Sparse representation-based classification scheme for motor imagery-based brain–computer interface systems

2012 J. Neural Eng. 9 056002

(http://iopscience.iop.org/1741-2552/9/5/056002)

J. Neural Eng. **9** (2012) 056002 (12pp) doi:10.1088/1741-2560/9/5/056002

Sparse representation-based classification scheme for motor imagery-based brain–computer interface systems

Younghak Shin, Seungchan Lee, Junho Lee and Heung-No Lee

School of Information and Communications, Gwangju Institute of Science and Technology, Gwangju, Korea

E-mail: heungno@gist.ac.kr

Received 22 March 2012
Accepted for publication 10 July 2012
Published 7 August 2012
Online at stacks.iop.org/JNE/9/056002

Abstract
Motor imagery (MI)-based brain–computer interface systems (BCIs) normally use a powerful spatial filtering and classification method to maximize their performance. The common spatial pattern (CSP) algorithm is a widely used spatial filtering method for MI-based BCIs. In this work, we propose a new sparse representation-based classification (SRC) scheme for MI-based BCI applications. Sensorimotor rhythms are extracted from electroencephalograms and used for classification. The proposed SRC method utilizes the frequency band power and CSP algorithm to extract features for classification. We analyzed the performance of the new method using experimental datasets. The results showed that the SRC scheme provides highly accurate classification results, which were better than those obtained using the well-known linear discriminant analysis classification method. The enhancement of the proposed method in terms of the classification accuracy was verified using cross-validation and a statistical paired t-test ($p < 0.001$).

(Some figures may appear in colour only in the online journal)

1. Introduction

Brain–computer interface systems (BCIs) can provide a new communication and control channel between the user's brain and an external device, such as a computer or prosthetic device. Electroencephalogram (EEG) signals noninvasively capture the brain waves generated by brain activity. These can be recorded using a set of sensitive electrodes placed on the scalp. These signals can then be mapped to different commands after a series of sophisticated signal processing procedures, such as feature extraction and the classification of the measured EEG signals. Thus, a subject can control a computer application by intentionally generating different EEG patterns. The use of BCIs can be an additional means to the brain's normal output pathways, such as the peripheral nerves and muscles of the body [1].

EEG-based BCIs widely use two methods to obtain EEG signals: P300 [2, 3] and the motor imagery (MI)-based BCI [4–7]. In P300, distinctive EEG signals are usually obtained from the parietal and occipital areas after 300 ms of a visual stimulus when the subject is paying attention to a specific target. A possible visual stimulus might involve staring at the blinking rows and columns in a matrix of letters [3]. In this study, we focused on MI-based BCIs, which use sensorimotor rhythms (SMRs), such as the Mu and/or Beta rhythms; these rhythms can be recorded on the scalp over the sensorimotor cortex area. A widely used feature in MI-based BCIs is event-related desynchronization (ERD) [4]. A significant decrease in the power level of SMRs can be observed on the contralateral hemisphere during the unilateral imagination of hand movements [6]. The different patterns present in EEG signals are detected and used for BCI control.

Sparse representation-based classification (SRC) has received much attention recently in the pattern recognition field [8–10]. The sparse representation problem involves finding the most compact representation of a given signal, where the representation is expressed as a linear combination of columns in an overcomplete dictionary matrix. The idea of

sparse representation has been used in compressive sensing (CS) theory [11–13]. CS theory suggests that many natural signals can be represented as sparse signals on a certain basis. A sparsely representable signal can be compressively sampled, and the number of compressed samples required for the perfect recovery of the original signal is smaller than that required under the conventional Shannon–Nyquist sampling theory.

Sparse representation can be used for signal classification when compact description of a test signal is produced using a set of training signals incorporated in the dictionary [8]. This SRC method has been widely used, and its high classification accuracy has been demonstrated in various applications, such as face recognition [9] and speech recognition [10].

The EEG signals acquired from scalp electrodes are usually very noisy and show a non-stationary characteristic. They contain signals from non-interesting physiological activities (e.g. electromyograms (EMGs)), the sensor noise present in any electrical system and environmental noise (e.g. power lines). Thus, the use of powerful signal processing and classification techniques plays a critical role. In addition, raw EEG scalp potentials have poor spatial resolution because of the volume conduction effect [14]. In MI-based BCIs, an electrode placed on the scalp measures the EEG signals generated not only from the motor cortex area but also from other cortical regions. Thus, it is important to isolate the MI signals from the others. Spatial filtering is a technique that is used to extract localized information from a set of measured EEG signals derived from multiple electrodes. MI-based BCIs use SMR features that are generated directly from the sensorimotor cortex, which means that the use of spatial filtering is essential [17, 18]. Well-known spatial filtering methods include the Laplacian, common average reference [15] and common spatial pattern (CSP) method [16, 17].

In this work, we apply a SRC method to EEG-based MI BCIs. To the best of our knowledge, the proposed method is a novel classification method for this application. We use a band power approach that involves extracting the power information from the signal for the SMRs [18]. In SRC, the design of a good dictionary matrix is critical, or the performance will be poor. We provide a detailed design procedure for constructing the dictionary matrix. To maximize the benefit of sparse representation, we propose to use CSP filtering and preprocess the raw training signals to construct the columns of the dictionary. This was a critical step for increasing the classification accuracy with this new approach.

We note that there are some related works [31–33] in the BCI application. The authors in [31], for example, use sparse representation as a source separation method at the preprocessing stage to enhance the efficiency of feature extraction. The authors in [32], on the other hand, aim at enforcing a sparsity condition on the selection of spatial filter coefficients, by adding a sparsity-regularization term in the CSP optimization problem. They have shown that this sparse spatial filtering enhances classification accuracy to a certain degree. In [33], the authors aim at detecting the vigilance state of a subject, and use an L1-based sparse representation of EEG signals. Compared to these works, our work here provides a principled approach to design an efficient sparse representation dictionary. Furthermore, detailed discussion on why and how the SRC method works better than the conventional linear discriminant analysis (LDA) method is provided in our paper.

The remainder of this paper is organized as follows. In section 2, the EEG datasets and experimental procedure are described. In section 3, we first outline the preprocessing and CSP filtering procedures and then introduce the proposed methods. Section 4 presents the experimental results. There are discussions and conclusions at the end of this paper.

2. Experimental procedure and dataset

In this section, we describe the information in the datasets used for evaluating the performance of the proposed method. We also discuss the BCI experimental procedure for each dataset.

2.1. EEG datasets

In this study, we used two different datasets. The first was the INFONET dataset obtained from our own MI-based BCI experiment. The other was the Berlin dataset downloaded from the website of BCI competition III (dataset IVa) [19]. The main difference between the two datasets was that they used different numbers of EEG channels and had different sizes (i.e. numbers of total trials). The Berlin dataset contained more trials, i.e. 80 for the INFONET dataset and 140 for the Berlin dataset per class. The number of EEG electrodes used to collect data was also different, i.e. 16 for the INFONET dataset and 118 for the Berlin dataset. Testing using these two different datasets, one of which is widely known, facilitated objective comparison between the different classifiers.

2.1.1. INFONET dataset. This dataset consisted of five different datasets obtained from five healthy subjects (five males, average age = 22 and SD = 6.85). They were all novice subjects in BCI experiments. There were two classes, i.e. the left- and right-hand motor imaginary movements. In this experiment, we used 16 EEG channels. The EEG signals were recorded from active electrodes in a cap (with the earlobe used as the reference) based on the international 10/20 standard. In our experiment, we used a g.EEGcap and g.ACTIVE electrodes made by G. Tec Inc. and a PZ3 amplifier from Tucker-Davis Technologies. We used a sampling rate of 256 samples s^{-1} with a band-pass filter of 1–100 Hz and a notch filter of 60 Hz. Figure 1 shows the electrode positions used for our dataset.

In our BCI experiments, the subjects were seated in a comfortable chair and asked to watch a monitor screen. Figure 2 shows the time procedure for one trial. At the beginning of each run, a 'Left Hand' or 'Right Hand' letter instruction randomly appeared for 4 s at the center of the screen. Then, subjects imagined a left- or right-hand movement after the instruction was given, i.e. repeated fist clenching. This was followed by a rest period of 3 s. One run consisted of 40 trials, i.e. 20 left- and 20 right-hand trials. With all subjects, we conducted six runs that consisted of two runs with real movements and four runs with imaginary movements. We used

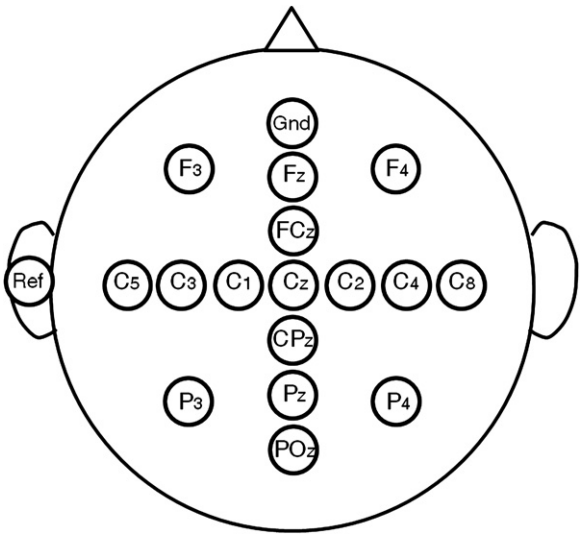

Figure 1. 16-channel EEG electrode positions used for INFONET dataset.

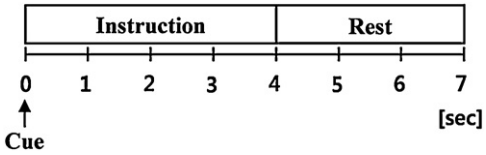

Figure 2. One trial time procedure for INFONET dataset.

only the imaginary data trials for further signal processing. To suppress electrooculogram (EOG) artifacts, the subjects were instructed not to blink or move their eyes during the instruction period. During the rest period, they could blink freely but were not allowed to move their body.

2.1.2. Berlin dataset. The Berlin dataset was produced in the BCI competition and is widely used in the BCI field for the analysis of EEG signal processing. It contains five datasets recorded from five different healthy subjects (aa, al, av, aw, and ay). The subjects followed the same procedure as the BCI experiment with three classes, i.e. left-hand, right-hand and right-foot MI movements. However, only the data corresponding to the right hand (R) and right foot (F) were used for analysis. These datasets only contain data from the four initial sessions without feedback. The data were recorded using BrainAmp amplifiers and a 128-channel Ag/AgCl electrode cap from ECI. 118 EEG channels were measured at the positions of the extended international 10/20 system. The exact electrode positions are provided in the data file [19]. The signals were band-pass filtered between 0.05 and 200 Hz and then digitized at 1000 Hz. The signals were downsampled to 100 Hz for offline analysis by the Berlin research team.

Figure 3 shows a timed trial procedure for the Berlin dataset. The subjects were seated in a comfortable chair with their arms resting on armrests. Visual cues were provided for 3.5 s that indicated the appropriate motor imagery the subject should perform, i.e. left hand, right hand, or right foot. The presentation of target cues alternated with periods of random length, i.e. 1.75–2.25 s, in which the subject could relax. Each of the five datasets consists of 140 trials for each class.

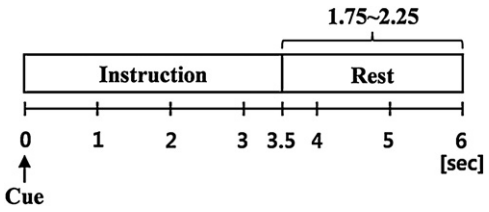

Figure 3. One trial time procedure for Berlin dataset.

3. Methods

In this study, we used CSP filtering to design the dictionary matrix of the SRC method. CSP filtering is used to distinguish between the features of two different classes of signals. In our method, the CSP filtering was used to preprocess the two classes of EEG training signals and the columns of the dictionary matrix were made of the CSP filtered signals. Thus the generated dictionary matrix became *incoherent*, i.e. columns from different classes are uncorrelated (details will be given briefly later in section 3.2.2). This was a crucial step in our proposed method. When a dictionary is *incoherent*, a test signal from one particular class can be predominantly represented by the columns of the same class. The *uncertainty principle* (UP) [29] in the sparse representation theory dictates that a signal cannot be sparsely represented in both classes simultaneously. This phenomenon intensifies as the degree of incoherence of the dictionary increases. We used this UP as the underlying principle in our design.

In this section, we explain the proposed method. Figure 4 shows the entire procedure, including the CSP filtering and classification steps. With the goal of providing a detailed explanation of how CSP filtering is incorporated into the SRC method, the principle of CSP filtering is briefly reviewed. We also outline the conventional CSP filtering-based LDA classification method to allow a comparison with the proposed method.

3.1. Preprocessing and CSP filtering

After the EEG data acquisition, we performed data segmentation for further analysis. We used 1–2 s time samples after the cue appeared in all of the experimental trials. To reduce the interference from other sources such as EOGs and EMGs, we used a band-pass filter with a passband of 8–15 Hz. We applied CSP filtering. CSP is a widely used and powerful signal processing technique that is suitable for two classes (conditions) of multi-channel EEG-based BCIs [17, 20]. To distinguish features, CSP filters maximize the variance of the spatially filtered signals for one class while minimizing it for the other class.

Here, we explain how to apply CSP filtering to the INFONET dataset. Let $\mathbf{X} \in \mathbb{R}^{C\times T}$ be a segment of EEG signals, where C is the number of EEG channels. In the INFONET dataset, C is 16 and T is the number of sampled time points collected in all of the trials for a single subject. In this study, we used 1 s data samples (256 samples). We had two classes of EEG training trials, each collected in a matrix, $\mathbf{X}_R \in \mathbb{R}^{C\times T}$ and $\mathbf{X}_L \in \mathbb{R}^{C\times T}$, corresponding to the

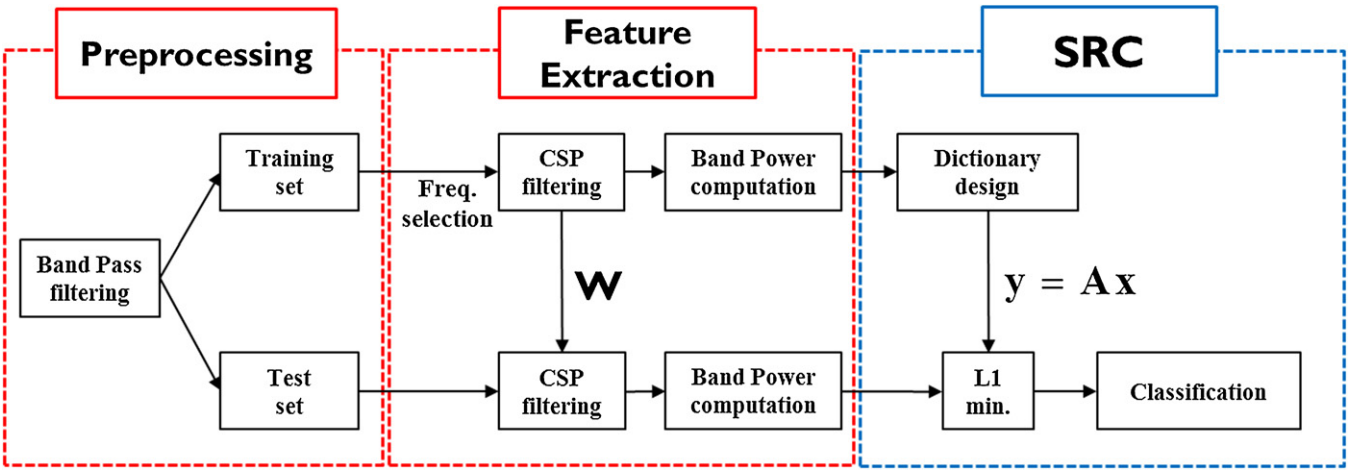

Figure 4. Proposed SRC scheme. We designed dictionary **A** using CSP filter **W** and the band power. To find coefficient vector **x**, we used the L1 minimization tool for test signal **y**.

right- and left-hand motor imaginary movements, respectively. Using the optimization procedure given below, we estimated the set of CSP filters stored in the filter matrix, $\mathbf{W} \in \mathbb{R}^{C\times C}$ [17]. Thus, we initially found a vector, $\mathbf{w}$, satisfying the following optimization problem:

$$\max_{\mathbf{w}} \left(\frac{\mathbf{w}^T C_R \mathbf{w}}{\mathbf{w}^T C_L \mathbf{w}} \right). \tag{1}$$

Equation (1) leads to the minimization problem given below:

$$\min_{\mathbf{w}} \left(-\mathbf{w}^T C_R \mathbf{w} \right) \quad \text{subject to } \mathbf{w}^T C_L \mathbf{w} = 1, \tag{2}$$

where we define $C_R = X_R X_R{}^T$ and $C_L = X_L X_L{}^T$. Using the Lagrangian method, we derive the following equation from (2):

$$L(\mathbf{w}, \lambda) = -\mathbf{w}^T C_R \mathbf{w} + \lambda(\mathbf{w}^T C_L \mathbf{w} - 1). \tag{3}$$

Taking the derivative of (3) and setting it equal to zero, $\frac{\partial}{\partial \mathbf{w}} L(\mathbf{w}, \lambda) = -C_R \mathbf{w} + \lambda C_L \mathbf{w} = 0$, we have

$$C_R \mathbf{w} = \lambda C_L \mathbf{w}. \tag{4}$$

Thus, the optimization problem in (1) can be solved as a generalized eigenvalue problem in (4). Using (4), we derive the objective function (1) as follows:

$$\max_{\mathbf{w}} \left(\frac{\mathbf{w}^T C_R \mathbf{w}}{\mathbf{w}^T C_L \mathbf{w}} \right) = \max_{\mathbf{w}} \left(\frac{\mathbf{w}^T \lambda C_L \mathbf{w}}{\mathbf{w}^T C_L \mathbf{w}} \right) = \max_{\mathbf{w}} \lambda. \tag{5}$$

Finally, $\mathbf{w}$ is an eigenvector corresponding to the largest eigenvalue. The eigenvalues of (4) are the roots of the characteristic equation, i.e. $|C_R - \lambda C_L|_{\text{det}} = 0$, and the eigenvector for each λ is the corresponding $\mathbf{w}$, obtained by solving $(C_R - \lambda C_L)\,\mathbf{w} = 0$. $\mathbf{w}$ maximizes the variance for the signals of the right-hand class, while at the same time minimizing it for the signals of the left-hand class. Note that this property produces the maximum incoherence between the two classes. We will discuss this in section 3.2.2. We used 16 EEG channels in the INFONET dataset. Thus, we had 16 eigenvalues and eigenvectors. The CSP filter matrix, $\mathbf{W}$, consists of column vectors, $\mathbf{w}_i \in \mathbb{R}^C (i = 1, 2, \ldots, C)$ where $C = 16$. The second and later components are derived from the rest of the eigenvectors. Namely, we obtain all of the other eigenvectors, $\mathbf{w}_i$, $i = 2, \ldots, C$, corresponding to the second largest to the smallest eigenvalue. The eigenvalues are indexed from 1 to C in decreasing order. We call each column vector $\mathbf{w}_i$ of $\mathbf{W}$ a spatial filter. Of these, we used $2n$ CSP filters, $n \leqslant 8$ for $C = 16$. The first n columns and the last n columns of $\mathbf{W}$ were selected as the $2n$ CSP filters. The filter $\mathbf{w}_{16}$, corresponding to the smallest eigenvalue, gave exactly the opposite effect to the first filter, $\mathbf{w}_1$. Thus, it minimized the variance of the data for the right-hand class while maximizing that of the left-hand class. Finally, we produced the CSP filtering matrix, $\mathbf{W}_{\text{CSP}} \in \mathbb{R}^{C\times 2n}$, as follows:

$$\mathbf{W}_{\text{CSP}} = [\mathbf{w}_1, \ldots, \mathbf{w}_n, \mathbf{w}_{C-n+1}, \mathbf{w}_C]. \tag{6}$$

Selecting the best number of filters, n, depended on the number of training trials and the subject. Given two classes of EEG training trials, $\mathbf{X}_R \in \mathbb{R}^{C\times T}$ and $\mathbf{X}_L \in \mathbb{R}^{C\times T}$, we defined the CSP filtered signals as follows:

$$\begin{aligned} \mathbf{X}_R^{\text{CSP}} &\in \mathbb{R}^{2n\times T} := \mathbf{W}_{\text{CSP}}^{\text{T}} \mathbf{X}_R, \\ \mathbf{X}_L^{\text{CSP}} &\in \mathbb{R}^{2n\times T} := \mathbf{W}_{\text{CSP}}^{\text{T}} \mathbf{X}_L. \end{aligned} \tag{7}$$

Figure 5 shows eight CSP filters that correspond to the four largest and four smallest eigenvalues of subject A from the INFONET dataset. These filters were used to project the original EEG signal matrix in (7) into the $2n \times T$ space.

The color in figure 5 denotes the significance of the corresponding channels. For example, the red color indicates the highest significance. Let us take the upper-left picture as an example, which is for the first filter $\mathbf{w}_1$. It is noticeable that there is a strong focus, a red dot, on the position of the C4 electrode in this case. Imaginary movement of a hand causes an ERD feature in the contralateral hemisphere. Namely, with a left-hand imaginary movement, a signal feature appears on the C4 electrode, while with a right-hand imaginary, it appears on the C3 electrode. Hand motor functions are controlled in a motor cortex region of the brain on which the C3 and C4 electrodes are placed. This contralateral manifestation of imaginary hand movements is a well-known neurophysiological phenomenon [4, 5]. From this discussion, it is clear that the first CSP filter amplifies the signal feature from the left-hand imaginary movement at the C4 electrode,

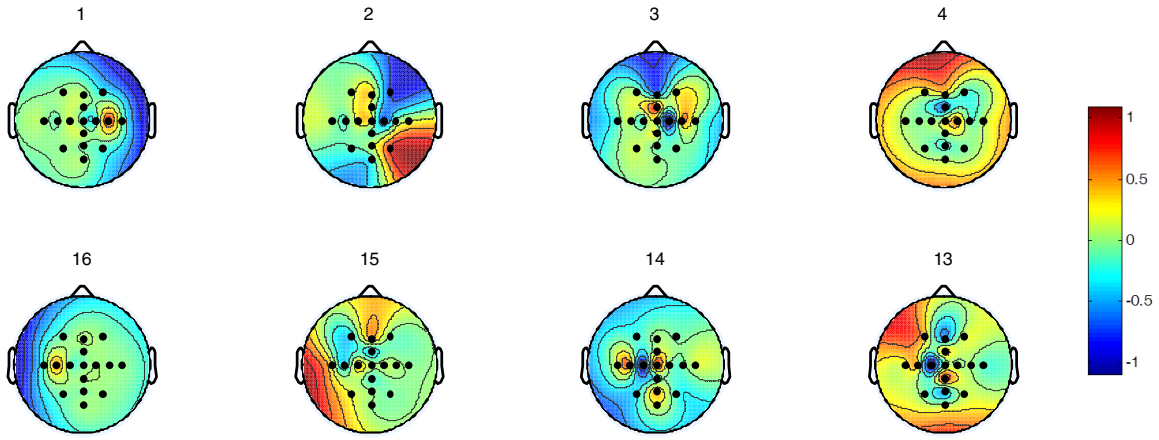

Figure 5. Color coded mapping of magnitudes of the coefficients of a filter, say $\mathbf{w}_i$ whose index i is indicated at the top of each figure. CSP filters are computed from the right-hand and the left-hand imaginary movement signals of subject A in INFONET dataset. The color map is projected onto the scalp.

while the last CSP filter does the one from the right-hand imaginary movement.

3.2. Proposed SRC method

In this section, we describe the design of the dictionary matrix using CSP filtering and the frequency band power of SMRs. We also give a detailed description of how our proposed linear sparse representation model performed as a classification tool for BCI application.

3.2.1. Feature extraction and dictionary design. An important physiological issue in the MI-based BCI field is that the SMRs include the Mu (8–14 Hz) and Beta (14–24 Hz) frequency bands, which are closely related to limb movements. The power of specific frequency bands (so-called band power) is a widely used feature in Graz BCI systems [4–7]. In this study, the power of the CSP filtered signal, i.e. the second moment of each row of $\mathbf{X}_R^{\mathrm{CSP}}$ and $\mathbf{X}_L^{\mathrm{CSP}}$, is the band power from 8 to 15 Hz. We fixed the frequency range for all of the subjects at 8 to 15 Hz to simplify the comparison. This frequency band was not optimally chosen and may depend on the subject.

An important step when applying the SRC method to the BCIs was the design of an appropriate dictionary matrix, $\mathbf{A}$. Figure 6 shows how this was performed. Let N_t be the total number of training signals for each class, i. That is, $i = L$ for the left-hand, $i = R$ for the right-hand. We define a component dictionary matrix $\mathbf{A}_i = [\mathbf{a}_{i,1}, \mathbf{a}_{i,2}, \ldots, \mathbf{a}_{i,N_t}]$ for each class i where each column vector $\mathbf{a}_{i,j} \in \mathbb{R}^{m\times 1}$, $j = 1, 2, \ldots, N_t$, having dimension $m = 2n$ is obtained by concatenating the number of $2n$ SMR band power features, i.e. $2n$ sum of frequency power from 8 to 15 Hz. Here, $2n$ was the number of CSP filters. The same procedure was repeated for the left-hand and right-hand classes. By concatenating the two matrices, we formed the complete dictionary, $\mathbf{A}: = [\mathbf{A}_L; \mathbf{A}_R]$ as shown in figure 6, where the dimension was $m \times 2N_t$.

3.2.2. Incoherent dictionary using CSP filter. In this subsection, we discuss why the CSP filtering method is a good technique to use for the design of the dictionary matrix. We also demonstrate how CSP filtering can be used to maximize the incoherence between the two classes in the dictionary.

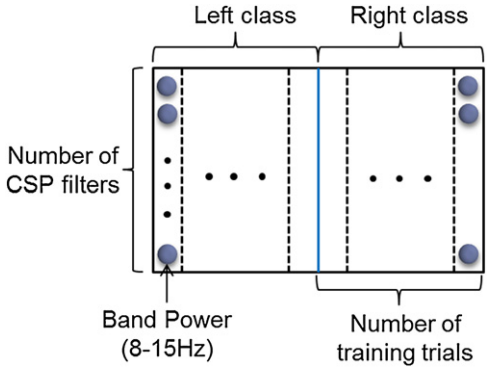

Figure 6. Dictionary design for the proposed SRC method. Each element of the dictionary is the band power from 8 to 15 Hz. Each column is obtained from a training trial for a separate class. The number of rows in the dictionary represents the number of CSP filters (i.e. feature dimensions).

The coherence measures the correlation between the two component dictionaries defined in the following way:

$$M(\mathbf{A}_L, \mathbf{A}_R) \triangleq \max\{ \mid \langle \mathbf{a}_{L,j}, \mathbf{a}_{R,k} \rangle \mid : j, k = 1, 2, \ldots, N_t\}. \quad (8)$$

The vector $\mathbf{a}_{L,j}$ is the jth column of $\mathbf{A}_L$; similarly, $\mathbf{a}_{R,k}$ is the kth column of $\mathbf{A}_R$. The notation $\langle \mathbf{a}_{L,j}, \mathbf{a}_{R,k} \rangle$ denotes the inner product of two vectors. We call M the measure of mutual coherence of the two component dictionaries; when M is small, we say that the complete dictionary is *incoherent* [29]. As we have mentioned already in the first paragraph of this section, an *incoherent* dictionary promotes the sparse representation of the test signal under the L1 minimization (see theorem 1 of [30]). As will be discussed in section 5, being able to sparsely represent a test signal in turn helps in boosting the classification accuracy of the proposed method.

Recall that we use the CSP filtering method. The CSP filter maximizes the variance of the spatially filtered signals for one class, while minimizing it for the other class. Figure 7 shows a two-dimensional example illustrating the effect of

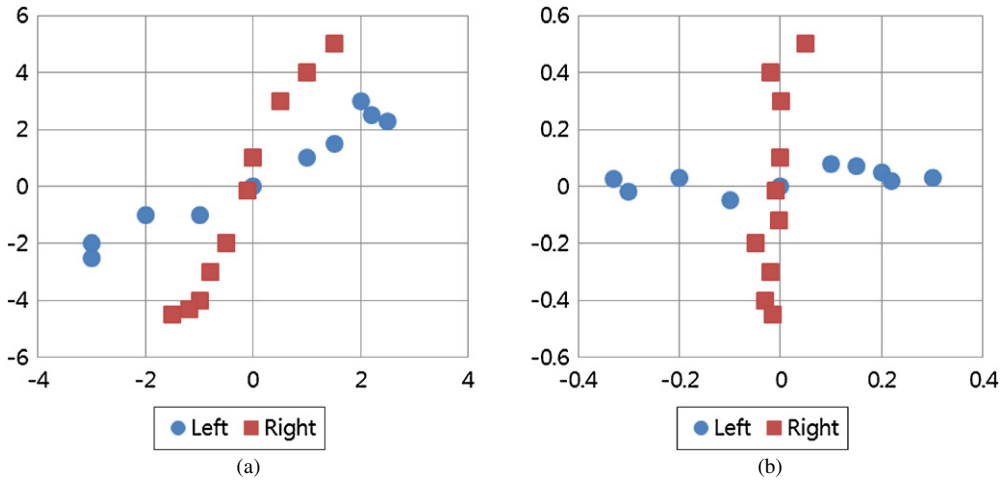

Figure 7. Example of CSP filtering effect: (a) before CSP filtering; (b) after CSP filtering.

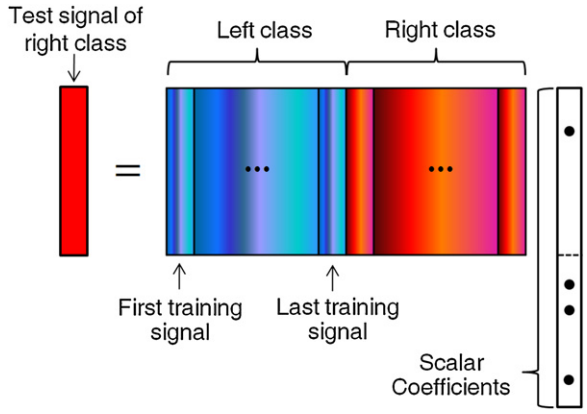

Figure 8. Linear sparse representation model.

CSP filtering and its relation to incoherence. Two classes of samples are expressed by the blue circles and red squares. Figure 7(a) shows the distribution of the samples before CSP filtering. Figure 7(b) shows the distribution of the samples after CSP filtering. In (b), the horizontal axis is $\mathbf{w}_1$, which is an eigenvector corresponding to the largest eigenvalue. The vertical axis is $\mathbf{w}_{16}$, which is an eigenvector corresponding to the smallest eigenvalue. This $\mathbf{w}_1$ has the property that after the samples are projected onto $\mathbf{w}_1$, the variance of the projected samples for the left class (blue circles) is maximized, while the variance of the projected samples for the right class (red squares) is simultaneously minimized. In addition, $\mathbf{w}_{16}$ does exactly the reverse of $\mathbf{w}_1$. Using the effect of the CSP filter, we simultaneously form maximally uncorrelated feature vectors between the two classes (see also figure 6 in [17]). Thus, if we calculate and compare the mutual coherence, M, between the two classes, before and after applying the CSP filtering, surely the mutual coherence after the filtering (figure 7(b)) is smaller.

3.2.3. Linear sparse representation model. In this section, we introduce our sparse representation model based on EEG training and test signals. We also demonstrate how this model works using an example. Figure 8 shows the proposed linear sparse representation model. First, we obtained the test signal, $\mathbf{y}$, using the same procedure used to obtain the columns of dictionary $\mathbf{A}$. Thus, a test signal was transformed into a vector, $\mathbf{y} \in \mathbb{R}^{m\times 1}$, via the processes of CSP filtering and band power computation. Thus, the dimension of $\mathbf{y}$ was the same as the dimension of the columns in dictionary $\mathbf{A}$. This test signal, $\mathbf{y}$, can be sparsely represented as a linear combination of some columns from $\mathbf{A}$:

$$\mathbf{y} = \sum_{i=L,R} x_{i,1}\mathbf{a}_{i,1} + x_{i,2}\mathbf{a}_{i,2} + \cdots + x_{i,n_t}\mathbf{a}_{i,N_t}, \tag{9}$$

where $x_{i,j} \in \mathbb{R}$, $j = 1, 2, \ldots, N_t$ are scalar coefficients. We represent this using a matrix algebraic form:

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \tag{10}$$

where $\mathbf{x} = [x_{L,1}, x_{L,2}, \ldots, x_{L,N_t} x_{R,1}, x_{R,2}, \ldots, x_{R,N_t}]^T \in \mathbb{R}^{2\times N_t}$. In this study, we assume a no-noise model for our sparse representation in equation (10). In figure 8, the test signal $\mathbf{y}$ of class R can be represented as the linear combination of the training signals of class R.

$$\mathbf{y}_R = \mathbf{A}\mathbf{x}_R \in \mathbb{R}^{m\times 1}, \tag{11}$$

where $\mathbf{x}_R = [0, \ldots, 0, \mathbf{a}_{R,1}, \mathbf{a}_{R,2}, \ldots, \mathbf{a}_{R,N_t}]^T \in \mathbb{R}^{2N_t}$ is a coefficient vector whose elements are zero, except for some elements that are associated with the test signals of class R. The sparse representation of test signal $\mathbf{y}$ is produced when the number of non-zero coefficients of $\mathbf{x}$ is much smaller than N_t [11]. It is useful to include as many training signals as possible.

3.2.4. Sparse representation by L1 minimization. Our sparse representation model in figure 8 shows that the number of total training signals was $2N_t$, which was much larger than the number of CSP filters ($m = 2n$). Thus, the linear equation (11) is under-determined ($m < 2N_t$). This problem can be solved using L0 norm minimization:

$$\min \|\mathbf{x}\|_0 \text{ subject to } \mathbf{y} = \mathbf{A}\mathbf{x}. \tag{12}$$

The L0 norm is equivalent to the number of non-zero components in vector $\mathbf{x}$, by definition. This involves a combinatorial search. Therefore, solving this L0 norm minimization problem is an NP-hard problem. However, recent studies in CS theory have shown that signal reconstruction

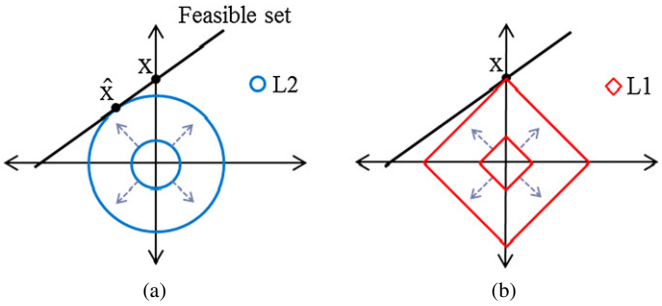

Figure 9. L1 and L2 norm minimization.

is conducted using an L1 minimization technique [11–13]. This is a suboptimal and relaxed approach to the optimal but intractable L0 minimization. The L1 norm minimization, given below, as one of the big surprises in CS theory, finds the exact sparse solution in polynomial time if sufficient samples are available [21],

$$\min \|\mathbf{x}\|_1 \text{ subject to } \mathbf{y} = \mathbf{A}\mathbf{x}, \tag{13}$$

unlike the conventional L2 norm minimization given below:

$$\min \|\mathbf{x}\|_2 \text{ subject to } \mathbf{y} = \mathbf{A}\mathbf{x}. \tag{14}$$

Figure 9 shows a two-dimensional example of why the L1 norm minimization finds the sparse solution efficiently, unlike the L2 norm minimization [12]. In figure 9, the black line represents the set of all feasible solutions. From the definition of the norm, the L2 and L1 norms can be individually represented as vectors on the surface of the circle and rhombus in (a) and (b), respectively. Using the L2 norm minimization in (14), when the L2 ball (circle) is grown in an equidistant manner, we can find the minimum L2 ball, which touches the feasible set first. As shown in (a), the L2 ball finds the non-sparse point, $\hat{\mathbf{x}}$, which lies in the two-dimensional non-zero space. Similarly, in (b), the L1 ball finds the sparse point, $\mathbf{x}$, which lies on the vertical axis.

There are many L1 minimization algorithms. In this study, we used a standard linear programming method called basis pursuit [22]. The 'SolveBP' function implements the basis pursuit method available in SparseLab, which is a free MATLAB software package [23]. This function solves equation (13) by reducing it to a linear program using an optimization technique such as the primal-dual log-barrier algorithm.

3.2.5. Sparse representation-based classification. In an ideal case, the L1 minimization solution of the linear equation (11), such as $\mathbf{x}_R$, should only have non-zero coefficients in the lower half, which corresponds to the columns of the right-hand class. However, the EEG signals are very noisy and non-stationary in MI-based BCIs. Thus, non-zero coefficients may appear in the indices that correspond to the columns of another left-hand class. To make use of the sparse representation results in classification, we need to specify a classification rule. One method is simply to count the number of non-zero coefficients in vector $\mathbf{x}$. Another method is to compute the energy of the coefficients, i.e.

$$\text{Class}(\mathbf{y}) = \arg\max_{i=L,R} \|\mathbf{x}_i\|_2. \tag{15}$$

An additional effective classification rule is to use the residuals, which was introduced in [9]. We used this method as the classification rule in this study. To explain this rule, it is useful to introduce a utility function. For each class i, we define its characteristic function $\delta_i(\cdot)\mathbb{R}^{2N_t} \rightarrow \mathbb{R}^{2N_t}$, which picks up only the coefficients associated with class i, while nullifying the coefficients of the other class. Thus, for $\mathbf{x} \in \mathbb{R}^{2N_t}$, $\delta_R(\mathbf{x}) \in \mathbb{R}^{2N_t}$ is a vector that is obtained by nulling all of the elements of $\mathbf{x}$ that are associated with the left-hand class. We can obtain the residual norm for the right-hand class:

$$r_R(\mathbf{y}) := \|\mathbf{y} - \mathbf{A}\delta_R(\mathbf{x})\|_2. \tag{16}$$

Similarly, the residual norm, $r_L(\mathbf{y})$, for the left class can be obtained for a given test vector, $\mathbf{y}$. The classification rule is given by

$$\text{Class}(\mathbf{y}) = \arg\min_i r_i(\mathbf{y}). \tag{17}$$

We determine the class, i, that has the minimum residual norm.

3.3. Linear discriminant analysis

To provide a fair comparison between the SRC and LDA (also known as Fisher's LDA) classification methods, we aim to explain how the LDA classification method works when CSP filtering is incorporated. LDA is widely used as a linear classification method in the BCI field [28], e.g. see [17] for MI-based and [27] for P300-based BCIs. The LDA approach introduced by Fisher aimed to find the optimal direction, $\mathbf{w}_L$, to project data upon and maximize Fisher's ratio [20, 25]:

$$J(\mathbf{w}_L) = \frac{\mathbf{w}_L{}^T \mathbf{S}_\mathbf{B} \mathbf{w}_L}{\mathbf{w}_L{}^T \mathbf{S}_\mathbf{w} \mathbf{w}_L}, \tag{18}$$

where $\mathbf{S}_\mathbf{B}$ and $\mathbf{S}_\mathbf{w}$ are called the *between-class* scatter matrix and *within-class* scatter matrix, respectively, which are obtained as follows:

$$\mathbf{S}_\mathbf{B} = (\mathbf{m}_2 - \mathbf{m}_1)(\mathbf{m}_2 - \mathbf{m}_1)^T$$

and

$$\mathbf{S}_\mathbf{w} = \sum_i (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T, \tag{19}$$

where $\mathbf{x}$ is the input feature vector and $\mathbf{m}_i$ is the group mean of the feature vectors in class i. We used the band power of the CSP filtered signals as a feature vector, $\mathbf{x}$, which was exactly the same feature used in the proposed SRC scheme (see section 3.2 and figure 6).

4. Results

In this section, we present the classification results with the proposed SRC method using the two datasets, INFONET and Berlin, as described in section 2. We also compare the results achieved with the SRC and the LDA methods. To evaluate the average classification accuracy using limited size datasets, we used the statistical leave-one-out (LOO) cross-validation method with the same total number of data trials for each subject [26].

Table 1 shows the classification accuracy for the INFONET dataset. There were a total of 160 trial signals

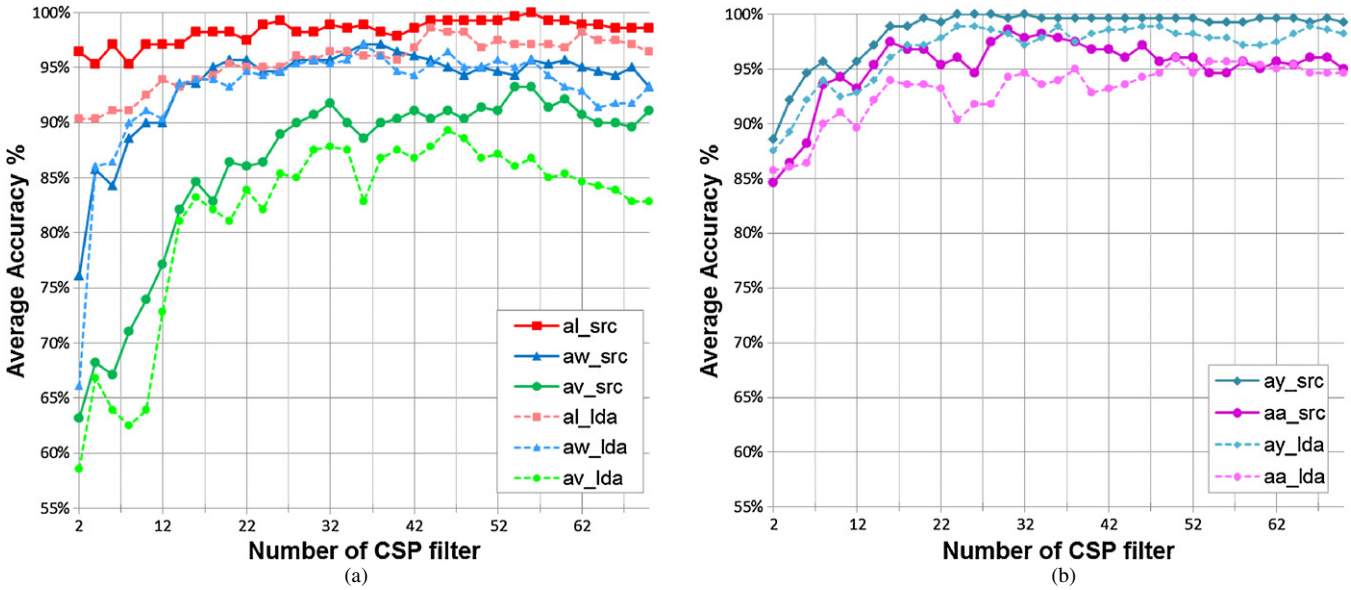

Figure 10. Classification accuracy (%) per subject with different number of CSP filters for Berlin dataset. (a) Classification accuracies for subjects al, aw and av. The solid lines represent SRC results and the dashed lines represent LDA results. (b) Classification accuracies for subjects ay and aa.

Table 1. Classification accuracy (correct recognition rate (%) = [correct number of assessments/total number of assessments (160)] × 100) with INFONET dataset for two different SRC and LDA classification methods.

Subject	SRC accuracy (%)	LDA accuracy (%)
A	95.63	93.13
B	63.75	61.87
C	68.14	67.50
D	80	76.25
E	71.25	68.12
Mean (SD)	75.75 (12.60)	73.37 (12.18)

Table 2. Comparison of the SRC and the LDA methods in terms of classification accuracy (%) with Berlin dataset.

Subject	SRC accuracy (%)	LDA accuracy (%)
al	98.93	96.43
ay	100	97.14
aw	95.71	95.36
aa	97.86	94.64
av	91.79	87.86
Mean (SD)	96.85 (3.25)	94.29 (3.72)

for each subject. In each assessment using the LOO cross-validation method, one trial signal out of the 160 trial signals was selected as the test signal, and the remainder (159 trial signals) were used as the training signals to produce the columns of the dictionary. This assessment is repeated for each trial signal. Thus, there were 160 assessments for each subject. We used the first and last CSP filters to produce the feature vectors and dictionary, i.e. $n = 1$ in (6), for subjects A, B and C, whereas we used four filters, $n = 2$, for subjects D and E. The number of CSP filters to use was determined empirically. We will return to this discussion in section 5.2. The results in table 1 show that the proposed SRC scheme delivered enhanced classification accuracy compared with the conventional LDA method for all of the subjects.

To further evaluate the SRC method, we extended our validation to the Berlin dataset (see section 2.1.1). This dataset was acquired from five subjects using 118 EEG channels. Table 2 shows the results of comparisons using this dataset. We used a total of 280 trial signals and the LOO method for all of the subjects. With this dataset, the number of available CSP filters was 118. We selected the number of CSP filters based on our experimental results in figure 10. Figure 10 shows the classification accuracy (%) of SRC and LDA as a function of the number of CSP filters for each subject. Figure 10(a) shows the results of subjects al, aw and av. The solid line represents the SRC accuracy and the dashed line represents the LDA accuracy. Figure 10(b) shows the results of subjects ay and aa. We compute the average accuracy of the 160 trials expressed in figure 10 using the LOO cross-validation method. As can be seen in these figures, there was no significant increase in accuracy when more than 32 CSP filters were used for both SRC and LDA methods. Thus, we used 32 CSP filters for feature extraction, which corresponded to the 16 largest and the 16 smallest eigenvalues. However, figure 10 shows that for each selection on the number of CSP filters, SRC performs better than LDA does with few exceptions. Thus, it can be said that SRC has better classification accuracy than LDA regardless of the number of CSP filters in figure 10. To investigate the statistical significance of the observed accuracies in figure 10, we performed a paired t-test for each subject. The obtained p-value of the t-test was less than 0.05 for all subjects, which indicates that the difference was significant.

Table 2 indicates that the proposed SRC scheme delivered higher average classification accuracy (96.85%) than the LDA method (94.29%). Moreover, for subject ay, the accuracy was 100%. Thus, the proposed SRC method had consistently higher classification accuracy than the LDA method in both datasets. We also performed a paired t-test on both datasets in tables 1

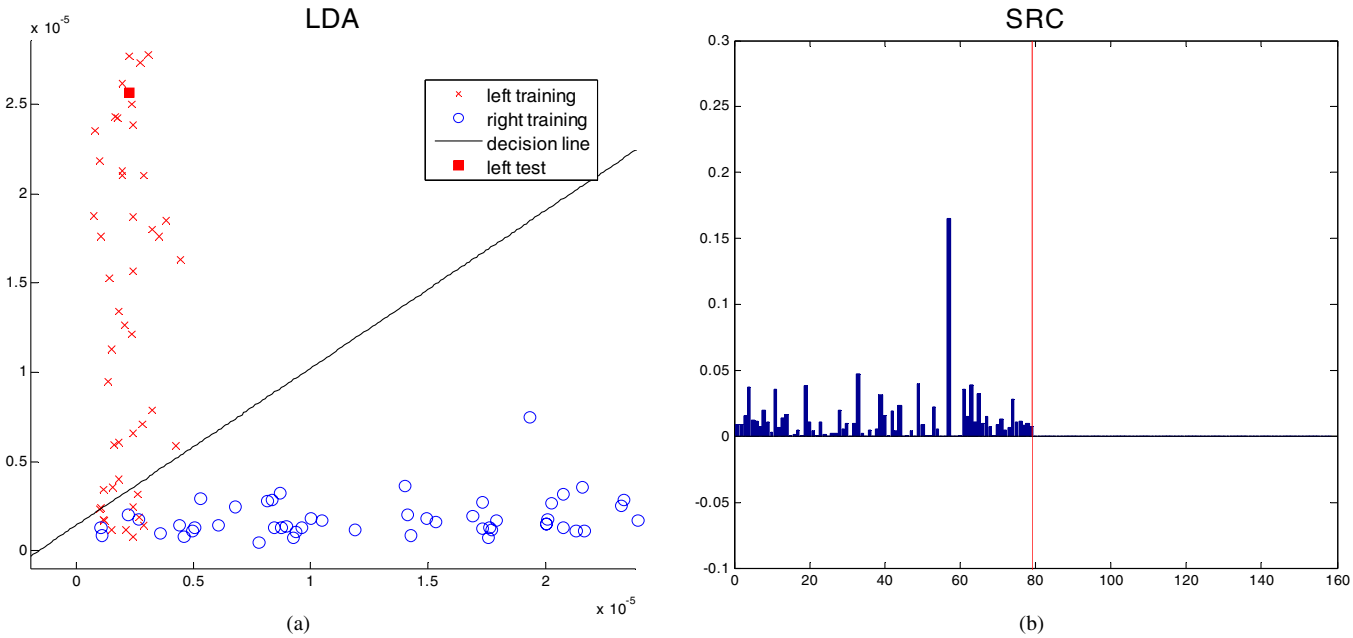

Figure 11. Example of LDA and SRC results when the test signal is easy to classify. (a) The two-dimensional scatter plot using two CSP filters ($n = 1$) of subject A from the INFONET dataset. Here, the x-axis represents the band power of the signal projected to the first CSP filter and the y-axis represents the last CSP filter. (b) Coefficient vector $\mathbf{x}$ recovered using the same test signal as in (a). The central vertical line in (b) indicates the boundary between the two classes. Here, the x- and y-axes represent the index and recovered coefficient of $\mathbf{x}$, respectively.

and 2. The obtained p-value was less than 0.001, which means that the difference was statistically significant.

An important issue for online BCI applications is the speed of the algorithm, as well as the classification accuracy. We compared the execution times of the algorithms. SRC and LDA took similar times to complete the classification job. The average difference between the execution times was negligible with the same computer and software (using MATLAB), i.e. LDA took 129.78 s and SRC took 129.99 s. This small difference is negligible.

5. Discussion

5.1. Why does SRC outperform LDA?

SRC is stronger than LDA as a classification tool. Both SRC and LDA work very well when the classification task is easy. However, SRC does better when the classification task is difficult such as in the 'gray' region. For easy exposition, we use two-dimensional signal sets (two CSP filters) obtained from subject A of INFONET dataset.

Let us consider an easy classification task first. Figure 11 shows that both LDA and SRC work well in this case. Locate the square shown in figure 11(a). It indicates a left class signal chosen as the test signal in this example. Note that it is located very far—thus an easy classification task—from the cloud of right class training signals, which are plotted as circles. The square is located above the decision line. Thus, the LDA method correctly classifies the test signal as a left class signal. We now consider the same test signal shown in (a) for the SRC method. Figure 11(b) shows the recovered coefficient vector, $\mathbf{x}$, in equation (13) using the L1 minimization method. SRC searches for a parsimonious representation of the test signal, the linear combination of a small number of training signals. The sparse representation is enforced by the L1 minimization routine. It is well known that L1 minimization is good at finding the sparsest representation among all feasible representations. In figure 11(b), the vertical line at the center indicates the boundary between the two classes. Thus, the part to the left of the boundary indicates the left class and the part to the right corresponds to the right class. Note that the coefficients belonging to the right class are all close to zero. In this case, the characteristic function $\delta_R(\mathbf{x})$ in (16) is close to zero as well. Thus, the residual norm for the right class in (16) is as large as the norm of the test signal. Moreover, it is easy to see that that of the left class should be somewhat smaller than this. Thus, the minimal residual norm classification criterion (17) decides that the test signal belongs to the left class.

Let us now consider a difficult case. Even though we use CSP filtering, for some trials, the physiological features are very similar to each other. Therefore, a 'gray' region exists where the two clouds of training signals are intermixed, and classification is difficult. This region is indicated by the dotted circle shown in figure 12(a). With LDA, a single decision line is drawn from the training signal set, but this line does not exactly divide the two classes of signals clearly, as indicated in (a). In (a), suppose the square inside the dotted circle is the test signal, a signal belonging to the left class. In this particular case, the test signal is located below the decision line, and LDA wrongly classifies it as the right class. This is an example of a misclassification by LDA.

On the other hand, figure 12(b) shows the result of the SRC method, the coefficient vector, $\mathbf{x}$, obtained from the L1 minimization routine. The test signal is the same as in (a). The

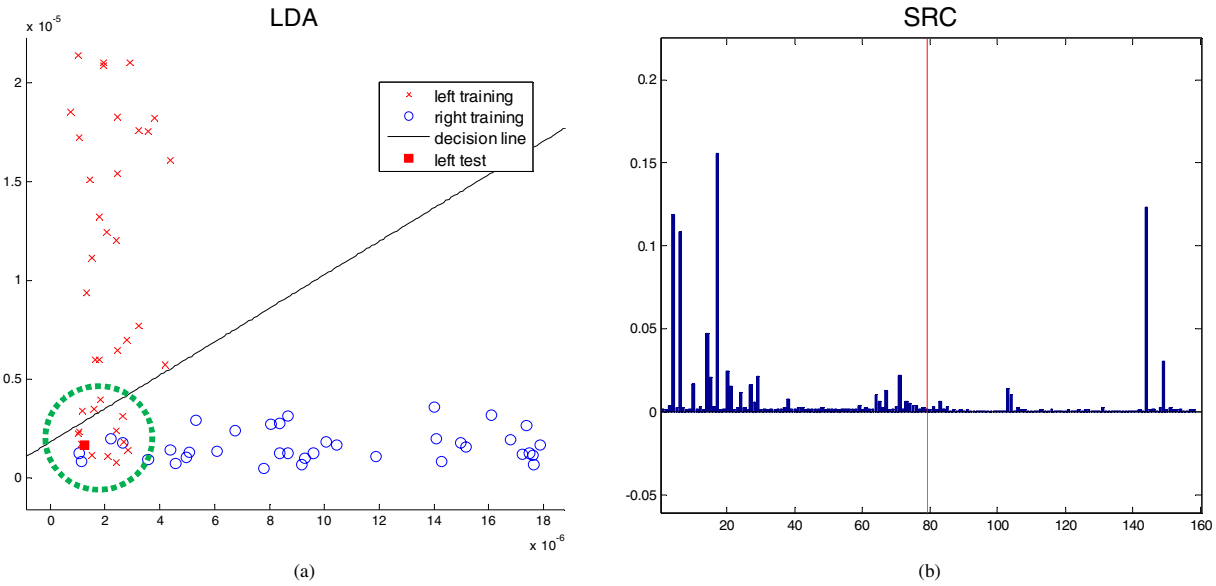

Figure 12. Example of LDA and SRC results when the test signal is difficult to classify. (a) A two-dimensional scatter plot using two CSP filters ($n = 1$) for subject A from the INFONET dataset. The dotted circle indicates the 'gray' region. (b) Coefficient vector **x** recovered using the same test signal as in (a).

test signal is located inside the dotted circle, which shows that the physiological features of that signal are difficult to classify. Note that the SRC method classifies it correctly as left class. This is what differentiates SRC from LDA.

The SRC method continues to work even in the 'gray' area where classification is difficult. LDA relies only on a single decision line. Once it is drawn, it is the only measure considered in classifying a test signal. In contrast, SRC aims to find a linear combination of a small number of training signals that can account for the given test signal with a sparse representation of the test signal. This sparse representation is enforced by the L1 minimization routine. It is known that L1 minimization using its key characteristic is good at finding the sparsest representation among many feasible representations. In addition, the dictionary matrix is *incoherently* designed. Therefore, there is a chance that the test signal will receive heavier representation with the correct class of training signals than with the wrong class. We mentioned that an *incoherent* dictionary has an intriguing relation with the UP in the sparse representation theory [29] at the start of section 3. Shedding light using UP on the class selecting property of L1 minimization results given in this subsection, we note the following: under L1 minimization a test signal does not yield a parsimonious representation with the training signals from both component dictionaries simultaneously, but does yield a partial representation from a selected component dictionary so as to bring forth as sparse an overall representation as possible.

The job of any classifier for test signals within the gray region is not easy. SRC is not an exception in this sense. It is also possible to find test signals for which LDA makes the correct decision while SRC does not. Figures 13(a) and (b) show such an example. Find the right test signal located in the gray region. It is under the decision line. Thus, LDA classifies it as right class, which is correct. Per the L1 minimization result of figure 13(b), however, SRC classifies it as left class, which is an error. Note that SRC still makes a parsimonious representation, even in this case where the decision is wrong.

On the one hand, SRC results for a test signal within the gray region are intermixed with successes and failures. In the case of LDA, on the other hand, all the left class test signals below the decision line lead to 100% classification failures.

5.2. Selecting number of CSP filters

In section 3.2, we used CSP filtering to produce the columns of the dictionary matrix, **A**. The number of rows in the dictionary was the same as the number of CSP filters used. It is known that using two or three CSP filters from both ends of the eigenvectors is usually appropriate [17]. However, there is no exact method for selecting the appropriate number of CSP filters. The literature suggests heuristics that determine the number of CSP filters based on experience and visual inspection [24]. One lesson we could glean from the literature is that the use of too many CSP filters could produce the problem of classifier overfitting. Namely, a classifier could become overfitted to the transient part of the training set and, in such a case, it might not work well with a test signal [17]. For example, with subject aa in the Berlin dataset, we used 118 EEG channels, and 140 training signals were available for each class. In this case, our experiment showed that the best classification accuracy was achieved with 32 ($n = 16$ in (6)) CSP filters.

In the meantime, it should be noted that there is a fundamental limitation on the size of the dictionary with L1 minimization [12]. The number of rows in the dictionary, i.e. the number of linear equations in the sparse representation,

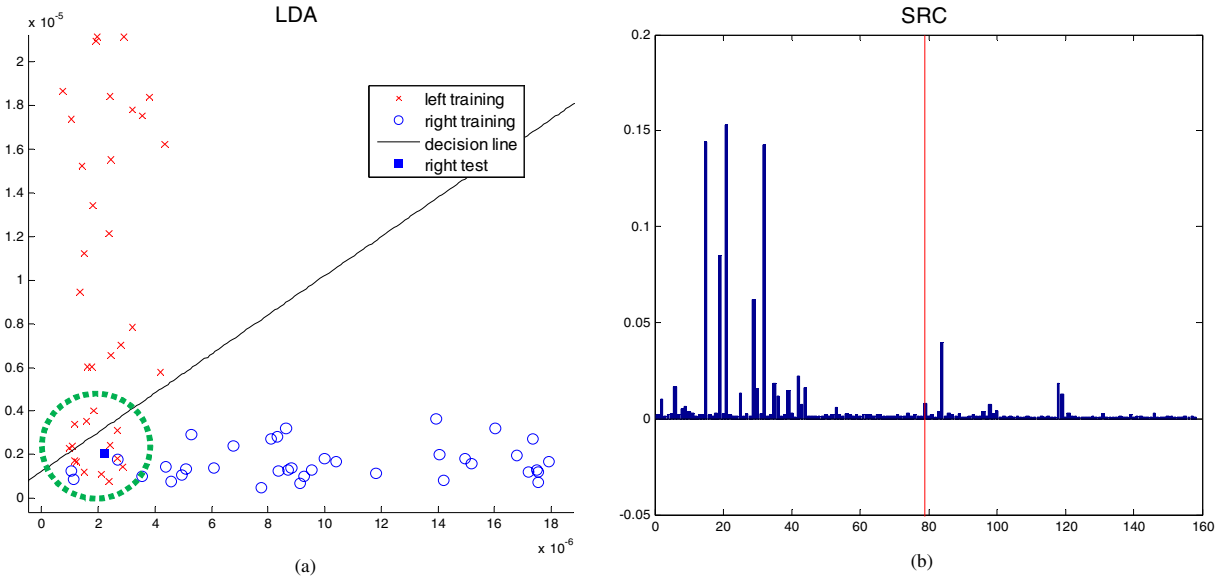

Figure 13. Example of LDA and SRC results when LDA classifies correctly while SRC does not. (a) A two-dimensional scatter plot using two CSP filters ($n = 1$) for subject A from the INFONET dataset. The dotted circle indicates the 'gray' region. (b) Coefficient vector **x** recovered using the same test signal as in (a).

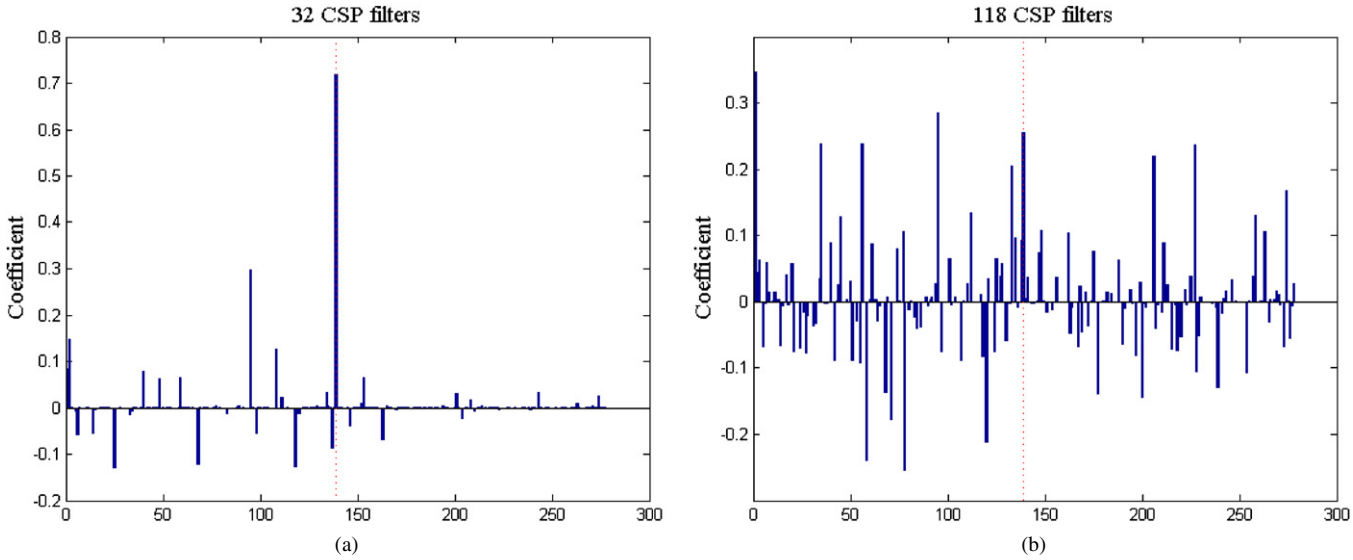

Figure 14. Coefficient vector **x** corresponding to the right-hand test signal of subject al from Berlin dataset. (a) 32 CSP filters and (b) 118 CSP filters were used to produce the dictionary. The central vertical line indicates the boundary between the right-hand and right-foot training signals.

should be larger than or equal to the number of unknown non-zero elements of vector **x** in (10). Therefore, in the proposed SRC method, the number of rows in the dictionary should be carefully selected to strike a balance between avoiding the overfitting problem and satisfying the fundamental limitation.

Figure 14 shows an example of coefficient vector **x**. Figures 14(a) and (b) were acquired using 32 and 118 CSP filters for subject al in the Berlin dataset, respectively. We can see that (a) provided a sparser representation than (b). In addition, (a) gave a higher classification accuracy than (b). We observe that the non-sparse result of (b) was obtained when all 118 CSP filters were used. In the above discussion, we also observed that the use of excessive CSP filters could lead to the classifier overfitting problem. While no direct and precise connection can be made in this work, it is interesting to note that an intriguing relationship might exist between the non-sparse representation result and the classifier overfitting problem when too many filters are used.

6. Conclusions

In this study, we applied the idea of sparse representation to the field of BCIs and proposed a new classification method that delivered good performance for an MI-based BCI application.

We used the well-known band power feature to utilize the ERD concept, which is the most widely used physiological feature in MI-based BCIs. This method required a well-designed dictionary matrix. We proposed a new procedure in which a CSP filtering technique is incorporated to produce the dictionary. We referred to this new classification system as the SRC method in this paper. To validate the SRC method, we applied it not only to an INFONET dataset that we obtained in our laboratory but also to the publicly available Berlin dataset. In addition, we compared the proposed method with one of the well-known classification methods, the LDA classification method. The results indicated that the proposed SRC scheme delivers classification accuracy higher than that of the LDA method. We noted that an incoherently designed dictionary, together with the use of L1 minimization, makes SRC competitive as a classification tool.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MEST) (Do-Yak Research Program, no 2012-0005656).

References

[1] Wolpaw J R, Birbaumer N, McFarland D J, Pfurtscheller G and Vaughan T M 2002 Brain–computer interfaces for communication and control *Clin. Neurophysiol.* **113** 767–91

[2] Farewell L A and Donchin E 1988 Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials *Electroencephalogr. Clin. Neurophysiol.* **70** 510–23

[3] Sellers E and Donchin E 2006 A P300-based brain–computer interface: initial tests by ALS patient *Clin. Neurophysiol.* **117** 538–48

[4] Pfurtscheller G, Flotzinger D and Kalcher J 1993 Brain–computer interface—a new communication device for handicapped persons *J. Microcomput. Appl.* **16** 293–9

[5] Wolpaw J R, McFarland D J, Neat G W and Forneris C A 1991 An EEG-based brain–computer interface for cursor control *Electroencephalogr. Clin. Neurophysiol.* **78** 252–9

[6] Pfurtscheller G, Neuper C, Flotzinger D and Pregenzer M 1997 EEG-based discrimination between imagination of right and left hand movement *Electroencephalogr. Clin. Neurophysiol.* **103** 642–51

[7] Pfurtscheller G and Neuper C 2001 Motor imagery and direct brain–computer communication *Proc. IEEE* **89** 1123–34

[8] Huang K and Aviyente S 2006 Sparse representation for signal classification *Adv. Neural Inf. Process. Syst.* **19** 609–16

[9] Wright J, Yang A Y, Ganesh A, Sastry S S and Ma Y 2009 Robust face recognition via sparse representation *IEEE Trans. Pattern Anal. Mach. Intell.* **31** 210–27

[10] Gemmeke J F, Virtanen T and Hurmalainen A 2011 Exemplar-based sparse representations for noise robust automatic speech recognition *IEEE Trans. Audio Speech Lang. Process.* **19** 2067–80

[11] Donoho D 2006 Compressed sensing *IEEE Trans. Inf. Theory* **52** 1289–306

[12] Baraniuk R 2007 Compressive sensing *IEEE Signal Process. Mag.* **24** 118–21

[13] Candes E J and Wakin M B 2008 An introduction to compressive sampling *IEEE Signal Process. Mag.* **25** 21–30

[14] Nunez P L, Srinivasan R, Westdorp A F, Wijesinghe R S, Tucker D M, Silberstein R B and Cadusch P J 1997 EEG coherency I: statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales *Electroencephalogr. Clin. Neurophysiol.* **103** 499–515

[15] McFarland D J, McCane L M, David S V and Wolpaw J R 1997 Spatial filter selection for EEG-based communication *Electroencephalogr. Clin. Neurophysiol.* **103** 386–94

[16] Ramoser H, Müller-Gerking J and Pfurtscheller G 2000 Optimal spatial filtering of single trial EEG during imagined hand movement *IEEE Trans. Rehabil. Eng.* **8** 441–7

[17] Blankertz B, Tomioka R, Lemm S, Kawanabe M and Müller K-R 2008 Optimizing spatial filters for robust EEG single-trial analysis *IEEE Signal Process. Mag.* **25** 41–56

[18] Graimann B, Allison B and Pfurtscheller G 2010 *Brain–Computer Interfaces: Revolutionizing Human–Computer Interaction* (Berlin: Springer)

[19] Blankertz B *Berlin Brain–Computer Interface* http://www.bbci.de/competition/iii/desc_IVa.html

[20] Fukunaga K 1990 *Introduction to Statistical Pattern Recognition* 2nd edn (Boston, MA: Academic)

[21] Candès E, Romberg J and Tao T 2006 Stable signal recovery from incomplete and inaccurate measurements *Commun. Pure Appl. Math.* **59** 1207–23

[22] Chen S, Donoho D and Saunders M 2001 Atomic decomposition by basis pursuit *SIAM Rev.* **43** 129–59

[23] Donoho D, Stodden V and Tsaig Y SparseLab http://sparselab.stanford.edu/

[24] Dornhege G, del R Millán J, Hinterberger T, McFarland D J and Müller K-R 2007 *Toward Brain–Computer Interfacing* (Cambridge, MA: MIT Press) pp 214–5

[25] Bishop C M 2008 *Neural Networks for Pattern Recognition* (Oxford: Oxford University Press) pp 105–8

[26] Wasserman L 2010 *All of Statistics: A Concise Course in Statistical Inference* (New York: Springer) pp 63–4

[27] Bostanov V 2004 BCI competition 2003—data sets Ib and IIb: feature extraction from event-related brain potentials with the continuous wavelet transform and the *t*-value scalogram *IEEE Trans. Biomed. Eng.* **51** 1057–61

[28] Lotte F, Congedo M, Lecuyer A, Lamarche F and Arnaldi B 2007 A review of classification algorithms for EEG-based brain–computer interfaces *J. Neural Eng.* **4** R1–13

[29] Donoho D L and Huo X 2001 Uncertainty principles and ideal atomic decomposition *IEEE Trans. Inf. Theory* **47** 2845–62

[30] Donoho D L and Elad M 2003 Optimally sparse representation in general (nonorthogonal) dictionaries via $\ell 1$ minimization *Proc. Natl Acad. Sci. USA* **100** 2197–202

[31] Li Y, Guan C and Qin J 2005 Enhancing feature extraction with sparse component analysis for brain–computer interface *Proc. 27th Annual Int. Conf. of the Engineering in Medicine and Biology Society (IEEE-EMBS 2005)* pp 5335–8

[32] Arvaneh M, Guan C, Ang K K and Quek H C 2011 Spatially sparsed common spatial pattern to improve BCI performance *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP 2011)* pp 2412–5

[33] Yu H, Lu H, Ouyang T, Liu H and Lu B L 2010 Vigilance detection based on sparse representation of EEG *Proc. 32nd Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC 2010)* pp 2439–42