Optimizing the Channel Selection and Classification Accuracy in EEG-Based BCI

Mahnaz Arvaneh*, *Student Member, IEEE*, Cuntai Guan, *Senior Member, IEEE*, Kai Keng Ang, and Chai Quek

***Abstract*—Multichannel EEG is generally used in brain–computer interfaces (BCIs), whereby performing EEG channel selection 1) improves BCI performance by removing irrelevant or noisy channels and 2) enhances user convenience from the use of lesser channels. This paper proposes a novel sparse common spatial pattern (SCSP) algorithm for EEG channel selection. The proposed SCSP algorithm is formulated as an optimization problem to select the least number of channels within a constraint of classification accuracy. As such, the proposed approach can be customized to yield the best classification accuracy by removing the noisy and irrelevant channels, or retain the least number of channels without compromising the classification accuracy obtained by using all the channels. The proposed SCSP algorithm is evaluated using two motor imagery datasets, one with a moderate number of channels and another with a large number of channels. In both datasets, the proposed SCSP channel selection significantly reduced the number of channels, and outperformed existing channel selection methods based on Fisher criterion, mutual information, support vector machine, common spatial pattern, and regularized common spatial pattern in classification accuracy. The proposed SCSP algorithm also yielded an average improvement of 10% in classification accuracy compared to the use of three channels (C3, C4, and Cz).**

***Index Terms*—Brain–computer interface (BCI), EEG channel selection, motor imagery, sparse common spatial pattern (SCSP) .**

I. Introduction

A BRAIN–COMPUTER INTERFACE (BCI) measures, analyzes, and decodes brain signals to provide a nonmuscular means of controlling a device. Thus, BCIs enable users with severe motor disabilities to use their brain signals for communication and control [1]–[5]. In BCI applications, the brain signals are generally measured by EEG, due to its low cost and high time resolution compared to other modalities, such as fMRI, fNIRS, etc.

Manuscript received December 14, 2010; revised February 17, 2011; accepted February 17, 2011. Date of publication March 22, 2011; date of current version May 18, 2011. This work was supported by the Agency for Science, Technology and Research (A.STAR), and the Nanyang Technological University, Singapore. *Asterisk indicates corresponding author.*

*M. Arvaneh is with the School of Computer Engineering, Nanyang Technological University, Singapore 639798 (e-mail: stuma@i2r.a-star.edu.sg).

C. Guan and K. K. Ang are with the Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), Singapore 138632 (e-mail: ctguan@i2r.a-star.edu.sg; kkang@i2r.a-star.edu.sg).

C. Quek is with the School of Computer Engineering, Nanyang Technological University, Singapore 639798 (e-mail: ashcqek@ntu.edu.sg).

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/TBME.2011.2131142

To achieve good performance, most BCIs require signals from multiple sites of the scalp [6]. However, applying a large number of EEG channels may include noisy and redundant signals that degrade the BCI performance [7], [8]. Moreover, using a large number of channels involves a prolonged preparation time that directly impacts the convenience in the use of the BCI. Therefore, selecting the least number of channels that yield the best or required accuracy can balance both needs for performance and convenience. However, how to perform optimal channel selection is not a trivial task, since selecting channels manually based on neurophysiologic knowledge does not necessarily yield optimal results compared to using all the EEG channels [6].

Various channel selection methods have been proposed in the literature. In [7]–[10], channel selection is embedded in a classifier such as support vector machine (SVM) [8], [9], which recursively eliminates the least-contributed channels in the classification accuracy. These methods generally rely on the performance of a specific classifier to evaluate the quality of a set of features. In [11] and [12], the channels are ranked based on the mutual information (MI) between the channels and the class labels. Although these methods are independent of the classifier, they rank the channels individually without considering the correlation between them. The common spatial pattern (CSP) algorithm is also used for channel selection [13], [14], whereby the channels are directly selected according to their CSP coefficients without deriving the features corresponding to each channel. The CSP algorithm is shown to be effective in discriminating two classes of EEG measurements in BCI applications [15]. It considers all the channels simultaneously, independent of any other applied machine learning algorithms. Since EEG measurements are generally contaminated by artifacts and noise, the CSP algorithm is, thus, highly sensitive to these contaminants [16]. This motivated the research for sparse solutions in the CSP algorithm [17]–[19]. In these methods, the sparse spatial filters project the signals in the most discriminative direction based on a smaller number of electrodes at the expense of lowering the accuracy.

Despite various studies, finding the optimal number and positions of EEG electrodes in a BCI application still remains a challenging issue. Although subject-independent channel sets are useful for a new subject, a relatively larger number of channels is required to suit all the subjects at the same time [10]. In contrast, subject-dependent channel selection usually incur inconvenience for the first calibration session for a subject whereby all the channels are used, but the inconvenience is alleviated once a reduced number of channels is used in the following sessions. This paper focuses on subject-dependent channel selection to answer the following research questions in EEG-based BCI applications: how many channels are required for the best classification accuracy? What is the minimum number of channels required to achieve the same accuracy as obtained by using all the channels? The term "all the channels" refers to the initial set of channels that are generally used in EEG-based BCI, for example, all the channels in a 10–20 system. In some studies [20]–[22], a minimum number of channels (namely C3, C4, and Cz) are deployed for motor imagery-based BCI. Using these three channels certainly alleviates the inconvenience of BCI preparation time, but at the expense of the accuracy of the system. This paper also investigates the extent of accuracy deterioration in the use of three channels compared to the use of optimally selected channels.

To address the research questions, a novel sparse common spatial pattern(SCSP) algorithm is proposed in this paper for optimal EEG channel selection. The proposed algorithm minimizes the number of channels by sparsifying the common spatial filters within a constraint of classification accuracy. In order to evaluate the performance of the proposed SCSP algorithm, two datasets from publicly available BCI competitions are used, one with moderate number of initial channels (22 channels) [23], and another with much denser electrodes (118 channels) [24]. The performance of the proposed algorithm is also compared with several other channel selection methods, based on the Fisher criterion (FC), MI, SVM, CSP coefficients, and the regularized common spatial pattern (RCSP) in [17].

The remainder of this paper is organized as follows. Section II describes the proposed method. The applied datasets and the performed experiments are explained in Section III. Section IV presents the experimental results and finally Section V concludes this paper.

II. Methodology

A. CSP Algorithm as an Optimization Problem

The CSP algorithm [15], [25] is effective in discriminating two classes of EEG data by maximizing the variance of one class while minimizing the variance of the other class. Let $\mathbf{X} \in \mathbf{R}^{\times}$ denotes a matrix that represents the EEG of a single trial, where and denote the number of channels and number of measurement samples, respectively. The CSP algorithm projects $\mathbf{X}$ to spatially filtered $\mathbf{Z}$ as

$$\mathbf{Z} \quad \mathbf{W}\mathbf{X} \tag{1}$$

where the rows of the projection matrix $\mathbf{W}$ are the spatial filters and the columns of $\mathbf{W}^{-}$ are the CSP.

The CSP algorithm computes $\mathbf{W}$ by simultaneous diagonalization of the covariance matrices from both classes. For each centered $\mathbf{X}$, the normalized covariance matrix can be obtained from

$$\mathbf{C} \quad \frac{\mathbf{X}\mathbf{X}}{\mathbf{X}\mathbf{X}} \tag{2}$$

where denotes the transpose operator, and $\mathbf{X}$ gives the sum of diagonal elements of $\mathbf{X}$. The covariance matrix of each class $\mathbf{C}$ and $\mathbf{C}$ are computed by averaging over the multiple trials of EEG data. The composite covariance matrix and its eigenvalue decomposition are given by

$$\mathbf{C} \quad \mathbf{C} \quad \mathbf{C} \quad \mathbf{F} \; \psi \mathbf{F} \tag{3}$$

where $\mathbf{F}$ is a matrix of normalized eigenvectors with corresponding matrix of eigenvalues ψ.

The whitening transformation matrix

$$\mathbf{P} \quad \psi^{-} \quad \mathbf{F} \tag{4}$$

transforms the covariance matrices as

$$\mathbf{C}' \quad \mathbf{P}\mathbf{C}\,\mathbf{P} \qquad \mathbf{C}' \quad \mathbf{P}\mathbf{C}\,\mathbf{P} \tag{5}$$

where $\mathbf{C}'$ and $\mathbf{C}'$ share common eigenvectors, and the sum of corresponding eigenvalues for the two matrices are always one, such that

$$\mathbf{C}' \quad \mathbf{U}\boldsymbol{\Lambda}\,\mathbf{U} \qquad \mathbf{C}' \quad \mathbf{U}\boldsymbol{\Lambda}\,\mathbf{U} \qquad \boldsymbol{\Lambda} \quad \boldsymbol{\Lambda} \quad \mathbf{I} \tag{6}$$

where $\mathbf{I}$ is the identity matrix. $\mathbf{U}$ and $\boldsymbol{\Lambda}$, respectively, denote the matrix of eigenvectors and the diagonal matrix of eigenvalues. The eigenvalues are sorted in descending order and the CSP projection matrix is defined as $\mathbf{W} \quad \mathbf{U} \; \mathbf{P}$, which projects the covariance matrix of each class as

$$\mathbf{C}'' \quad \mathbf{U}\;\mathbf{P}\mathbf{C}\,\mathbf{P}\;\mathbf{U} \quad \boldsymbol{\Lambda} \qquad \mathbf{C}'' \quad \mathbf{U}\;\mathbf{P}\mathbf{C}\,\mathbf{P}\;\mathbf{U} \quad \boldsymbol{\Lambda} \tag{7}$$

where $\boldsymbol{\Lambda} \quad \boldsymbol{\Lambda} \quad \mathbf{I}$. Since the sum of two corresponding eigenvalues is always one, the maximum variance of one class leads to the minimum variance of the other class. This property makes the CSP effective for classification of two distributions. The projection of the whitened EEG signals onto the eigenvectors corresponding to the largest eigenvalues of $\boldsymbol{\Lambda}$ and $\boldsymbol{\Lambda}$ gives feature vectors that are optimal for discriminating two groups of EEG in the least-squares sense [26].

The CSP algorithm, in computing the projection matrix $\mathbf{W}$, can be formulated as an optimization problem given by

$$\begin{aligned} &_{\mathbf{w}} \quad \sum \mathbf{w}\,\mathbf{C}\,\mathbf{w} \qquad \sum \; \mathbf{w}\,\mathbf{C}\,\mathbf{w} \\ &\qquad \mathbf{w} \;\; \mathbf{C} \quad \mathbf{C} \;\; \mathbf{w} \qquad\qquad \{ \qquad\qquad \} \\ &\mathbf{w} \;\; \mathbf{C} \quad \mathbf{C} \;\; \mathbf{w} \qquad\qquad\qquad \{ \qquad\qquad \} \;\; / \end{aligned} \tag{8}$$

where $\mathbf{C}$ denotes the covariance matrix of class . Unknown weights $\mathbf{w} \in \mathbf{R}^{\times}$, $\{ \quad \}$, respectively, indicate the first and last rows of CSP projection matrix, corresponding to the largest eigenvalues of $\boldsymbol{\Lambda}$ and largest eigenvalues of $\boldsymbol{\Lambda}$.

The CSP algorithm is formulated as a quadratically constrained quadratic programming problem in (8) in order to formulate the SCSP algorithm in the following section. In this way, the CSP algorithm optimizes $\mathbf{W}$ using constraints to keep the covariance matrices of both classes diagonal.

B. SCSP

The rows of the CSP projection matrix give nonuniform weights to channels, so that the differences between two classes of the EEG signals are maximized. Hence, the CSP spatial filters can be seen as source distribution vectors.

The use of the CSP algorithm for EEG channel selection was proposed by Wang *et al.* [13]. In the proposed method, four channels corresponding to the maximal CSP vector coefficients were selected as the optimal channels. However, the weights of the CSP are dense (not sparse), and only a few number of channels may have negligible weights compared to the rest. Thus, by eliminating other channels, the remaining signals can no longer be projected onto the direction that best discriminates the two classes of EEG signals. Moreover, since EEG measurements are generally contaminated by artifacts and noise, the CSP algorithm that is based on the covariance matrices of EEG trials, can be distorted by these contaminants [16].

These issues motivated the approach to sparsify the CSP spatial filters to emphasize on a limited number of channels with high variances between the classes, and to discard the rest of the channels with low or irregular variances that may be due to noise or artifacts.

Sparsity can be induced in the CSP algorithm by adding an norm regularization term into the optimization problem given in (8). $\|\mathbf{x}\|$, the norm of $\mathbf{x}$, is the measure giving the number of nonzero elements of $\mathbf{x}$. However, solving a problem with the norm is combinatorial in nature and, thus, computationally prohibitive. Furthermore, since an infinitesimal value is treated the same as a large value, the presence of noise in the data may render the norm completely ineffective in inducing sparsity [27]. Therefore, several alternative measures were proposed as approximations of norm [27].

Among the proposed measures, the norm, ; and norm of $\mathbf{x}$ are commonly used in place of the norm. The norm of $\mathbf{x}$, , is defined as

$$\|\mathbf{x}\| \qquad \left(\sum |\mathbf{x} |\right) \tag{9}$$

where denotes the total number of elements of the vector $\mathbf{x}$. Although the and norms are very popular in regularization algorithms, the norm defined in (10) is sometimes used in place of them

$$\frac{\|\mathbf{x}\|}{\|\mathbf{x}\|} \tag{10}$$

The use of the norm in (10) ensures that the sparsest possible vector whereby only a single element is nonzero has a sparseness of one, and a vector with all equal nonzero elements has a sparseness of $\sqrt{\ }$. Since the norm increases when the sparsity decreases, it can be considered as a nonsparsity measure.

Hurely and Rickard [27] compared commonly used nonsparsity measures based on intuitive attributes, and revealed that the norm satisfied more desirable attributes of nonsparsity compared to the and lp norms. The norm is limited to the range $\sqrt{\ }$, while the and norms are affected by the magnitude of the components. Moreover, according to Dalton's first law [27], a representation is sparser if it has one large component rather than dividing up the large component into two

TABLE I
OUTCOME OF THE NON-SPARSITY MEASURES FOR THREE EXAMPLE

	[0 0 3 5]	[0 1 3 4]	[0 0 6 10]	Range
l_1	8	8	16	$[0,\infty)$
$l_p, p = 0.5$	15.74	22.39	31.49	$[0,\infty)$
l_1/l_2	1.372	1.56	1.372	$[0,\sqrt{n}]$

smaller ones. The and norms preserve this, whereas the norm gives the same sparseness for both conditions. Furthermore, the norm is scale invariant while the two norms are not.

Table I illustrates the differences between these norms using a toy example. Although the vector [0 0 3 5] is sparser than the vector [0 1 3 4] due to Dalton's first law [27], gives the same sparseness. Since the norm is scale invariant, it gives the same sparseness for [0 0 3 5] and [0 0 6 10], but the and norms do not give the same sparseness.

The properties of the norm motivated its use as the nonsparsity measure in the proposed SCSP algorithm:

$$\mathbf{w} \quad - \left(\sum \mathbf{w}\ \mathbf{C}\ \mathbf{w} \quad \sum \mathbf{w}\ \mathbf{C}\ \mathbf{w}\right) \quad \sum \frac{\|\mathbf{w}\ \|}{\|\mathbf{w}\ \|}$$
$$\mathbf{w}\ \mathbf{C} \quad \mathbf{C}\ \mathbf{w} \qquad \{ \qquad \}$$
$$\mathbf{w}\ \mathbf{C} \quad \mathbf{C}\ \mathbf{w} \qquad \{ \qquad \} \ / \tag{11}$$

where $\leq \ \leq$ is a regularization parameter that controls the sparsity (number of removed channels) and the classification accuracy. When , the solution is essentially the same as the CSP algorithm. The methodology to find the optimal is discussed in Section IV. The proposed SCSP algorithm is a nonlinear optimization problem, and due to the equality constraints it is a nonconvex optimization problem. It can be solved using several methods such as sequential quadratic programming (SQP) and augmented Lagrangian methods. In this study, for $/$, spatial filters obtained from the CSP algorithm are used as the initial point.

In (11), the constraints lead to diagonal covariance matrices in both classes. The projected signals obtained from sparse spatial filters are, thus, uncorrelated in both classes [25], while the RCSP algorithm in [17] optimizes the spatial filters of different classes independently without considering the correlation between them. Our proposed SCSP algorithm considers the correlation between the spatial filters of different classes in order to achieve a better discrimination. To investigate this, the results of using the proposed algorithm are compared with the results of the RCSP algorithm [17] in Section IV.

C. SCSP-Based Channel Selection

To select channels using the proposed method, first two sparse spatial filters corresponding to two motor imagery tasks are obtained by solving the optimization problem given in (11) with . Since the value controls the number of selected channels and the achieved classification accuracy, it should be carefully chosen to fulfill the application needs. The effects of on the system performance will be discussed in detail in Section IV.

After obtaining the sparse filters, channels corresponding to the zero elements in both of the spatial filters are discarded, and the rest are defined as the selected channels.

To compare and consider the importance of each selected channel, a ranking method is proposed as follows: first, the top-ranked channels for each motor imagery task are determined from the maximum of the absolute value of the corresponding sparse spatial filter. Let $|\mathbf{w}\ |$ and $|\mathbf{w}\ |$, respectively, denote the *i*th best channel of the first and second motor imagery tasks, with corresponding absolute spatial filter coefficients $|\mathbf{w}\ |$ and $|\mathbf{w}\ |$. Consequently

$$
\begin{aligned}
&|\mathbf{w}\ | \geq |\mathbf{w}\ | \qquad \leq \\
&|\mathbf{w}\ | \geq |\mathbf{w}\ | \qquad \leq
\end{aligned}
\tag{12}
$$

where *p* and *q*, respectively, denote the total number of remaining channels in the first and second sparse spatial filters. Thereafter, the overall channel ranks are obtained as

$$
\begin{cases}
\quad - \qquad |\mathbf{w}\ |\ |\mathbf{w}\ | \qquad \leq\ \leq \\
\qquad |\mathbf{w}\ |\ |\mathbf{w}\ | \qquad \leq\ \leq \\
\text{If} \qquad \text{then} \qquad |\mathbf{w}\ | \qquad \leq\ \leq \\
\text{If} \qquad \text{then} \qquad |\mathbf{w}\ | \qquad \leq\ \leq
\end{cases}
\tag{13}
$$

where denotes the minimum of and . Finally, since some channels may have been iterated twice in , the lower rank is discarded. As shown in (13), in this method channels are pairwisely ranked from both motor imagery areas.

III. Experiments

A. Data Description

In this study, the EEG data of 14 subjects from two publicly available datasets, one with a moderate number of channels and another with a large number of channels, were used. These two datasets are described as follows.

1) *Dataset IIa [23] from BCI competition IV:* This dataset contains EEG data of nine subjects recorded using 22 channels. During the recording session, the subjects were instructed with visual cues to perform one of the four motor imagery tasks: left hand, right hand, feet, or tongue. In this study, only the EEG data from right and left hand motor imagery tasks were used. The EEG data for each subject comprised of a training and testing sets of which each set included 72 trials for each motor imagery task. The testing set was recorded in another day.

2) *Dataset IVa [24] from BCI competition III [28]:* This dataset contains EEG data of five subjects recorded using 118 channels. During the recording session, the subjects were instructed to perform one of two motor imagery tasks: right hand or foot. The EEG data for each subject comprised of a training and testing sets of which both included 280 single-trials. Since the purpose of this study is not to investigate the performance of the proposed algorithm on a small training dataset, the 280 single-trials data were equally distributed so that the training set included 140 single-trials and the testing set included the remaining 140 single-trials.

B. Data Processing

For each dataset, the EEG data from 0.5 to 2.5 s after the visual cue were used whereby the selected time segment was used by the winner of the BCI competition IV dataset IIa [29]. The EEG data were band-pass filtered using elliptic filters from 8 to 35 Hz, since this frequency band included the range of frequencies that are mainly involved in performing motor imagery. The filtered EEG data from the training set were then used to select the optimal channels. The optimal channels were selected using the first and the last sparse spatial filters obtained in (11) using . Subsequently, the classification accuracy was evaluated on the testing set as follows: first, the CSP was retrained over the selected channels. Subsequently, the signals from selected channels were spatially filtered using the first and last three spatial filters of the retrained CSP using . Finally, the variance of the spatially filtered signals were applied as the inputs of the SVM classifier [26].

For the purpose of benchmarking, the classification accuracies of several EEG channel selection methods based on FC [8], MI [11], SVM [8], CSP [13], [19], and RCSP [17] were also evaluated on the two datasets. The FC- and MI-based channel selection methods are filter approaches that rank the channels based on maximizing the MI and FC between the channels and the class labels. The SVM-based channel selection method is a wrapper approach that recursively eliminates the least contributed channels based on the classification accuracy from the SVM classifier. The CSP-based channel selection method uses the CSP coefficients to select the channels. The RCSP channel selection method selects the channels by inducing the sparsity in the spatial filters. Compared to the proposed algorithm, the RCSP algorithm sparsifies the spatial filters without keeping the covariance matrices of the projected signals diagonal.

The SVM classifier was used in the classification step for all the mentioned channel selection methods. The radial basis function was used as the kernel function of the SVM as suggested in [30], and the hyperparameters were obtained by a grid search using cross validation on the training data [30].

IV. Results and Discussion

A. SCSP Optimization Problem and the Regularization Term

The optimization problem in (11) is a nonconvex programming problem because of the quadratic equality constraint. It can be solved using several methods such as SQP. In this study, the package fmincon available in MATLAB based on the SQP method was used to solve this optimization problem [31], [32]. The convergence tolerance was manually set to $^{-}$. Hence, the algorithm converges when the maximum constraints violation is less than $^{-}$, and the relative change in the objective function is less than $^{-}$ [32].

Fig. 1 compares the performance of the SCSP algorithm using the norm as the regularization term with the performance of the SCSP algorithm that uses the norm instead. For each

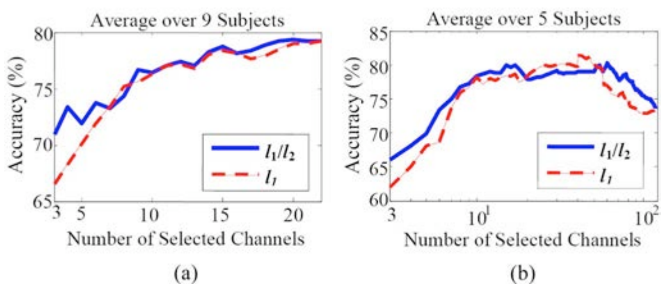

Fig. 1. Performance comparison of SCSP channel selection algorithm based on two different regularization terms for (a) dataset IIa, BCI Competition IV and (b) dataset IVa, BCI Competition III.

TABLE II
COMPARING THE AVERAGE TIME AND NUMBER OF ITERATIONS TO CONVERGE FOR THE SCSP ALGORITHM BASED ON TWO REGULARIZATION TERMS

	Dataset IVa,Competition III		Dataset IIa,Competition IV	
	Time(s)	**♯Iter**	**Time(s)**	**♯Iter**
l_1	32.7	630	1.37	230
l_1/l_2	50.1	1001	2.63	454

algorithm, different numbers of channels, ranged from three to all the channels, were selected by varying the *r* value. Thereafter, the classification accuracy corresponding to each set of selected channels was calculated on the testing data. Fig. 1 shows that the use of the norm instead of the norm in the proposed SCSP algorithm leads better channel selection accuracies, particularly when the number of selected channels is relatively small. The *x*-axis in Fig. 1(b) has been drawn in log scale to emphasize on the small number of channels and present them in a more informative way.

Table II compares the averaged elapsed time and number of iterations required to converge for the SCSP algorithm with two different regularization terms using MATLAB 7.5 and an Intel Quad 2.83 GHz CPU. Table II shows that using the norm as the regularization term resulted in faster convergence with less number of iterations. Moreover, the increase of the number of channels from 22 to 118 yielded considerable increases in the calculation time and number of iterations for both regularization terms.

Since the computation time is not an important issue for off-line channel selection in BCI applications, considering the better accuracy of using the norm compared to the norm, the former is, therefore, used in the proposed SCSP algorithm. However, the proposed SCSP algorithm using either regularization terms may not be practical for online BCI applications that requires a fast response.

B. SCSP-Based Channel Selection With Different Criteria

The regularization parameter *r*, given in (11), controls the classification accuracy and the number of selected channels. Increasing *r* results in selecting less number of channels, but may also result in decreasing the classification accuracy due to the exclusion of some informative channels. Therefore, the optimal *r* value should be calibrated according to the requirements of a specific BCI application.

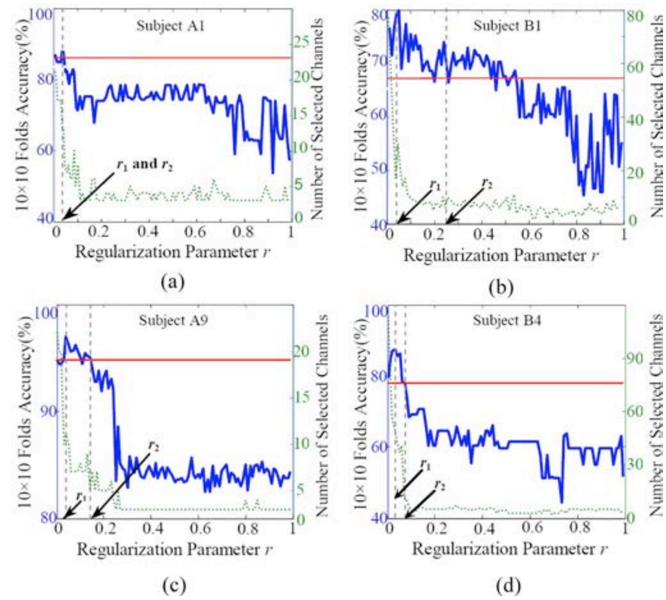

Fig. 2. Effects of varying *r* on the accuracy and number of selected channels for four subjects. The dotted, narrow, and thick lines, respectively, denote number of selected channels, all the channels accuracy and × -fold accuracy. and indicate the optimal *r* values for SCSP1 and SCSP2, respectively.

In this study, the optimal subject-specific *r* was chosen from a set of values ∈ { } applied on the training data. First, for each *r*, a set of selected channels was determined by solving the optimization problem in (11). Subsequently, the classification accuracy corresponding to each set of the selected channels was computed using × -fold cross validation on the training data. Finally, the optimal *r* was selected based on the cross-validation accuracy and the given criterion.

Two channel selection criteria were investigated. The first criterion maximizes the accuracy by removing noisy and irrelevant channels. The second criterion minimizes the number of selected channels while maintaining the classification accuracy such that it is greater or equal to the classification accuracy from using all the channels. Based on the cross-validation accuracies obtained from testing different *r* values on the training data, the optimal *r* is selected to meet these criteria. In this paper, the SCSP algorithm using the first and second criteria are, respectively, abbreviated as SCSP1 and SCSP2.

Fig. 2 illustrates how to select optimal *r* values satisfying the mentioned criteria. It also presents the effect of varying *r* on the accuracy and the number of selected channels for four subjects: (a) A1, (b) B1, (c) A9, and (d) B4. The optimal *r* values for the two mentioned criteria were also indicated for each subject. In Fig. 2, the set of selected channels corresponding to each *r* was found by solving (11), and subsequently the classification accuracy was calculated over each set of selected channels using × -fold cross validation on the training data.

Fig. 2 shows that the use of small values of *r* improved the accuracy by removing some noisy and redundant EEG channels, while increased values of *r* reduced the number of channels but also decreased the classification accuracy. When *r* is increased further to approach the value 1, increased variations in the

TABLE III
PERFORMANCE COMPARISON OF DIFFERENT METHODS APPLIED ON FIRST DATASET WITH OVERALL 22 CHANNELS

Dataset IIa, BCI Competition IV						
			SCSP1		SCSP2	
Subject	**All Ch Acc(%)**	**(C3,C4,Cz) Acc(%)**	**Acc (%)**	**♯ Selected Ch**	**Acc (%)**	**♯ Selected Ch**
A1	90.97	75.69	91.66	13	91.66	13
A2	56.25	53.47	67.36	9	60.41	4
A3	96.52	93.05	97.91	14	97.14	12
A4	72.91	68.05	72.22	14	70.83	11
A5	63.88	53.47	65.27	11	63.19	9
A6	63.88	61.11	66.67	14	61.11	10
A7	79.86	57.63	84.72	19	78.47	15
A8	97.22	86.80	97.22	15	95.13	5
A9	91.66	88.88	91.66	10	93.75	5
Mean	79.23	70.90	81.63	13.22	79.07	8.55
Std	15.63	15.72	13.7	2.99	15.61	3.90
p-value	0.006	–	0.003	–	0.004	–

P-value denotes the paired T-test between results of (C3,C4,CZ) and other results.
(CH: Channels, ACC: Accuracy, ♯ : Number).

TABLE IV
PERFORMANCE COMPARISON OF DIFFERENT METHODS APPLIED ON FIRST DATASET WITH OVERALL 118 CHANNELS

Dataset IVa, BCI Competition III						
			SCSP1		SCSP2	
Subject	**All Ch Acc(%)**	**(C3,C4,Cz) Acc(%)**	**Acc (%)**	**♯ Selected Ch**	**Acc (%)**	**♯ Selected Ch**
B1	74.28	54.28	80.71	17	71.42	7
B2	94.28	80	97.14	12	95.71	10
B3	49.28	55	57.14	33	57.14	3
B4	77.14	70	85	36	77.85	10
B5	72.85	87.14	91.42	15	94.28	10
Mean	73.56	69.28	82.28	22.6	79.28	7.6
Std	16.06	14.69	15.38	11.05	16.19	3.08
p-value	0.535	–	0.043	–	0.023	–

P-value denotes the paired T-test between results of (C3,C4,CZ) and other results.
(CH: Channels, ACC: Accuracy, ♯ : Number).

classification accuracy were observed. The floor effect on the number of selected channels, observed for some r values greater than , showed that further increase in the r value did not yield further reduction in the number of selected channels. According to Fig. 2, evaluating a small subset of r values suffices to find the optimal r to meet the mentioned criteria.

Tables III and IV summarize the performance (classification accuracy and number of selected channels) of all subjects. The SCSP results based on two mentioned criteria are compared with the results of using all the channels, and in particular benchmarked with three typical motor imagery channels (C3, C4, and Cz). The last row of the tables presents the p-values obtained from the paired t-test between the results of three (C3, C4, and Cz) channels and the other results.

Table III shows the results on the dataset with moderate number of channels. According to the results, the proposed SCSP1 algorithm yielded an average improvement of 2.4% in the classification accuracy by decreasing the number of channels to 13.22 from 22. The improvement in the classification accuracy for some subjects such as A2 is substantial (around 11%). This showed that the proposed SCSP1 algorithm is capable of removing redundant and noisy channels. On average, the proposed SCSP1 algorithm reduced 40% of the channels and achieved a marginal improvement in the classification accuracy (although the improvement is not statistically significant,

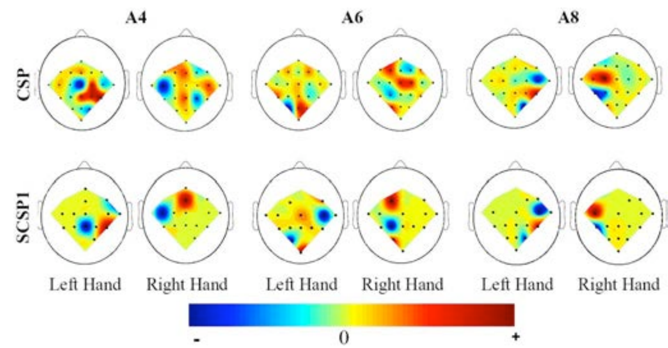

Fig. 3. Spatial filters obtained from CSP and SCSP1 algorithms, for subjects A4, A6, and A8 (22 electrodes). The dots denote the selected channels.

p-value 0.08). Likewise, the proposed SCSP2 algorithm reduced on average 61.2% of the channels with only 0.16% drop in the classification accuracy. Although the selected channels were evaluated on the data recorded from another session in a different day, the obtained accuracies fulfill both mentioned criteria. This shows that the selected channels were transferable to another BCI session recorded on a different day. The results also showed that the proposed SCSP algorithm using both criteria yielded significantly better classification accuracies (average 9.45% more) compared to the use of three typical channels.

Table IV shows the results on the dataset with large number of channels. The proposed SCSP1 algorithm yielded an average improvement of 8.6% in the classification accuracy with the use of only 22.6 from 118 channels. On average the SCSP1 algorithm reduced 81% of the channels and achieved significantly better classification accuracies than using all the channels (p-value 0.04). Similarly, the proposed SCSP2 algorithm decreased on average 93% of channels with interestingly 5.72% improvement in the classification accuracy. The results also present that the proposed SCSP algorithm using both criteria yielded an average improvement of 11.5% in the classification accuracy compared to the use of the fixed (C3, C4, and Cz) layout.

Comparison between Table III and Table IV reveals that in both datasets, the proposed SCSP channel selection significantly reduced the number of channels with fulfilling the chosen criteria. Moreover, the improvement in accuracy and the reduction in number of channels were more salient when the SCSP algorithm applied on the dataset with large number of channels. Comparing Tables III and IV also shows that on average, dataset IVa required more channels to achieve higher accuracy compared to dataset IIa. This may be due to the performance of different motor imagery actions in these two datasets: right hand and foot in dataset IVa, right and left hand in dataset IIa.

Figs. 3 and 4 present some examples of the spatial filters obtained from CSP and the proposed SCSP1 algorithm. The results showed that CSP filters have large weights in several unexpected locations from a neurophysiological point of view. On the contrary, the SCSP filters have strong weights over the motor cortex areas and smooth weights over the other areas.

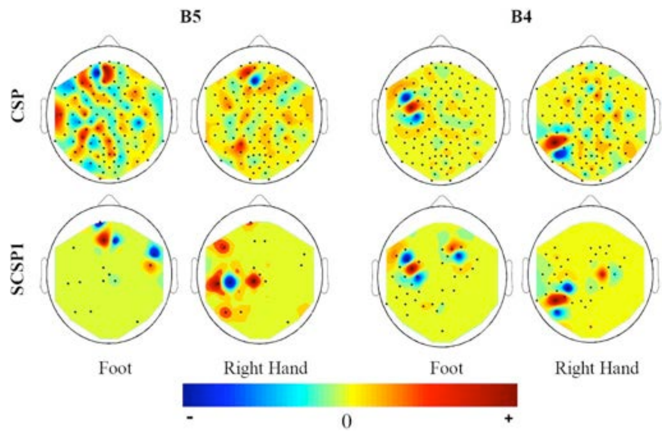

Fig. 4. Spatial filters obtained from CSP and SCSP1 algorithms, for subjects B4, B5 (118 electrodes). The dots denote the selected channels.

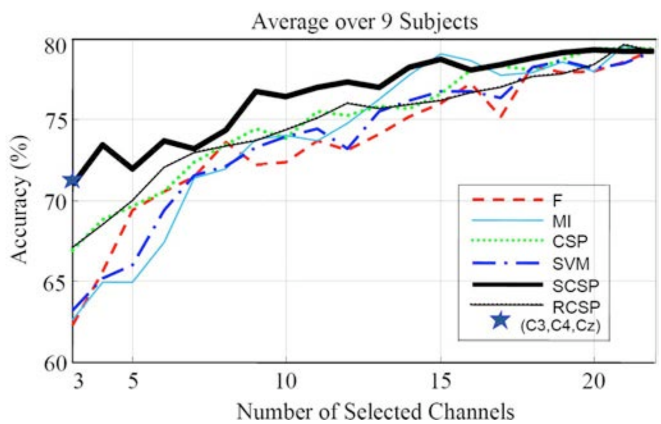

Fig. 5. Comparison of different EEG channel selection algorithms applied on Dataset IIa, BCI Competition IV.

This showed that the proposed SCSP yielded filters that are neurophysiologically more relevant and interpretable.

C. Comparing with Other Channel Selection Methods

As explained in Section II, in the proposed SCSP optimization problem given in (11), the regularization parameter r controls the number of selected channels. Hence, change of r results in selecting different number of channels. To consider the performance of the proposed algorithm in selecting different number of channels, a set of r values were applied on the training data, and to select each specific number of channels (from 3 to all the channels) the optimal r was defined. Finally for comparison purpose, the classification accuracies of testing data obtained from optimal r values were compared with the results of other channel selection methods based on the FC, MI, SVM, CSP, and the RCSP algorithm in [17]. In the RCSP algorithm, the first and last obtained spatial filters were used for channel selection.

Figs. 5 and 6 depict averaged accuracy versus different number of channels selected by six different channel selection algorithms. The x-axis in Fig. 6 has been drawn in log scale to emphasize on small number of channels and present them in a more informative way.

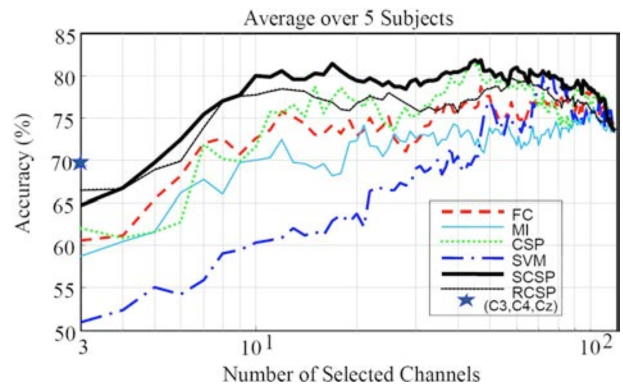

Fig. 6. Comparison of different EEG channel selection algorithms applied on Dataset IVa, BCI Competition III.

The results in Figs. 5 and 6 show that the proposed algorithm outperformed the other channel selection methods, particularly when the number of selected channels are relatively small. Fig. 6 also shows that the three selected channels using SCSP algorithm yielded less averaged accuracy compared to using (C3,C4, and Cz). It may be because that the proposed SCSP needed a big value of r to select only three channels among 118 channels. As can be seen in the optimization problem 11, a big value of r highly increases the weight of the sparsity term versus the other term that controls the separability of two classes. Therefore, although the proposed algorithm outperformed the other introduced methods, it may not achieve the highest accuracy in selecting a very few number of channels among a large number of channels.

V. Conclusion

This paper focused on subject-dependent channel selection in motor imagery-based BCI applications. For this purpose, we investigated the reduction of channels whereby the classification accuracy is constrained to an acceptable range. This is achieved by solving an optimization problem that induces sparsity in the common spatial filters.

For benchmarking purpose, the proposed SCSP algorithm using two different criteria were applied on two datasets with 22 and 118 channels each. The results showed that the proposed SCSP algorithm using the first criterion yielded the best classification accuracy by removing the most number of channels, and using the second criterion retained the least number of channels without compromising the classification accuracy from using all the channels. The proposed SCSP algorithm using both the criteria, also yielded an average improvement of 10% in classification accuracy compared to the use of three channels (C3, C4, and Cz).

A comparative study of the proposed algorithm with other channel selection methods using FC, MI, SVM, CSP, and RCSP showed that our method outperformed the others, especially in the case where the number of selected channels is relatively small.

A visualization of the obtained sparse spatial filters showed that the proposed algorithm improved the results by emphasizing on a limited number of channels with high variances between the classes, and discarding the rest of the channels with low or irregular variances that may be due to noise or artifacts.

The extension of the proposed SCSP algorithm for channel selection to multiclass paradigms can be either performed by computing the SCSP using the one-versus-rest approach [33], or by using joint approximate diagonalization [34]. The former approach is conceptually identical to the SCSP for two-class paradigms. For multiclass paradigms, the proposed SCSP algorithm for channel selection can be performed on each of the classes versus the rest, and subsequently the selected channels are consolidated from all the sparse spatial filters. The latter approach finds sparse spatial filters that approximately diagonalize multiple covariance matrices from all the classes to maximize the separation between all the classes.

ACKNOWLEDGMENT

The authors would like to thank Graz and Berlin BCI groups for providing the BCI competition datasets, and Mr. H. Ahmadi, Dr. F. Lotte, and anonymous reviewers for their constructive comments.

REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, “Brain-computer interfaces for communication and control,” *Clin. Neurophysiol.*, vol. 113, pp. 767–791, 2002.

[2] N. Birbaumer, “Brain-computer-interface research: Coming of age,” *Clin. Neurophysiol.*, vol. 117, pp. 479–483, 2006.

[3] E. A. Curran and M. J. Stokes, “Learning to control brain activity: A review of the production and control of EEG components for driving brain-computer interface (BCI) systems,” *Brain Cogn.*, vol. 51, pp. 326–336, 2003.

[4] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, D. J. McFarland, P. H. Peckham, G. Schalk, E. Donchin, L. A. Quatrano, C. J. Robinson, and T. M. Vaughan, “Brain-computer interface technology: A review of the first international meeting,” *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 2, pp. 164–173, Jun. 2000.

[5] J. R. Wolpaw, D. J. McFarland, and T. M. Vaughan, “Brain-computer interface research at the Wadsworth Center,” *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 2, pp. 222–226, Jun. 2000.

[6] B. Blankertz, F. Losch, M. Krauledat, G. Dornhege, G. Curio, and K. R. Müller, “The Berlin brain-computer interface: Accurate performance from first-session in BCI-naive subjects,” *IEEE Trans. Biomed. Eng.*, vol. 55, no. 20, pp. 2452–2462, Oct. 2008.

[7] F. Popescu, S. Fazli, B. Yakob, B. Blankertz, and K. R. Müller, “Single trial classification of motor imagination using 6 dry EEG electrodes,” *PLoS ONE*, vol. 2, 2007, pp. 1–5.

[8] T. N. Lal, M. Schröder, T. Hinterberger, J. Weston, M. Bogdan, N. Birbaumer, and B. Schölkopf, “Support vector channel selection in BCI,” *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1003–1010, Jun. 2004.

[9] M. Schröder, T. N. Lal, T. Hinterberger, M. Bogdan, N. J. Hill, N. Birbaumer, W. Rosenstiel, and B. Schölkopf, “Robust EEG channel selection across subjects for brain-computer interfaces,” *EURASIP J. Appl. Signal Process.*, vol. 19, pp. 3103–3112, 2005.

[10] C. Sannelli, T. Dickhaus, S. Halder, E. M. Hammer, K. R. Müller, and B. Blankertz, “On optimal channel configurations for SMR-based brain-computer interfaces,” *Brain Topogr.*, vol. 23, pp. 186–193, 2010.

[11] T. Lan, D. Erdogmus, A. Adami, M. Pavel, and S. Mathan, “Salient EEG channel selection in brain computer interfaces by mutual information maximization,” in *Proc. 27th Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2005, pp. 7064–7067.

[12] T. Lan, D. Erdogmus, A. Adami, S. Mathan, and M. Pavel, “Channel selection and feature projection for cognitive load estimation using ambulatory EEG,” *Comput. Intell. Neurosci.*, vol. 2007, pp. 8–20, 2007.

[13] Y. Wang, S. Gao, and X. Gao, “Common spatial pattern method for channel selelction in motor imagery based brain-computer interface,” in *Proc. 27th Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2005, pp. 5392–5395.

[14] J. Meng, G. Liu, G. Huang, and X. Zhu, “Automated selecting subset of channels based on CSP in motor imagery brain-computer interface system,” in *Proc. IEEE Int. Conf. Rob. Biomimetics*, 2009, pp. 2290–2294.

[15] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K. R. Müller, “Optimizing spatial filters for robust EEG single-trial analysis,” *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, Jan. 2008.

[16] M. Krauledat, G. Dornhege, B. Blankertz, and K. R. Müller, “Robustifying EEG data analysis by removing outliers,” *Chaos Complexity Lett.*, vol. 2, pp. 259–274, 2007.

[17] J. Farquhar, N. J. Hill, T. N. Lal, and B. Schölkopf, “Regularised CSP for sensor selection in BCI,” in *Proc. 3rd Int. Brain-Computer Interface Workshop Training Course*, 2006, pp. 14–15.

[18] X. Yong, R. K. Ward, and G. E. Birch, “Sparse spatial filter optimization for EEG channel reduction in brain-computer interface,” in *Proc. IEEE Int. Conf. Acoustics, Speech Signal Process.*, 2008, pp. 417–420.

[19] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek, “Optimizing EEG channel selection by regularized spatial filtering and multi band signal decomposition,” in *Proc. Int. Assoc. Sci. Technol. Develop. Int. Conf. Biomed. Eng.*, Innsbruck, 2010, pp. 86–90.

[20] G. Pfurtscheller, C. Neuper, D. Flotzinger, and M. Pregenzer, “EEG-based discrimination between imagination of right and left hand movement,” *Electroencephalogr. Clin. Neurophysiol.*, vol. 103, pp. 642–651, 1997.

[21] R. Scherer, G. R. Müller, C. Neuper, B. Graimann, and G. Pfurtscheller, “An asynchronously controlled EEG-based virtual keyboard: Improvement of the spelling rate,” *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 979–984, Jun. 2004.

[22] R. Leeb, F. Lee, C. Keinrath, R. Scherer, H. Bischof, and G. Pfurtscheller, “Brain-computer communication: Motivation, aim, and impact of exploring a virtual apartment,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 15, no. 4, pp. 473–482, Dec. 2007.

[23] M. Naeem, C. Brunner, R. Leeb, B. Graimann, and G. Pfurtscheller, “Seperability of four-class motor imagery data using independent components analysis,” *J. Neural Eng.*, vol. 3, pp. 208–216, 2006.

[24] G. Dornhege, B. Blankertz, G. Curio, and K. R. Müller, “Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms,” *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 993–1002, Jun. 2004.

[25] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. New York: Academic Press, 1990, pp. 26–34..

[26] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, “Optimal spatial filtering of single trial EEG during imagined hand movement,” *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 441–446, Dec. 2000.

[27] N. Hurley and S. Rickard, “Comparing measures of sparsity,” *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4723–4741, Oct. 2009.

[28] B. Blankertz, K. R. Müller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlögl, G. Pfurtscheller, J. R. Millan, M. Schröder, and N. Birbaumer, “The BCI competition III: Validating alternative approaches to actual BCI problems,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 153–159, Jun. 2006.

[29] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, “Filter bank common spatial pattern (FBCSP) in brain-computer interface,” in *Proc. IEEE Int. Joint Conf. Neural Networks*, Hong Kong, 2008, pp. 2390–2397.

[30] C. Hsu, C. Chang, and C. Lin. (Jul. 2003). A practical guide to support vector classification [Online]. Available: http://www.csie.ntu.edu. tw/cjlin/papers.html

[31] M. Powell, “A fast algorithm for nonlinearly constrained optimization calculations,” in *Numerical Analysis*, vol. 630, G. Watson, Ed., Berlin, Germany: Springer-Verlag, 1978, pp. 144–157.

[32] M. Powell, “Variable metric methods for constrained optimization,” in *Computing Methods Applied Sciencesand Engineering*, vol. 704, R. Glowinski, J. L. Lions, and Iria Laboria, Eds., Berlin/Springer, 1979, pp. 62–72.

[33] W. Wei, G. Xiaorong, and G. Shangkai, “One-versus-the-rest(OVR) algorithm: An extension of common spatial patterns(CSP) algorithm to multi-class case,” in *Proc. 27th Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2005, pp. 2387–2390.

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

Mahnaz Arvaneh (S'11) received the B.Sc. degree in electrical engineering from K. N. Toosi University, Tehran, Iran, in 2005 and the M.Sc. degree in control engineering from the Ferdowsi University of Mashhad, Mashhad, Iran, in 2007. She is currently working toward the Ph.D. degree at Nanyang Technological University, Singapore.

From 2008 to 2009, she was with the R&D department of Mega-motor Company, Tehran. Her research interest includes brain–computer interface, signal processing, machine learning, and pattern recognition.

Cuntai Guan (SM'03) received the Ph.D. degree in electrical and electronic engineering from Southeast University, Jiangsu, China, in 1993.

He is currently a Principal Scientist and Program Manager at the Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), Singapore. From 1993 to 1994, he was at the Southeast University, where he focused on speech vocoder, speech recognition, and text-to-speech. In 1995, he was a Visiting Scientist at the CRIN/CNRS INRIA, Nancy, France, where he was focusing on key word spotting. From 1996 to 1997, he was with the City University of Hong Kong, where he was developing robust speech recognition under noisy environment. From 1997 to 1999, he was with the Kent Ridge Digital Laboratories, Connexis, Singapore, where he was focusing on multilingual, large vocabulary, continuous speech recognition. He was then a Research Manager and the R&D Director for five years in industries (L&H Asia Pacific, InfoTalk Technology), focusing on the research and development of spoken dialogue technologies. Since 2003, he established the Brain–computer Interface (BCI) Laboratory at Institute for Infocomm Research. He has published more than 110 refereed journal and conference papers, and holds 10 granted patents and applications. His current research interests include brain–computer interface, neural signal processing, machine learning, pattern classification, and statistical signal processing, with applications to assistive device, rehabilitation, and health monitoring.

Dr. Guan is a recipient several awards for his contributions in research and development to brain–computer interfaces, including Annual BCI Research Award, in 2010, IES Prestigious Engineering Achievement Award, in 2009, etc. He is now the Programme Manager of Intelligent Systems for Personalized and Connected Healthcare (INSPECH) programme, which covers neurotechnology, medical image processing, ambient intelligence, activity tracking, biosignal processing, health monitoring, diagnosis, rehabilitation and treatment, and decision support. He is an Associate Editor for IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING and Frontiers in Neuroprosthetics.

Kai Keng Ang received the B.A.S.c. (First Hons.), M.Phil. and Ph.D. degrees in computer engineering from Nanyang Technological University, Singapore, in 1997, 1999, and 2008, respectively.

From 1999 to 2003, he was a Senior Software Engineer with Delphi Automotive Systems Singapore Pvt. Ltd. focusing on embedded software for automotive engine controllers. He is now a Senior Research Fellow focusing on brain–computer interface in the Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), Singapore. His current research interests include computational intelligence, machine learning, pattern recognition, and signal processing.

Chai Quek received the B.Sc. degree in electrical and electronics engineering and the Ph.D. degree in intelligent control both from Heriot-Watt University, Edinburgh, U.K.

He is currently an Associate Professor and a Member of the Centre for Computational Intelligence (formerly the Intelligent Systems Laboratory) and the Assistant Chair with the School of Computer Engineering, Nanyang Technological University, Singapore. His current research interests include intelligent control, intelligent architectures, artificial intelligence in education, neural networks, fuzzy neural systems, neurocognitive informatics, and genetic algorithms.

Dr. Quek is a member of the IEEE Technical Committee on Computational Finance and Economics.