J. Sejnowski, "Estimating alertness from the EEG power spectrum," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 1, pp. 60–69, Jan. 1997.

[38] C. Lin, L. Ko, J. Chiou, J. Duann, R. Huang, S. Liang, T. Chiu, and T. Jung, "Noninvasive neural prostheses using mobile and wireless EEG," *Proc. IEEE*, vol. 96, no. 7, pp. 1167–1183, Jul. 2008.

[39] C.-T. Lin, R.-C. Wu, S.-F. Liang, W.-H. Chao, Y.-J. Chen, and T.-P. Jung, "EEG-based drowsiness estimation for safety driving using independent component analysis," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 52, no. 12, pp. 2726–2738, Dec. 2005.

[40] T.-W. Lee, M. Girolami, and T. J. Sejnowski, "Independent component analysis using an extended infomax algorithm for mixed sub Gaussian and super Gaussian sources," *Neural Computat.*, vol. 11, no. 2, pp. 417–441, 1999.

Zhilin Zhang (M'13) received the Ph.D. degree in electrical engineering from University of California at San Diego, La Jolla, CA, USA, in 2012.

He is currently a Senior Research Engineer in the Emerging Technology Lab in Samsung Research America, Dallas, TX, USA. His research interests include sparse Bayesian learning, sparse signal recovery, signal separation and decomposition, machine learning, and their applications to biomedicine, healthcare, and smart-home. He has authored or coauthored about 40 peer-reviewed journal and conference papers.

Dr. Zhang is a technical committee member in Bio-Imaging and Signal Processing of the IEEE Signal Processing Society (from January 2014 to December 2016), and a technical program committee member of a number of international conferences. He received the Excellent Master Thesis Award in 2005, Second Prize in College Student Entrepreneur Competition (on fetal heart rate monitor) in 2005, and the Samsung Achievement Award in 2013 and 2014. He is currently an Associate Editor of IEEE JOURNAL OF TRANSLATIONAL ENGINEERING IN HEALTH AND MEDICINE.
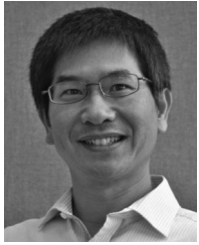
Tzyy-Ping Jung (SM'06) received the B.S. degree in electronics engineering from National Chiao-Tung University, Hsinchu, Taiwan, in 1984, and the M.S. and Ph.D. degrees in electrical engineering from The Ohio State University, Columbus, OH, USA, in 1989 and 1993, respectively.

He is currently a Research Scientist and the co-Director of the Center for Advanced Neurological Engineering, Institute of Engineering in Medicine, University of California at San Diego (UCSD), La Jolla, CA, USA. He is also an Associate Director of the Swartz Center for Computational Neuroscience, Institute for Neural Computation, and an Adjunct Professor of the Department of Bioengineering at UCSD. In addition, he is a Professor of Department of Computer Science, National Chiao Tung University, Hsinchu, Taiwan. His research interests are in the areas of biomedical signal processing, cognitive neuroscience, machine learning, time-frequency analysis of human EEG, functional neuroimaging, and brain–computer interfaces and interactions.

Dr. Jung received the Unsupervised Learning Pioneer Award from the Society for Photo-Optical Instrumentation Engineers in 2008. He is currently an Associate Editor of IEEE TRANSACTIONS ON BIOMEDICAL CIRCUITS AND SYSTEMS.

Zhouyue Pi (SM'13) received the B.E. degree (with honor) from Tsinghua University, Beijing, China, the M.S. degree from the Ohio State University, Columbus, OH, USA, and the MBA degree (with distinction) from Cornell University, Ithaca, NY, USA.

He is a Senior Director at Samsung Research America, Dallas, TX, USA, where he leads the Emerging Technology Lab doing research in next generation mobile devices, smart home solutions, and mobile health technologies. Before joining Samsung in 2006, he was with Nokia Research Center in Dallas, TX, USA and San Diego, CA, USA, where he led 3G wireless standardization and modem development for 3GPP2 1xEV-DV, 1xEV-DO, and ultra mobile broadband (UMB). In 2006–2009, he was a leading contributor to Samsung's 4G standardization efforts in 3GPP LTE and IEEE 802.16m, and to IEEE 802.11ad for 60 GHz communication. In 2009–2012, he pioneered 5G mm-wave massive MIMO technology and led the development of the world's first baseband and RF system that demonstrated the feasibility of Gb/s mobile communication at 28 GHz. He has authored more than 30 technical journal and conference papers and is the inventor of more than 150 patents and applications.

Scott Makeig received the B.S. degree, "Self in Experience," from the University of California Berkeley, Berkeley, CA, USA, in 1972, and the Ph.D. degree in music psychobiology from the University of California at San Diego, La Jolla, CA, USA, in 1985.

After spending a year in Ahmednagar, India, as a American India Foundation Research Fellow, he became a Psychobiologist at University of California at San Diego (UCSD), La Jolla, CA, USA, and then a Research Psychologist at the Naval Health Research Center, San Diego, CA, USA. In 1999, he became a Staff Scientist at the Salk Institute, La Jolla, CA, USA, and moved to UCSD as a Research Scientist in 2002 to develop and direct the Swartz Center for Computational Neuroscience. His research interests are in high-density electrophysiological signal processing and mobile brain/body imaging to learn more about how distributed brain activity supports human experience and behavior.

Bhaskar D. Rao (F'00) received the Ph.D. degree from the University of Southern California, Los Angeles, CA, USA, in 1983.

Since 1983, he has been with the University of California at San Diego (UCSD), La Jolla, CA, USA, where he is currently a Professor with the Department of Electrical and Computer Engineering and the holder of the Ericsson Endowed Chair in wireless access networks. His interests are in the areas of digital signal processing, estimation theory, and optimization theory, with applications to digital communications, speech signal processing, and human-computer interactions.

Dr. Rao has been a member of the Statistical Signal and Array Processing Technical Committee, the Signal Processing Theory and Methods Technical Committee, the Communications Technical Committee of the IEEE Signal Processing Society. His work has received several paper awards. His paper received the Best Paper Award at the 2000 Speech Coding Workshop and his students have received the Best Student Paper Awards at both the 2005 and 2006 International Conference on Acoustics, Speech and Signal Processing (ICASSP), as well as the Best Student Paper Award at Neural Information Processing Systems Conference (NIPS) in 2006. A paper he co-authored with B. Song and R. Cruz received the 2008 Stephen O. Rice Prize Paper Award in the Field of Communications Systems. A paper co-authored by him and his student received the 2012 Signal Processing Society (SPS) Best Paper Award. He was elected to the IEEE Fellow in 2000 for his contributions to the statistical analysis of subspace algorithms for harmonic retrieval.