

Research Article

Common Spatio-Time-Frequency Patterns for Motor Imagery-Based Brain Machine Interfaces

Hiroshi Higashi^{1,2} and Toshihisa Tanaka^{1,2}

¹ Department of Electronic and Information Engineering, Tokyo University of Agriculture and Technology, 2-24-16 Nakacho, Koganei-shi, Tokyo 184-8588, Japan

² RIKEN Brain Science Institute, 2-1 Hirosawa, Wako-shi, Saitama 351-0106, Japan

Correspondence should be addressed to Hiroshi Higashi; higashi@sip.tuat.ac.jp

Received 7 May 2013; Revised 17 August 2013; Accepted 31 August 2013

Academic Editor: Daoqiang Zhang

Copyright © 2013 H. Higashi and T. Tanaka. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

For efficient decoding of brain activities in analyzing brain function with an application to brain machine interfacing (BMI), we address a problem of how to determine spatial weights (spatial patterns), bandpass filters (frequency patterns), and time windows (time patterns) by utilizing electroencephalogram (EEG) recordings. To find these parameters, we develop a data-driven criterion that is a natural extension of the so-called common spatial patterns (CSP) that are known to be effective features in BMI. We show that the proposed criterion can be optimized by an alternating procedure to achieve fast convergence. Experiments demonstrate that the proposed method can effectively extract discriminative features for a motor imagery-based BMI.

1. Introduction

Brain machine/computer interfacing (BMI/BCI) is a challenging technology of signal processing, machine learning, and neuroscience [1]. BMIs capture brain activities associated with mental tasks and external stimuli, realize nonmuscular communication, and control channel for conveying messages and commands to the external world [1–3]. Basically, noninvasively measured data such as electroencephalogram (EEG), magnetoencephalogram (MEG), and functional magnetic resonance imaging (fMRI) are widely used to observe brain activities. Among them, because of its simplicity and low cost, EEG is practical for use in engineering applications [4, 5].

Efficient decoding around motor-cortex is a crucial technique for the realization of BMI associated with motor-imagery (MI-BMI) [6, 7] with the application to controlling external devices [7], prostheses [4], rehabilitation [8], and so forth. For instance, it is also known that the real and imaginary movements of hands and feet evoke the change of the so-called mu rhythm in different brain regions [2, 3]. Therefore, the accurate extraction of these changes from the measured EEG signals in the presence of measurement noise and spontaneous components which are related to other brain

activities enables us to classify the EEG signal associated with the different motor (imagined) actions such as movement of the right hand, left hand, or feet.

In classification of EEG signals in MI-BMI and analyzing of the brain activities during motor imagery, signal processing techniques such as bandpass filtering and spatial weighting are used [1]. For the processing, presuming the parameters such as coefficients of the filters and weights that extract the related components is a crucial issue. Moreover, the optimum parameters in classification are highly dependent on users and measurement environments [9].

In order to determine the parameters, data-driven techniques that exploit observed data are widely used [1, 2]. The observed data essentially include class labels corresponding to the tasks. The techniques should find the parameters that extract discriminative features as much as possible. For example, the well-known common spatial pattern (CSP) method finds the spatial weights by using the observed signals [1, 9, 10] in such a way that the variances of the signals extracted by the linear combination of a multichannel signal and the spatial weights differ as much as possible between two classes. The standard CSP method has been extended

to methods to estimate the other parameters, such as the frequency bands [11–16], and methods to select the CSP features extracted with various parameters [17, 18].

Besides, one of the parameters to be decided by data-driven techniques is a time window because of the following reasons. A kind of BMIs is implemented based on cues which a user follows. In the BMI, the user begins to perform a task when the cue is given. Therefore, the time when the user begins to perform the task is known. However, the time when the brain activity associated with the task occurs is unknown. The time windows working to remove samples that do not contain the brain activity will not match the period when the cues are showed. For instance, the samples for a few hundreds of milliseconds after the cues are assumed not to be used to extract the features in previous works [13, 17, 19, 20], which heuristically determined the time window. Contrary to these works, this paper hypothesizes that an optimal observation period in classification depends on users. For example, reaction time defined as the elapsed time between the presentation of a sensory stimulus and the subsequent behavioral response is strongly associated with age [21]. The reaction time can be related to the time of response to the cues. Therefore, the time window should also be designed by using the observed signals or data-driven.

In this paper, we propose a method for finding the time windows as well as the aforementioned parameters (CSP and temporal filters) by extending a framework proposed in [15]. The proposed method enables us to find these parameters that separately extract several components that are observed in different spatial patterns, frequency bands, and periods of time. We call these simultaneously designed parameters common spatio-time-frequency patterns (CSTFP). As compared to the methods, in which the parameters are selected out of the predefined candidates fixed in advance [18], the CSTFP method has a higher degree of freedom for the parameters reducing the computational costs. In CSTFP, the coefficients of the temporal filters and the spatial weights are searched in the set of real numbers. Moreover, although the time windows are selected out of a set of candidates, the computational cost of the CSTFP method does not grow rapidly even with a large number of candidates.

The rest of this paper is organized as follows. Section 2.1 reviews the CSP method. Next we, illustrate the CSTFP method from Section 2.2 to Section 2.4. We experimentally analyze the CSTFP method by using artificial data in Section 3. Section 4 presents experimental results of classification of EEG signals during motor imageries to show the performance of CSTFP. Finally, the conclusions of this paper are presented in Section 5.

2. Common Spatio-Time-Frequency Patterns (CSTFP)

We describe a novel method called the CSTFP method to simultaneously find the parameters for the spatial weights, the temporal filters, and the time windows. In this section, before introducing CSTFP, CSP is reviewed. Next, we define a signal extraction model using the spatial weighting, filtering,

and time windowing. Then, we propose a criterion designing the parameters motivated by that for the CSP method. The proposed criterion evaluates not only the spatial weights but also the coefficients of the time windows and the temporal filters. Next, the optimization method for the proposed criterion with alternate updating is presented. Finally, we define feature vector with the feature extraction model for classification of unlabeled EEG signals.

2.1. Common Spatial Pattern (CSP): A Review. Let $\mathbf{X} \in \mathbb{R}^{M \times N}$ be an observed signal, where M is the number of channels and N is the number of samples. In BMI application, we do not directly use \mathbf{X} , but we use the filtered signal described as $\widehat{\mathbf{X}} = \mathcal{H}(\mathbf{X})$ to find the CSP, where \mathcal{H} is a bandpass filter which passes the frequency components related to brain activity of motor imagery. Denote the components of $\widehat{\mathbf{X}}$ by $\widehat{\mathbf{X}} = [\widehat{\mathbf{x}}_1, \dots, \widehat{\mathbf{x}}_N]$, where $\widehat{\mathbf{x}}_n \in \mathbb{R}^M$ and n is the time index. We assume the sets of the observed signals, \mathcal{C}_1 and \mathcal{C}_2 , where \mathcal{C}_d contains the signals belonging to class d , $d \in \{1, 2\}$ is a class label, $\mathcal{C}_1 \cap \mathcal{C}_2 = \emptyset$, and \emptyset is a set having no elements. CSP is defined as the weight vector that maximizes the intra class variance in \mathcal{C}_c under the normalization of samples, where c is a class label. More specifically, for c being fixed, the weight vector is found by solving the following optimization problem [9, 10]:

$$\begin{aligned} \max_{\mathbf{w}} \quad & E_{\mathbf{X} \in \mathcal{C}_c} \left[\frac{1}{N} \sum_{n=1}^N \left| \mathbf{w}^T (\widehat{\mathbf{x}}_n - \boldsymbol{\mu}) \right|^2 \right], \\ \text{subject to} \quad & \sum_{d=1,2} E_{\mathbf{X} \in \mathcal{C}_d} \left[\frac{1}{N} \sum_{n=1}^N \left| \mathbf{w}^T (\widehat{\mathbf{x}}_n - \boldsymbol{\mu}) \right|^2 \right] = 1, \end{aligned} \quad (1)$$

where $E_{\mathbf{X} \in \mathcal{C}_d}[\cdot]$ denotes the expectation over \mathcal{C}_d , $\boldsymbol{\mu}$ is the time average of \mathbf{X} given by $\boldsymbol{\mu} = (1/N) \sum_{n=1}^N \widehat{\mathbf{x}}_n$, \cdot^T is the transpose of a vector or a matrix, and $|\cdot|$ is the absolute value of a scalar. The solution of (1) is given by the generalized eigenvector corresponding to the largest generalized eigenvalue of the generalized eigenvalue problem described as

$$\boldsymbol{\Sigma}_c \mathbf{w} = \lambda (\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2) \mathbf{w}, \quad (2)$$

where $\boldsymbol{\Sigma}_d$, for $d = 1, 2$, are defined as $\boldsymbol{\Sigma}_d = E_{\mathbf{X} \in \mathcal{C}_d} [(1/N) \sum_{n=1}^N (\widehat{\mathbf{x}}_n - \boldsymbol{\mu})(\widehat{\mathbf{x}}_n - \boldsymbol{\mu})^T]$.

2.2. Signal Extraction Model. The target signal and signal extraction procedure are formulated in this section. The filtered signal of a target signal, \mathbf{X} , denoted as $\widehat{\mathbf{x}} = [\widehat{x}_1, \dots, \widehat{x}_K]^T$, is defined as

$$\widehat{x}_n = b_n \sum_{m=1}^M w_m \sum_{p=1}^P h_p [\mathbf{x}_{n+P-p}]_m, \quad (3)$$

for $n = 1, \dots, K$, $K = N - P + 1$, where $[\cdot]_i$ is the i th entry of a vector, P is the filter order of an FIR filter of which the coefficients are denoted by h_1, \dots, h_P , w_m is a spatial weight for m th channel, and b_n is a time window for n th sample that takes a binary value of either 0 or 1. The structure of

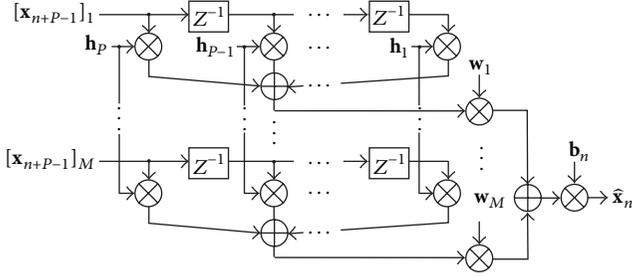


FIGURE 1: The filtering of an observed signal by a temporal filter, \mathbf{h} , spatial weights, \mathbf{w} , and a time window, \mathbf{b} . Z^{-n} is an operator of n samples delay.

temporally filtering, spatial weighting, and windowing for \mathbf{X} is illustrated in Figure 1. In this model, \mathbf{w}_m is regarded as the spatial pattern, \mathbf{h}_p is regarded as the frequency pattern, and \mathbf{b}_n is regarded as the time pattern of the extracted signal.

The sample variance of \hat{x}_n over time $n = 1, \dots, K$ is described as

$$\alpha_{\mathbf{X}}(\mathbf{w}, \mathbf{h}, \mathbf{b}) = \frac{1}{\|\mathbf{b}\|} \sum_{n=1}^K \left| \hat{x}_n - \frac{1}{K} \sum_{i=1}^K \hat{x}_i \right|^2, \quad (4)$$

where \mathbf{w} is defined as $\mathbf{w} = [w_1, \dots, w_M]^T$, \mathbf{h} is defined as $\mathbf{h} = [h_1, \dots, h_p]^T$, \mathbf{b} is defined as $\mathbf{b} = [b_1, \dots, b_N]^T$, and $\|\cdot\|$ is the Euclidean norm of a vector.

The variance defined in (4) can be transformed to matrix-vector form as follows. We define \mathbf{A}_n , $n = 1, \dots, K$, whose elements are from \mathbf{X} as

$$[\mathbf{A}_n]_{m,p} = [\mathbf{X}]_{m,n+p-p}, \quad (5)$$

for $m = 1, \dots, M$ and $p = 1, \dots, P$, where $[\cdot]_{i,j}$ is the element at the i th row and the j th column of a matrix. Then, (4) can be modified to

$$\alpha_{\mathbf{X}}(\mathbf{w}, \mathbf{h}, \mathbf{b}) = \frac{1}{\|\mathbf{b}\|} \sum_{n=1}^K b_n |\mathbf{w}^T \widehat{\mathbf{A}}_n \mathbf{h}|^2, \quad (6)$$

where $\widehat{\mathbf{A}}_n$ is defined as $\widehat{\mathbf{A}}_n = \mathbf{A}_n - \|\mathbf{b}\|^{-1} \sum_{m=1}^K b_m \mathbf{A}_m$.

2.3. Optimization for Sets of Parameters. We consider the problem of the design of F sets of the FIR filter, the spatial weights, and the time window represented by $\{\mathbf{w}_i, \mathbf{h}_i, \mathbf{b}_i\}_{i=1}^F$. These sets of the parameters are designed in such a way that \mathbf{w}_i , \mathbf{h}_i , and \mathbf{b}_i maximize expectation of $\alpha_{\mathbf{X}}(\mathbf{w}_i, \mathbf{h}_i, \mathbf{b}_i)$ with respect to $\mathbf{X} \in \mathcal{C}_c$ samples under the normalization of an expectation of $\alpha_{\mathbf{X}}(\mathbf{w}_i, \mathbf{h}_i, \mathbf{b}_i)$ over all of the observation. Additionally, we impose the orthonormality on \mathbf{h}_i , $i = 1, \dots, F$ to avoid the trivial solution. Moreover, the time windows are chosen from given candidates for efficient

optimization. Therefore, we formulate the following maximization problem:

$$\begin{aligned} & \max_{\mathcal{P}_i, i=1, \dots, F} \sum_{i=1}^F \widehat{J}(\mathcal{P}_i) + \frac{\epsilon}{K} \|\mathbf{b}_i\|, \\ & \text{subject to } \frac{\mathbf{h}_i^T \mathbf{h}_j}{\|\mathbf{h}_i\| \|\mathbf{h}_j\|} = \delta_{ij}, \quad i, j = 1, \dots, F, \quad (7) \\ & \mathbf{h}_k \in \mathcal{S}_k^\perp, \quad k = 1, \dots, F, \\ & \mathbf{b}_l \in \mathcal{B}, \quad l = 1, \dots, F, \end{aligned}$$

where \mathcal{P}_i represents set $\{\mathbf{w}_i, \mathbf{h}_i, \mathbf{b}_i\}$ and $\widehat{J}(\mathcal{P}_i)$ is the cost evaluating the ratio of the feature value in all samples defined as

$$\widehat{J}(\mathcal{P}_i) = \frac{E_{\mathbf{X} \in \mathcal{C}_c} [\alpha_{\mathbf{X}}(\mathbf{w}_i, \mathbf{h}_i, \mathbf{b}_i)]}{\sum_{d=1,2} E_{\mathbf{X} \in \mathcal{C}_d} [\alpha_{\mathbf{X}}(\mathbf{w}_i, \mathbf{h}_i, \mathbf{b}_i)]}, \quad (8)$$

where \mathcal{S}_i is any subspace in \mathbb{R}^P , \mathcal{B} is a candidate set for the time windows, defined as $\mathcal{B} = \{\mathbf{b}_l\}_{l=1}^L$, c is a class label chosen from 1 and 2, ϵ is a regularization parameter, and δ_{ij} is the Kronecker delta defined as 1 for $i = j$ and 0, otherwise.

Since it is difficult to simultaneously find all parameters, we consider sequential optimization to find the parameters with respect to each filter index i . That is, we first find \mathcal{P}_1 , and then find \mathcal{P}_2 under the constraint on \mathbf{h}_1 . This sequential optimization is represented with respect to each i as

$$\begin{aligned} & \max_{\mathcal{P}_i} \widehat{J}(\mathcal{P}_i) + \frac{\epsilon}{K} \|\mathbf{b}_i\|, \quad (9) \\ & \text{subject to } \mathbf{h}_i \in \widehat{\mathcal{S}}_i^\perp, \quad \mathbf{b}_i \in \mathcal{B}, \end{aligned}$$

where $\widehat{\mathcal{S}}_i$ is a subspace defined as

$$\widehat{\mathcal{S}}_i = \text{Span}\{\mathbf{h}_1, \dots, \mathbf{h}_{i-1}\} \oplus \mathcal{S}_i \in \mathbb{R}^P, \quad (10)$$

where $\text{Span}(\dots)$ represents a subspace spanned by vectors and the operator denoted by \oplus gives the direct sum of two subspaces. Methods for choosing $\widehat{\mathcal{S}}_i$ has been discussed in [15]. In (9), to optimize the parameters indexed with i , we adopt an alternating optimization procedure based on alternating least square (ALS). In the optimization, we separate the problem of (9) into three subproblems for \mathbf{w}_i , \mathbf{h}_i , and \mathbf{b}_i , respectively. Then, we update the parameters by alternating solving the subproblems. The three subproblems and these solutions are as follows.

The first subproblem is to optimize \mathbf{w}_i . While fixing \mathbf{h}_i and \mathbf{b}_i , \mathbf{w}_i maximizing (9) is found as the generalized eigenvector corresponding to the largest generalized eigenvalue of the generalized eigenvalue problem [15] described as

$$\mathbf{R}_c \mathbf{w}_i = \lambda (\mathbf{R}_1 + \mathbf{R}_2) \mathbf{w}_i, \quad (11)$$

where

$$\mathbf{R}_d = E_{\mathbf{X} \in \mathcal{C}_d} \left[\frac{1}{\|\mathbf{b}_i\|} \sum_{n=1}^K [\mathbf{b}_i]_n \widehat{\mathbf{A}}_n \mathbf{h}_i \mathbf{h}_i^T \widehat{\mathbf{A}}_n^T \right], \quad (12)$$

for $d = 1, 2$, and λ is an eigenvalue.

Input: \mathcal{E}_1 and \mathcal{E}_2 : the sets of observed signals.
Parameters: F : the number of FIR filters, P : the filter order,
 \mathcal{B} : the set of the candidates for the time window.
Output: $\{\mathbf{w}_i\}_{i=1}^F$: the spatial weights, $\{\mathbf{h}_i\}_{i=1}^F$: the vectors of
FIR filter coefficients, $\{\mathbf{b}_i\}_{i=1}^F$: the time window.
for $i = 1, \dots, F$ **do**
 Initialize \mathbf{h}_i and \mathbf{b}_i .
 Set the index of iteration as $k = 0$.
 repeat
 $k \leftarrow k + 1$.
 Update \mathbf{w}_i by solving (11).
 Update \mathbf{h}_i by solving (13).
 Update \mathbf{b}_i by solving (19).
 Calculate cost, C_k from the cost function, $\hat{J}(\mathcal{P}_i)$.
 until $C_k - C_{k-1}$ is sufficiently small.
end for

ALGORITHM 1: Design of the FIR filters, the spatial weights, and the time windows.

The second subproblem is to optimize \mathbf{h}_i . While fixing \mathbf{w}_i and \mathbf{b}_i , \mathbf{h}_i maximizing (9) is found as the generalized eigenvector corresponding to the largest generalized eigenvalue of the generalized eigenvalue problem [15] described as

$$\mathbf{G}\mathbf{Q}_c\mathbf{h}_i = \zeta(\mathbf{Q}_1 + \mathbf{Q}_2)\mathbf{h}_i, \quad (13)$$

where

$$\mathbf{Q}_d = E_{\mathbf{X} \in \mathcal{E}_d} \left[\frac{1}{\|\mathbf{b}_i\|} \sum_{n=1}^K [\mathbf{b}_i]_n \widehat{\mathbf{A}}_n^T \mathbf{w}_i \mathbf{w}_i^T \widehat{\mathbf{A}}_n \right], \quad (14)$$

for $d = 1, 2$,

$$\mathbf{G} = \mathbf{I}_P - \mathbf{V}(\mathbf{V}^T(\mathbf{Q}_1 + \mathbf{Q}_2)^{-1}\mathbf{V})^{-1}\mathbf{V}^T(\mathbf{Q}_1 + \mathbf{Q}_2)^{-1}, \quad (15)$$

$$\mathbf{V} = [\mathbf{v}_{i1}, \dots, \mathbf{v}_{iD_i}] \in \mathbb{R}^{P \times D_i},$$

$\mathbf{v}_{i1}, \dots, \mathbf{v}_{iD_i}$ are vectors spanning $\widehat{\mathcal{S}}_i$ described as

$$\text{Span} \{ \mathbf{v}_{i1}, \dots, \mathbf{v}_{iD_i} \} = \widehat{\mathcal{S}}_i, \quad (16)$$

where \mathbf{I}_P is the $P \times P$ identity matrix and ζ is an eigenvalue.

The third subproblem is to optimize \mathbf{b}_i while fixing \mathbf{w}_i and \mathbf{h}_i . The cost of (9) can be reduced to

$$J_3(\mathbf{b}_i | \mathbf{w}_i, \mathbf{h}_i) = \frac{g_1(\mathbf{b}_i)}{g_1(\mathbf{b}_i) + g_2(\mathbf{b}_i)} + \frac{\epsilon}{K} \|\mathbf{b}_i\|. \quad (17)$$

Then,

$$g_d(\mathbf{b}) = E_{\mathbf{X} \in \mathcal{E}_d} \left[\frac{1}{\|\mathbf{b}\|} \sum_{n=1}^K \left(b_n \tilde{x}_n - \frac{1}{\|\mathbf{b}\|} \sum_{k=1}^K b_k \tilde{x}_k \right)^2 \right] \quad (18)$$

$$= \frac{1}{\|\mathbf{b}\|} E_{\mathbf{X} \in \mathcal{E}_d} [\mathbf{b}^T (\boldsymbol{\xi} + 2\mu\tilde{\mathbf{x}}) + \mu],$$

where $\tilde{x}_n = \mathbf{w}_i^T \mathbf{A}_n \mathbf{h}_i$, $\tilde{\mathbf{x}} = [\tilde{x}_1, \dots, \tilde{x}_K]^T$, $\boldsymbol{\xi} = [\tilde{x}_1^2, \dots, \tilde{x}_K^2]^T$, and $\mu = \|\mathbf{b}\|^{-1} \mathbf{b}^T \tilde{\mathbf{x}}$. Because \mathbf{b}_i are chosen out of \mathcal{B} , we

calculate the values $J_3(\widehat{\mathbf{b}}_i)$ for all candidates in \mathcal{B} and the optimal \mathbf{b}_i can be chosen as the candidates that maximize J_3 . This formulates that

$$\mathbf{b}_i = \arg \max_{\mathbf{b} \in \mathcal{B}} J_3(\mathbf{b}). \quad (19)$$

The procedure to design the spatial weights, the temporal filters, and the time windows is summarized in Algorithm 1 as a pseudocode.

2.4. Feature Vector Definition. With designed \mathbf{w}_i , \mathbf{h}_i , and \mathbf{b}_i , the feature vector of an EEG signal, \mathbf{X} , for classification is defined as

$$\mathbf{y} = [\alpha_{\mathbf{X}}(\widehat{\mathbf{w}}_1^{(1)}, \mathbf{h}_1, \mathbf{b}_1), \dots, \alpha_{\mathbf{X}}(\widehat{\mathbf{w}}_1^{(2r)}, \mathbf{h}_1, \mathbf{b}_1), \dots, \alpha_{\mathbf{X}}(\widehat{\mathbf{w}}_F^{(1)}, \mathbf{h}_F, \mathbf{b}_F), \dots, \alpha_{\mathbf{X}}(\widehat{\mathbf{w}}_F^{(2r)}, \mathbf{h}_F, \mathbf{b}_F)]^T, \quad (20)$$

where the set of $\{\widehat{\mathbf{w}}_i^{(1)}, \dots, \widehat{\mathbf{w}}_i^{(2r)}\}$ is the CSPs corresponding to \mathbf{h}_i and \mathbf{b}_i . $\widehat{\mathbf{w}}_i^{(m)}$, $i = 1, \dots, F$, $m = 1, \dots, 2r$, are decided as follows. By solving (11) with \mathbf{h}_i and \mathbf{b}_i , we obtain M spatial patterns as $\widehat{\mathbf{w}}_i^{(m)}$ for $i = 1, \dots, F$ and $m = 1, \dots, M$, where $\widehat{\mathbf{w}}_i^{(m)}$ is the unit-length eigenvector corresponding to the m th largest eigenvalue of (11). Then, $\widehat{\mathbf{w}}_i^{(m)}$ are defined as $\widehat{\mathbf{w}}_i^{(j)} = \widehat{\mathbf{w}}_i^{(j)}$ and $\widehat{\mathbf{w}}_i^{(r+j)} = \widehat{\mathbf{w}}_i^{(M-j+1)}$ for $i = 1, \dots, F$, $j = 1, \dots, r$.

3. Experimental Analysis of Artificial Signal

We give an analysis of CSTFP by a toy experiment with an artificial signal in this section. We assume a 2-class BMI where the observed EEG signals are modeled by a mixture of narrow-band signals (see Figure 2). In this model, a trial signal belonging to class d is given by

$$\mathbf{x}[n] = \sum_{i=1}^{N_s} s_i[n] \mathbf{a}_i^{(d)} + \boldsymbol{\eta}, \quad n = 0, \dots, N-1, \quad (21)$$

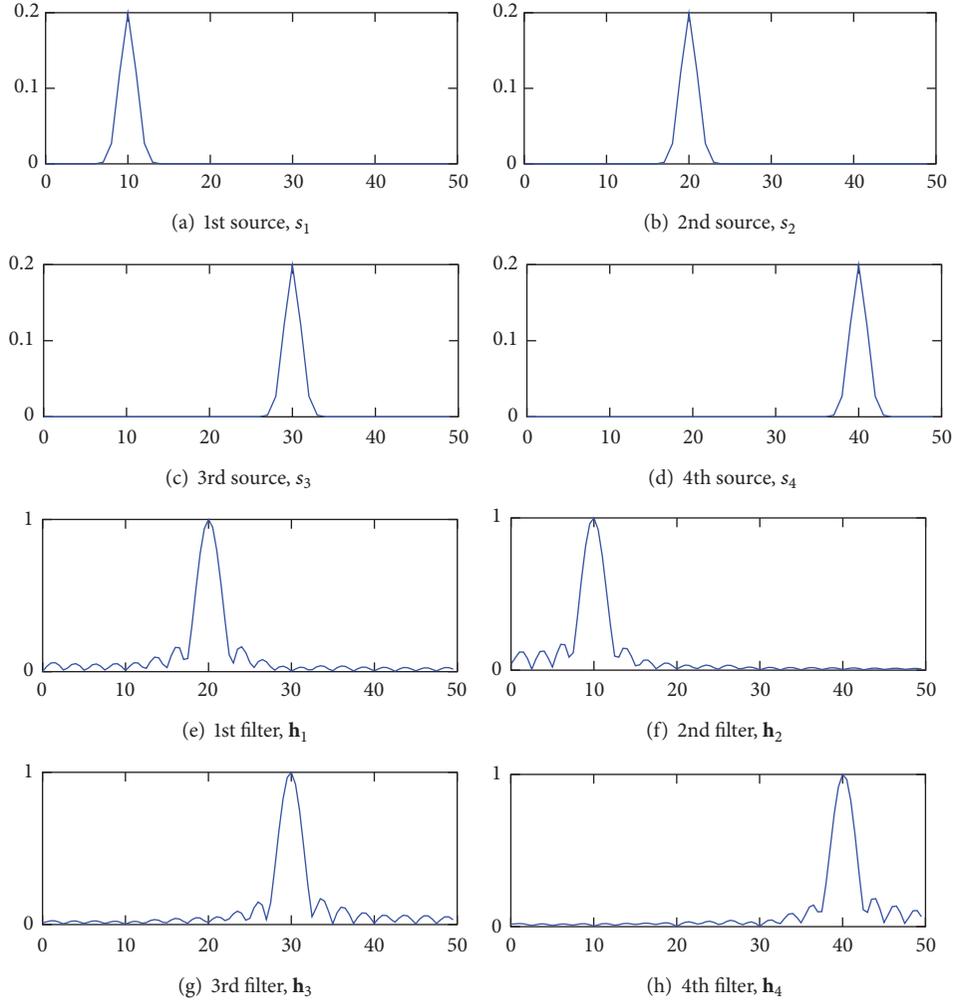


FIGURE 2: Frequency patterns of the spectra of the source signals ((a)–(d)) and the amplitude characteristics of the designed frequency patterns ((e)–(h)). The horizontal axis represents frequency and its unit is Hz.

where d is a class label taking either 1 or 2, $\mathbf{x}[n] \in \mathbb{R}^M$ is a vector representing a signal at a discrete time point n , N is the number of time samples for a trial, M is the number of channels, $s_i[n] \in \mathbb{R}$ is an i th source signal of feature components, N_s is the number of the source signals, $\mathbf{a}_i^{(d)} \in \mathbb{R}^M$ is a vector defined as $\mathbf{a}_i^{(d)} = [a_{i1}^{(d)}, \dots, a_{iM}^{(d)}]^T$, $a_{im}^{(d)} \in \mathbb{R}$ is the amplitude of $s_i[n]$ in the m th channel for class d , and $\boldsymbol{\eta} \in \mathbb{R}^M$ is a stochastic noise. The source signals, $s_i[n]$, are generated as

$$s_i[n] = t_i[n] \Re \left[\sum_{k=0}^{N-1} S_i[k] e^{j\theta} e^{j(2\pi k/N)n} \right], \quad (22)$$

where $S_i[k] \in \mathbb{R}$ represents a discrete spectrum, $t_i[n]$ represents the time window that decides the period when the i th source signal is generated, $\theta \in \mathbb{R}$ is a stochastic phase of the source signals, and the operator denoted by \Re takes a real part of a complex number.

In particular, 200 artificial signals that we used in the experiment were generated with the conditions shown in

TABLE 1: The conditions for generating the artificial signals.

Parameter	Value or distributions
Number of channel, M	10
Number of samples, N	100
Number of trials for each class	100
Sampling frequency	100 Hz
Number of sources, N_s	4
Spectra of sources, $ S_i[k] $	Figures 2(a)–2(d)
Stochastic phase, θ	$\mathcal{U}(0, 2\pi)$
Time windows, $t_i[n]$	Figures 3(a)–3(d)
Amplitudes, $\mathbf{a}_i^{(d)}$	Figures 4(a)–4(d)
Stochastic noise, $[\boldsymbol{\eta}]_m$	$\mathcal{N}(0, 0.1)$

Table 1 where $\mathcal{N}(m, \sigma^2)$ is a Gaussian distribution with a mean, m , and a variance, σ^2 , and $\mathcal{U}(a, b)$ is a uniform distribution whose minimum and maximum values are denoted by a and b , respectively.

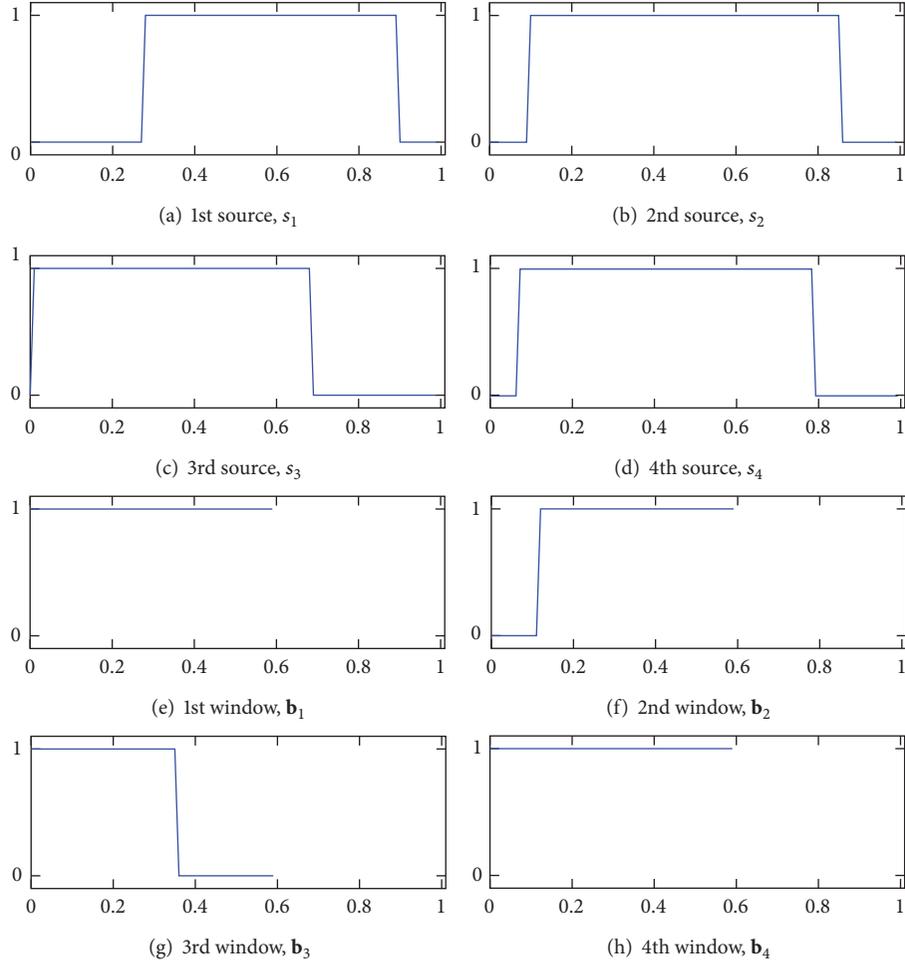


FIGURE 3: Time patterns as the time windows of the source signals ((a)–(d)) and the designed time windows ((e)–(h)). The horizontal axis represents time and its unit is the second.

We applied the CSTFP method to the artificial signals as follows. The class label represented by c in (8) was set to 1. The number of the sets of the parameters, F , was set to 4. The order of the temporal filters was set to 41, yielding that $P = 41$; the length of a filtered signal, K , is 60. For the given candidates for the time windows, we define the following set. First, we define ten K -dimensional vectors as

$$\mathbf{d}_j = \left[\underbrace{0, \dots, 0}_{D(j-1)}, \underbrace{1, \dots, 1}_D, 0, \dots, 0 \right]^T, \quad j = 1, \dots, 10, \quad (23)$$

where $D = 6$. Then, we use all combinations of $\{\mathbf{d}_j\}_{j=1}^{10}$ represented by

$$\widehat{\mathbf{b}}_l = p_1 \mathbf{d}_1 + \dots + p_{10} \mathbf{d}_{10} \quad (24)$$

and $p_j \in \{0, 1\}$, $j = 1, \dots, 10$, as the given candidate set. Therefore, the number of the candidates, L , was 1023. Moreover, for $\widehat{\mathcal{S}}_i$ that determines the search space for \mathbf{h}_i , we used $\widehat{\mathcal{S}}_i = \text{Span}\{\mathbf{h}_1, \dots, \mathbf{h}_{i-1}, \mathbf{h}'_{i-1}, \dots, \mathbf{h}'_1\}$ [15], where each element of \mathbf{h}'_j is defined as $[\mathbf{h}'_j]_k = [\mathbf{h}_j]_{((k+1) \bmod P)}$, $k =$

$1, \dots, P$, and the operator denoted by $a \bmod b$ takes a residue of dividing a by b . In addition, the regularization parameter, ϵ , was set to 0.05.

The optimization resulted in Figures 2(e)–2(h) for the amplitude characteristics of the FIR filters, Figures 3(e)–3(h) for the time windows, and Figures 4(e)–4(h) for the normalized spatial weights. The centers of passbands of the filters shown in Figures 2(e)–2(h) coincide with the centers of the source signals shown in Figures 2(a)–2(d). Moreover, in the spectrum of the source and the amplitude characteristic of the designed FIR filter that have similar center frequencies, the spatial amplitude corresponding to the source and the spatial weight vector corresponding to the FIR filter are similar to each other. For instance, the correlation coefficient between $\mathbf{a}_1^{(1)}$ shown in Figure 4(a) with circles and $\mathbf{w}_2^{(1)}$ shown in Figure 4(f) with circles is 0.971.

The results also suggest that the time windows designed by the CSTFP method can remove samples observed in the periods of time that do not contain the source signals. For instance, hence, the source signal, $s_1[n]$, is not observed in the first 25 samples according to Figure 3(a), the time window for extracting $s_1[n]$ is expected to remove the first 25 samples.

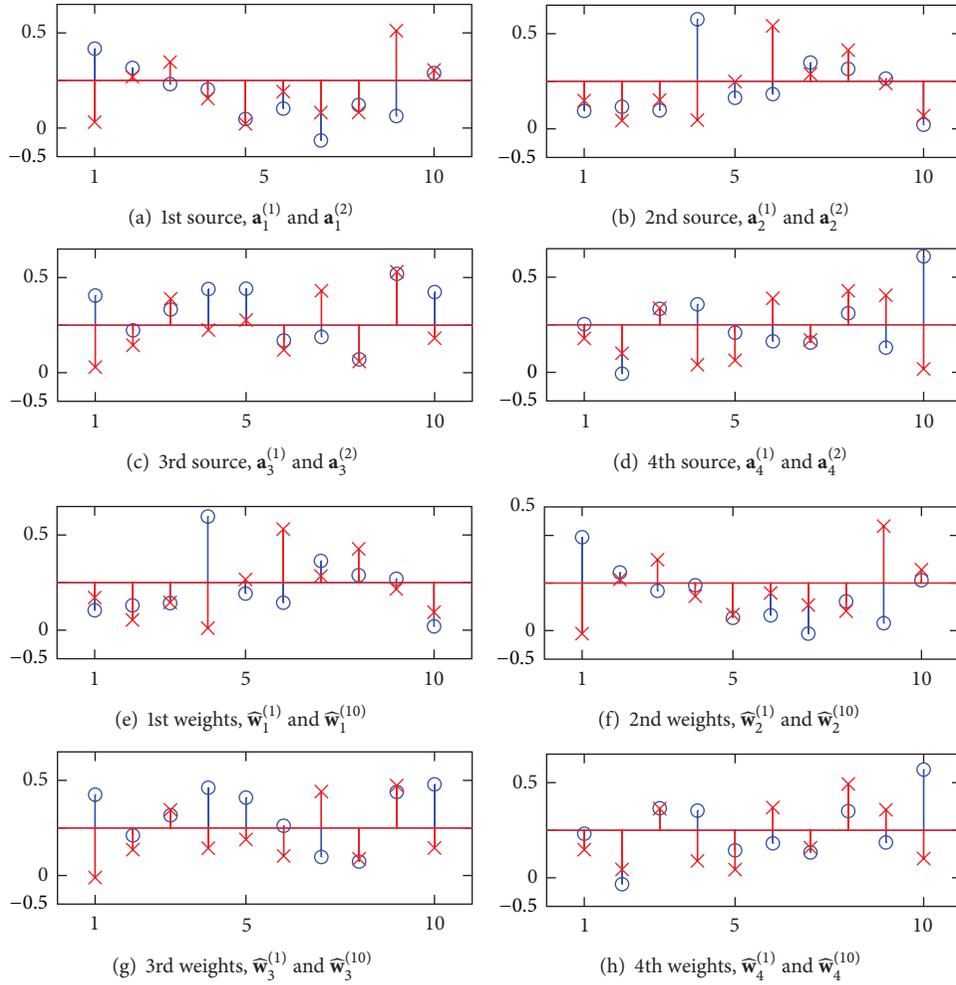


FIGURE 4: Spatial patterns as the spatial amplitudes of the source signals ((a)–(d)) and the designed spatial weights ((e)–(h)). The amplitudes plotted by circles are $\mathbf{a}_i^{(1)}$ and the amplitudes plotted by cross are $\mathbf{a}_i^{(2)}$ for $i = 1, \dots, 4$. The weights plotted by circles are $\mathbf{w}_i^{(1)}$ and the weights plotted by cross are $\mathbf{w}_i^{(10)}$ for $i = 1, \dots, 4$. The horizontal axis represents the channel number.

Because $|S_1[k]|$ and the amplitude characteristics shown in Figure 2(f) have similar center frequency, we can decide that the time window for extracting $s_1[n]$ is the time window shown in Figure 3(f). Although the designed time windows do not coincide with $t_i[n]$ because \mathbf{b} is applied to an FIR-filtered signal that is shorter than that of the original one, the time window removes samples in the first 10 samples, as we expected. Moreover, we can decide that the time window shown in Figure 3(c) is for extracting $s_3[n]$ due to the same reason. As the observed signals do not have $s_3[n]$ in the last 25 samples (see Figure 3(c)), the time window shown in Figure 3(c) removes the samples in the last 20 samples.

4. Experiment of EEG Signal Classification

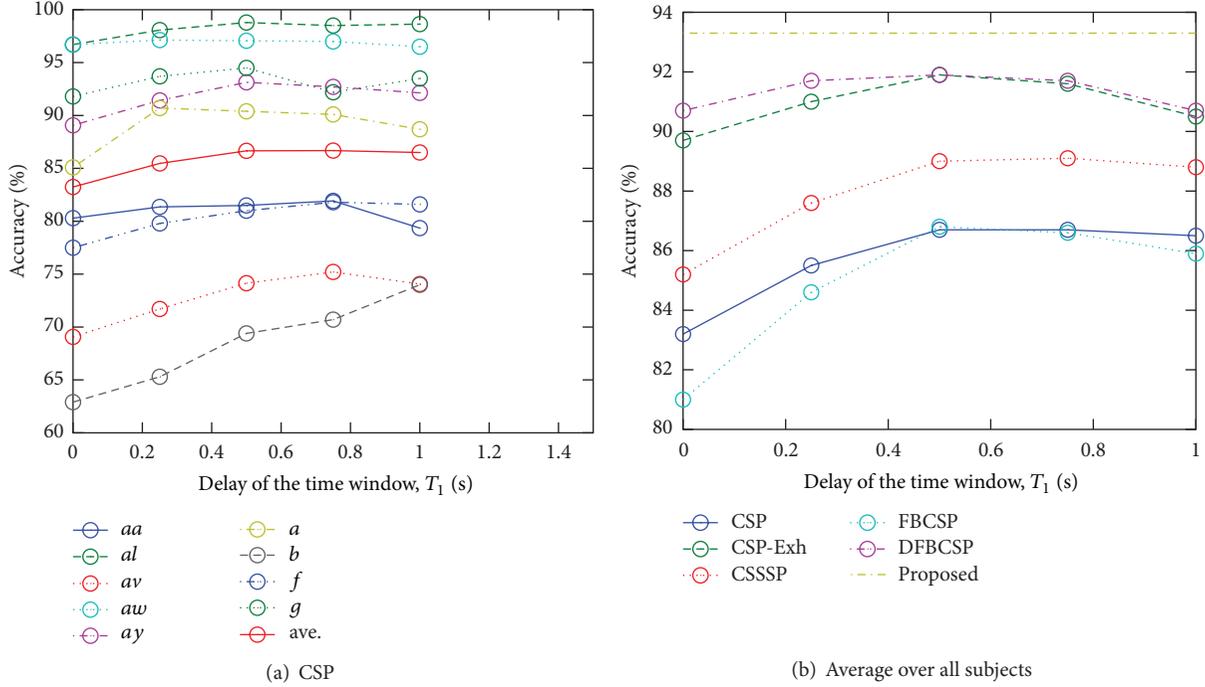
A comprehensive comparative study was performed to illustrate the ability of the CSTFP method to produce more accurate classification of EEG signals during motor imagery over several conventional methods (CSP [10], common sparse spectral patterns (CSSSP) [12], filter bank CSP

TABLE 2: Description of the datasets.

	Dataset IVa	Dataset 1
Classes	Right hand and right foot	2 tasks from foot, left hand, and right hand
Subject labels	<i>aa, al, av, aw, ay</i>	<i>a, b, f, g</i>
Number of channels	118	59
Signal length	3.5 secs	4 secs
Sampling rate	100 Hz	100 Hz
Number of the trials per class	140	100

(FBCSP) [17], and discriminative filter bank CSP (DFBCSP) [15]).

4.1. Data Description. We used dataset IVa from BCI competition III [22], which was provided by Fraunhofer FIRST

FIGURE 5: The variation of the accuracy rates by various T_1 .

(Intelligent Data Analysis Group) and Campus Benjamin Franklin of the Charité - University Medicine Berlin (Department of Neurology, Neurophysics Group) [23] and dataset 1 from BCI competition IV, which was provided by Berlin Institute of Technology (Machine Learning Laboratory), Fraunhofer FIRST (Intelligent Data Analysis Group), and Campus Benjamin Franklin of the Charité - University Medicine Berlin (Department of Neurology, Neurophysics Group) [24]. The condition for each dataset is shown in Table 2. They have two classes of motor imagery. The signals in the provided datasets were recorded with the sampling rate of 1000 Hz.

We furthermore applied to this dataset a Butterworth low-pass filter whose cutoff frequency is 50 Hz and the filter order is 4, and downsampled to 100 Hz.

4.2. Result. For the experiments, as a sample for each trial, we used a signal observed in the period from T_1 to T_2 [sec] after the cue that directs the subject to perform the task. In the experiments, T_1 was tuned by a method we mention later. T_2 was set to 3.5 and 4 seconds for BCI competition III dataset IVa and IV dataset 1, respectively.

In order to compare the classification abilities for the methods, we obtained the classification accuracy rates by 5×5 cross-validation (CV). In each classification in the CV, we separated learning samples for selecting the parameters of the feature extraction and a linear discriminant analysis (LDA) classifier and test samples for obtaining classification accuracy rates.

For the methods to be compared (CSP, CSP-Exh, CSSSP, FBCSP, DFBCSP, and CSTFP), the parameters for the feature extraction were obtained as follows.

TABLE 3: The parameters decided by the nested CV in the learning data in the classification experiments.

Method	Parameters and candidates
CSP	$T_1 \in \{0, 0.25, 0.5, 0.75\}$ $r \in \{1, 2, \dots, 10\}$
CSP-Exh	$T_1 \in \{0, 0.25, 0.5, 0.75\}$ $r \in \{1, 2, \dots, 10\}$ $f_i \in \{1, 2, \dots, 48\}$ $f_u \in \{f_i + 1, f_i + 2, \dots, 49\}$
CSSSP	$T_1 \in \{0, 0.25, 0.5, 0.75\}$ $r \in \{1, 2, \dots, 10\}$ $C \in \{0, 0.01, 0.1, 0.2, 0.5, 1, 2, 5\}$
FBCSP	$T_1 \in \{0, 0.25, 0.5, 0.75\}$ $r \in \{1, 2, \dots, 10\}$
DFBCSP	$T_1 \in \{0, 0.25, 0.5, 0.75\}$ $r \in \{1, 2, \dots, 10\}$ $F \in \{1, 2, \dots, 5\}$
CSTFP	$r \in \{1, 2, \dots, 10\}$ $F \in \{1, 2, \dots, 5\}$

- (i) CSP: the parameters determined in this method are spatial weights. Before obtaining the spatial weights by the CSP method, we applied the Butterworth bandpass filter with the passband of 7–30 Hz. In the CSP method, we minimized the variance cost of the right hand class in (1). The eigenvectors corresponding to the r largest and r smallest eigenvalues of the eigenvalue problem (2) were given as the spatial weights.

TABLE 4: Classification accuracy [%] given by 5×5 CV in dataset IVa from BCI competition III. The figure with \pm represents the standard deviation (S.D.) over CV.

Method	Subject					Average
	<i>aa</i>	<i>al</i>	<i>av</i>	<i>aw</i>	<i>ay</i>	
CSP	79.9 \pm 4.6	98.4 \pm 1.6	71.9 \pm 3.5	96.8 \pm 2.6	92.5 \pm 3.1	87.9
CSP-Exh	90.8 \pm 3.6	99.1 \pm 1.4	74.4 \pm 5.5	99.1 \pm 1.4	93.0 \pm 4.2	91.3
CSSSP	91.5 \pm 3.7	99.1 \pm 1.6	71.1 \pm 6.9	98.8 \pm 1.4	93.1 \pm 3.5	90.7
FBCSP	91.3 \pm 3.4	99.3 \pm 1.5	70.7 \pm 7.4	98.2 \pm 1.6	88.4 \pm 4.8	89.6
DFBCSP	91.9 \pm 2.6	98.9 \pm 1.5	74.8 \pm 4.0	98.9 \pm 1.6	96.0 \pm 3.1	92.1
CSTFP	92.6 \pm 2.1	98.9 \pm 1.5	75.4 \pm 5.3	99.0 \pm 1.3	96.0 \pm 2.4	92.5

TABLE 5: Classification accuracy [%] given by 5×5 CV in dataset 1 from BCI competition IV. The figure with \pm represents the standard deviation (S.D.) over CV.

Method	Subject				Average
	<i>a</i>	<i>b</i>	<i>f</i>	<i>g</i>	
CSP	89.6 \pm 5.1	68.8 \pm 6.7	78.1 \pm 6.5	93.9 \pm 3.4	82.6
CSP-Exh	92.3 \pm 2.5	85.3 \pm 6.9	78.1 \pm 6.5	93.9 \pm 3.6	90.2
CSSSP	89.8 \pm 5.5	70.4 \pm 9.5	85.3 \pm 5.3	93.4 \pm 4.1	84.7
FBCSP	87.3 \pm 8.6	63.9 \pm 14.1	78.6 \pm 6.9	94.1 \pm 4.1	81.0
DFBCSP	89.7 \pm 5.0	85.6 \pm 6.9	92.5 \pm 4.4	93.7 \pm 4.3	90.4
CSTFP	91.4 \pm 4.0	90.6 \pm 7.2	93.3 \pm 3.6	93.5 \pm 4.1	92.2

- (ii) CSP-Exh: the parameters determined in this method are spatial weights and a passband of the Butterworth filter. The passband of the Butterworth filters was tuned as $f_l - f_u$ Hz by an exhaustive search by the CSP method and the learning samples. After the filtering with the passband, the spatial weights were given by the same manner as that of the CSP method.
- (iii) CSSSP: the parameters determined in this method are spatial weights and a bandpass filter. The bandpass filter between 7 and 30 Hz was applied as preprocessing [12]. CSSSP was applied with regularization parameters, C , and the parameter for the number of the spatial weights, r . The order of the filter was fixed to 16 [12].
- (iv) FBCSP: the parameters determined in this method are r bandpass filters out of a filter bank and associated spatial weights. FBCSP was applied with the mutual information based best individual feature and a naïve Bayesian Parzen window (NBPW) classifier [17]. The filter bank comprising 9 bandpass filters covering 4–40 Hz was used. All filters were Chebyshev type II filters with a bandwidth of 4 Hz each. In FBCSP, the number of the spatial weights, N_M , in each band was set to 8. These parameters were decided by referring to [17].
- (v) DFBCSP: the parameters determined in this method are F FIR filters and r spatial weights associated to each FIR filter. DFBCSP was applied with the FIR filter order of 41 as done in [15]. In optimization, we stopped iteration when error of the cost function between successive iterations becomes under 10^{-5} .

- (vi) CSTFP: the parameters determined in this method are F FIR filters, the corresponding F time windows, and r spatial weights associated with each FIR filter. We fixed $T_1 = 0$ to observe the behavior of the resulting time window. CSTFP is applied with the FIR filter order, 41 [15], and the following candidate set for the time windows. The candidate set for the time windows, \mathcal{B} , consists of the vectors defined as

$$\hat{\mathbf{b}}_l = \left[\underbrace{0, \dots, 0}_D, \underbrace{1, \dots, 1}_O, 0, \dots, 0 \right]^T, \quad (25)$$

where we choose D and O out of a set $\{0, 5, 10, \dots, K\}$ such that $O > 50$ and $D + O \leq K$. The regularization parameter, ϵ , was set to 0.1. In alternating optimization, we initialize \mathbf{h}_i as a random vector which is orthonormalized from $\mathbf{v}_1, \dots, \mathbf{v}_D$, and \mathbf{b}_i as a vector all the elements of which are one. We stopped iteration when error of the cost function between successive iterations becomes under 10^{-5} .

For the obscure parameters such as r in the above list, we furthermore tuned them by 5×5 CV using the learning samples as done in [25]. We conducted the nested CV [25] with all of the combinations of these parameters and obtained the classification accuracy rates. We adopted the combination that performed the highest rates as the parameters. The parameters tuned by the nested CV in the learning data and the candidates for them are summarized in Table 3.

After we obtained the feature vectors extracted by the filters, spatial weights, and the time windows that are designed by each listed method, we calculated the logarithm

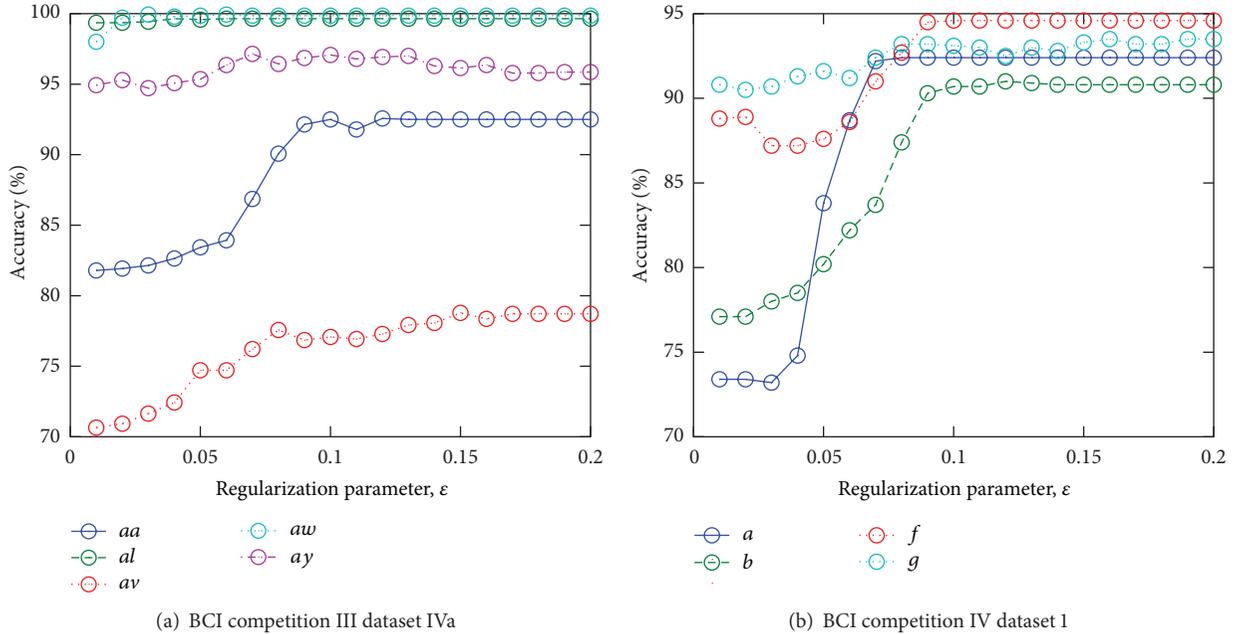


FIGURE 6: The variation of the accuracy rates by various regularization parameters, ϵ .

of the feature vectors. Then, the LDA for the learning samples was used to obtain a projector onto the 1-dimensional space. The threshold for classification was determined as the middle point of two class averages over the projected learning samples. The feature vectors from the test samples were classified by the projection and the threshold and we obtained the classification accuracy rate in each CV. In Tables 4 and 5, CSTFP results in the highest accuracy rate in the average over whole subjects.

Moreover, we conducted the classification experiments, in which the parameters shown in Table 3 were fixed to show the effects of the changes of the parameters on the classification accuracy rates. Figure 5 shows the accuracy rates of CSP and the rates of each method averaged over all subjects with each T_1 . For each subject, the parameters shown in Table 3 except for T_1 were fixed to the combinations of the parameters that performed the highest accuracy rates with each T_1 . In Figure 5(a), T_1 that perform the highest accuracy rates are different among the subjects. Moreover, Figure 5(b) shows that the accuracy rates highly depend on T_1 in the conventional methods. Figure 6 shows the variation of the accuracy rates by the various regularization parameters, ϵ , in CSTFP. For each subject, the parameters, r and F , were fixed to the combinations of the parameters that performed the highest accuracy rates with each ϵ . CSTFP performs higher accuracy as ϵ goes higher than 0.1.

We show examples of the spatial patterns, the frequency patterns (the amplitude characteristics of the FIR filters), and the time patterns designed by CSTFP in Figures 7 and 8. As we can observe in Figures 6, 7(a), and 8(a), the short time windows caused by small ϵ result in poor classification accuracy rates. All of the time windows shown in Figures 7 and 8 remove the samples observed within a few hundreds of milliseconds after the cue. The results suggest that the brain activities related to the task cannot be observed just

after the cue. In Figures 7(c)–7(e) and 8(c)–8(e), the time windows do not significantly change in around 0.1–0.4 of ϵ . This result of the relations between ϵ and the time windows corresponds with the result shown in Figure 6 in which the classification accuracy rates do not strongly depend on the regularization parameter, ϵ , by more than 0.12. Moreover, the FIR filters do not highly differ even with various ϵ . However, the spatial weights with the small windows (Figures 7(a) and 8(a)) differ from the other weights with the longer windows. Furthermore, there are slight differences in the time patterns and the frequency patterns between subject *aa* and subject *av*.

5. Conclusion

We have proposed a novel method called CSTFP for the classification of EEG signals during motor imagery. Our objective is to design the time windows that are adopted in signal processing for a cue-based BMI. Incorporating the idea into DFBCSP has allowed us to simultaneously design the parameters for the time windows, the spatial weights, and the FIR filters. These parameters are optimized in a single criterion based on the CSP method. We have shown the optimization procedure for the problem of the CSTFP method that is conducted by sequentially and alternatively solving the subproblems into which the original problem is divided. Through the experiments of the artificial signals and actual EEG signals, we have shown the performance of CSTFP. In the experiment, we have demonstrated that CSTFP achieves high classification accuracy rate. Our experimental results also suggest that the CSTFP method can find the frequency bands and the time periods in which brain activities associated with a mental task can be observed.

Methods for finding time intervals in which the feature components associated to some brain activities are observed are needed for accurate classification in an asynchronous

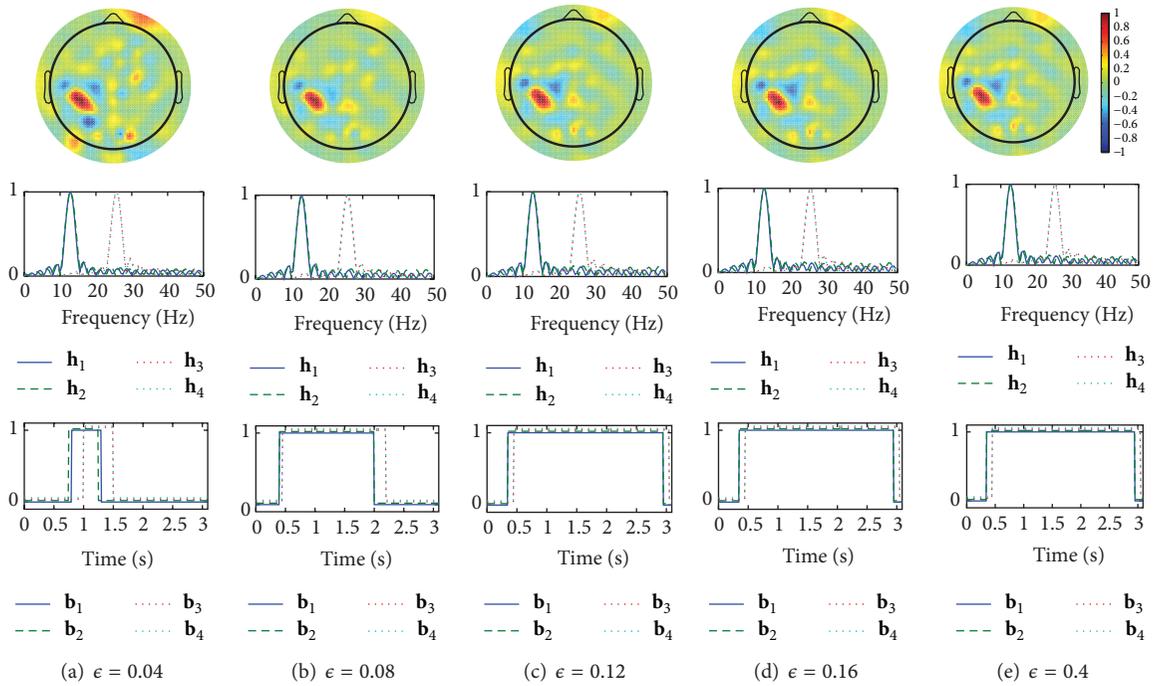


FIGURE 7: Examples of spatio-time-frequency patterns as the spatial weights, $\widehat{w}_1^{(1)}$ (top), the amplitude response of the filter, $\mathbf{h}_1, \dots, \mathbf{h}_4$ (center), and the time windows, $\mathbf{b}_1, \dots, \mathbf{b}_4$ (bottom), designed by the CSTFP method with various ϵ in the result of subject *aa*. The vertical axis represents normalized amplitude.

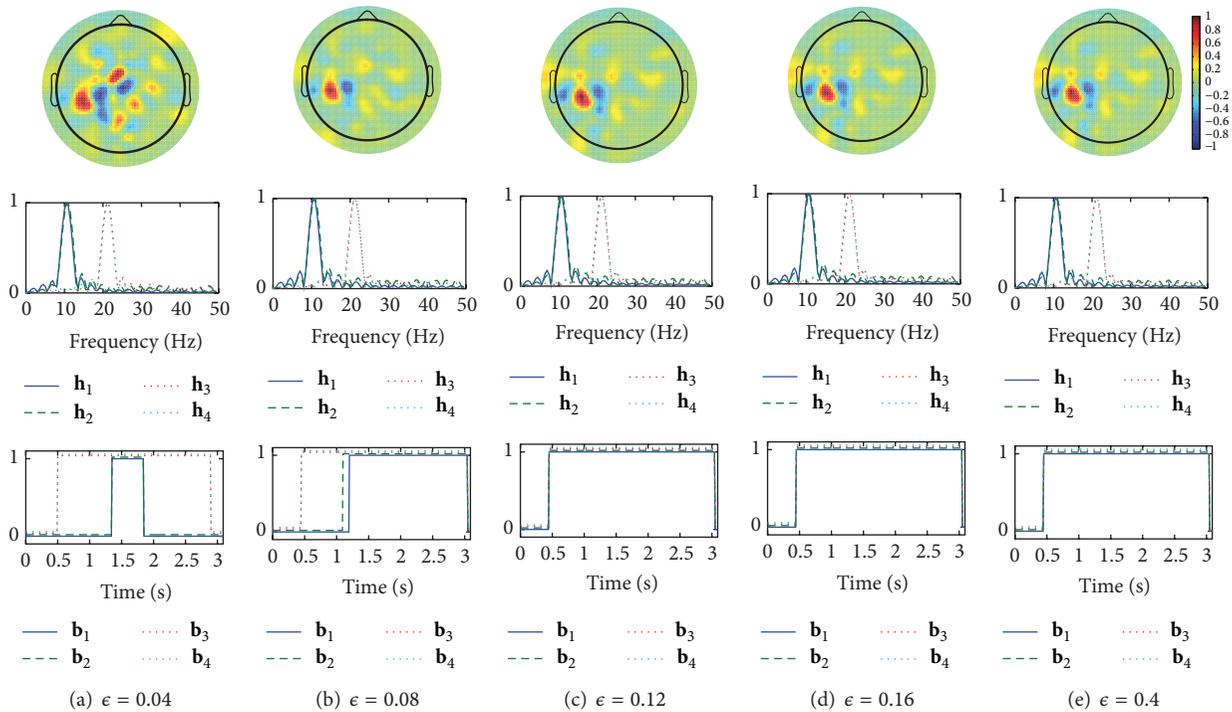


FIGURE 8: Examples of spatio-time-frequency patterns as the spatial weights, $\widehat{w}_1^{(1)}$ (top), the amplitude response of the filter, $\mathbf{h}_1, \dots, \mathbf{h}_4$ (center), and the time windows, $\mathbf{b}_1, \dots, \mathbf{b}_4$ (bottom), designed by the CSTFP method with various ϵ in the result of subject *av*. The vertical axis represents normalized amplitude.

BCI/BMI [26]. We would like to develop such methods by applying the proposed CSTFP algorithm to our future works.

Acknowledgments

This work is supported in part by KAKENHI, Grant-in-Aid for Scientific Research (B), 21360179 and 24360146, and Grant-in-Aid for JSPS Fellows, 24686.

References

- [1] G. Dornhege, J. R. Millan, T. Hinterberger, D. McFarland, and K.-R. Müller, Eds., *Toward Brain-Computer Interfacing*, The MIT Press, Cambridge, Mass, USA, 2007.
- [2] S. Sanei and J. Chambers, *EEG Signal Processing*, Wiley-Interscience, New York, NY, USA, 2007.
- [3] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.
- [4] D. J. McFarland and J. R. Wolpaw, "Brain-computer interface operation of robotic and prosthetic devices," *Computer*, vol. 41, no. 10, pp. 52–56, 2008.
- [5] C. Zhang, Y. Kimura, H. Higashi, and T. Tanaka, "A simple platform of brain-controlled mobile robot and its implementation by SSVEP," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN '12)*, pp. 1–7, 2012.
- [6] D. J. McFarland and J. R. Wolpaw, "Brain-computer interfaces for communication and control," *Communications of the ACM*, vol. 54, no. 5, pp. 60–66, 2011.
- [7] J. R. Wolpaw and D. J. McFarland, "Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 51, pp. 17849–17854, 2004.
- [8] J. J. Daly and J. R. Wolpaw, "Brain-computer interfaces in neurological rehabilitation," *The Lancet Neurology*, vol. 7, no. 11, pp. 1032–1043, 2008.
- [9] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," *Clinical Neurophysiology*, vol. 110, no. 5, pp. 787–798, 1999.
- [10] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 4, pp. 441–446, 2000.
- [11] S. Lemm, B. Blankertz, G. Curio, and K.-R. Müller, "Spatio-spectral filters for improving the classification of single trial EEG," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 9, pp. 1541–1548, 2005.
- [12] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, and K.-R. Müller, "Combined optimization of spatial and temporal filters for improving brain-computer interfacing," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 11, pp. 2274–2281, 2006.
- [13] R. Tomioka, G. Dornhege, G. Nolte, B. Blankertz, K. Aihara, and K. R. M. Müller, "Spectrally weighted common spatial pattern algorithm for single trial eeg classification," Tech. Rep. 40, Department of Mathematical Engineering, University of Tokyo, Tokyo, Japan, 2006.
- [14] H. Higashi and T. Tanaka, "Classification by weighting for spatio-frequency components of EEG signal during motor imagery," in *Proceedings of the 36th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '11)*, pp. 585–588, May 2011.
- [15] H. Higashi and T. Tanaka, "Simultaneous design of FIR filter banks and spatial patterns for EEG signal classification," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 4, pp. 1100–1110, 2013.
- [16] N. Tomida, H. Higashi, and T. Tanaka, "A joint tensor diagonalization approach to active data selection for EEG classification," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '13)*, pp. 983–987, 2013.
- [17] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter Bank Common Spatial Pattern (FBCSP) in brain-computer interface," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN '08)*, pp. 2390–2397, June 2008.
- [18] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Mutual information-based selection of optimal spatial-temporal patterns for single-trial EEG-based BCIs," *Pattern Recognition*, vol. 45, no. 6, pp. 2137–2144, 2012.
- [19] H. Lu, H.-L. Eng, C. Guan, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularized common spatial pattern with aggregation for EEG classification in small-sample setting," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 12, pp. 2936–2946, 2010.
- [20] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 2, pp. 355–362, 2011.
- [21] G. Der and I. J. Deary, "Age and sex differences in reaction time in adulthood: results from the United Kingdom health and lifestyle survey," *Psychology and Aging*, vol. 21, no. 1, pp. 62–73, 2006.
- [22] B. Blankertz, K.-R. Müller, D. J. Krusienski et al., "The BCI competition III: validating alternative approaches to actual BCI problems," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 2, pp. 153–159, 2006.
- [23] 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 Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 993–1002, 2004.
- [24] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Müller, and G. Curio, "The non-invasive Berlin brain-computer interface: fast acquisition of effective performance in untrained subjects," *NeuroImage*, vol. 37, no. 2, pp. 539–550, 2007.
- [25] S. Varma and R. Simon, "Bias in error estimation when using cross-validation for model selection," *BMC Bioinformatics*, vol. 7, no. 1, article 91, 2006.
- [26] J. D. R. Millán and J. Mouriño, "Asynchronous BCI and local neural classifiers: an overview of the adaptive brain interface project," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2, pp. 159–161, 2003.