Study of Classification Methods for EEG Signals During Motor-Imagery

Hiroshi Higashi

Department of Electrical and Electronic Engineering,

The Graduate School of Engineering,

Tokyo University of Agriculture and Technology

Superviser: Toshihisa Tanaka

Abstract

We propose novel methods for finding a frequency filter for classification of EEG in motor imagery based brain computer interface (MI-BCI). For MI-BCI, spatial weights for electrodes called common spatial pattern (CSP) are known effective in classification. To find spatial weights associated with motor imagery, EEG signals should be bandpass-filtered. However, the best filter depends on subjects and trials. This thesis solves the problem to find parameters of the filter, which is parameterized in the frequency or time domain. Finding the optimal parameters is formulated as a constraint minimum variance problem as a natural extension of CSP. Then, the spatial weight and the filter are sought by alternately solving the generalized eigenvalue problem, and the cost function monotonically decreases by the alternative optimization. In our experiment of classification for MI-BCI, the proposed methods outperform conventional methods in the classification accuracy for most of subjects.

Contents

Chapter 1

Introduction

Brain computer interface (BCI) is a challenging application of signal processing and neuroscience. A promising realization of BCI is so-called motor imagery based BCI (MI-BCI) [1, 2]. Basically, noninvasive measurement devices such as electroencephalogram (EEG), magnetoencephalogram (MEG), and functional magnetic response imaging (fMRI) are widely used to observe the brain activity of motor imagery. Among them, because of its simplicity and low cost, EEG is a practical measurement device for use in engineering applications. MI-BCI uses as features the mu rhythm, which disappears around motor cortex when a body moves. Left hand movement affects the right motor cortex in brain, and vise versa [1]. Therefore, by accurately capturing these changes from measured EEG signals in the presence of measurement noise and spontaneous components related other brain activities, the motor imagery can be used for BCI control [1, 2].

A promising method for extracting the brain activity for MI-BCI is the so-called common spatial pattern (CSP) [3, 4]. CSP gives a coefficient corre-

sponding to each electrode in a multichannel EEG measurement system. The coefficients, that are the spatial weights, are determined from learning data in such a way that the variance of the extracted signals differ between two classes (e.g. left and right hand movement imageries). Although together with the spatial weights, an input EEG signal is successfully classified, for the effective implementation, the observed signal should be bandpass-filtered to extract frequency components associated with motor imagery activities [3]. For the bandpass filtering, a passband of 7–30Hz is widely chosen in MI-BCI, because mu and beta rhythms which are associated with motor imagery, are included in this band [1]. However, the optimum frequency band for classification is highly dependent on users and measurement environments. Therefore, for accurate classification, it is necessary to find the best frequency band or the filter for each user and/or dataset. To this end, it would be simple to search the parameters by using cross validation (CV) over learning samples. In this method, we search by CV for the filter that gives the best classification accuracy out of a finite number of candidates which have various passbands. Hence performance in classification depends on the choice of candidate filters, and the selected filter is not always "optimal." Moreover, a large size of the candidate set leads to higher computational cost. Therefore, it is crucial to establish a finding the optimal filter for classification in MI-BCI.

Recently, several approaches to this problem have been proposed [5–7]. Common spatio-spatial pattern (CSSP) [5] and common sparse spectral spatial pattern (CSSSP) [6] are the methods that obtain coefficients of a finite impulse response (FIR) filter by applying CSP to the combination of observed signals with the time-delayed signals. However, CSSP provides very

poor frequency selectivity due to the limitation of a single delay. Moreover, CSSSP needs the computationally expensive optimization because the optimization problem for the filter should be solved by the gradient method and an extensive parameter tuning is needed. Spectrally weighted CSP (SPEC-CSP) uses an iterative procedure of optimizing spatial weights and a filter [7], where the spatial weights are optimized by CSP and the filter parameterized by weights for the spectrum is found by an optimization problem based on Fisher's criterion. These two optimization problems are alternately solved, however this alternative iteration is not guaranteed to be converged, because the cost functions for these two problems are different. In summary, the these methods have problem on optimization processes such as computational inefficiency and no guarantee of convergence.

This thesis proposes novel methods for finding the optimal parameters of the filter. The filter is parameterized by i) parameterization as weights for the spectrum in the frequency domain, ii) parameterization as a temporal filter in the FIR filter in the time domain. In both parameterization, the optimal parameters, that is the spatial weight and the filter, are sought by alternately minimizing the proposed variance cost, which is a natural extension of the cost function for CSP. Unlike SPEC-CSP, it is guaranteed that the cost function monotonically decreases by iterations, because the both of the proposed methods optimize two optimization parameters in the single cost function. Both of the optimization problems can be reduced to generalized eigenvalue problems, which can be solved by well-established optimization methods unlike gradient method that CSSSP uses.

This thesis is organized as follows. Chapter 3 reviews the CSP method

and the CSP based filter design methods (CSSP and SPEC-CSP). Next, we propose the spectral weight design method in Sec. 4.1 and the FIR filter design method in Sec. 4.2. Chapter 5 presents experimental results of classification of EEG signals during motor imageries to show the effectiveness of the proposed methods. Finally, the conclusion of this thesis is presented in chapter 6.

1.1 Notations

The following notation and mathematical operations are used in the thesis.

- Scalar is denoted by a italic letter e.g. x or X.
- Vector is denoted by a italic bold lower-case letter e.g. $\boldsymbol{x}$. $\boldsymbol{x}[i]$ denotes the ith entry of $\boldsymbol{x}$, where i is a positive integer.
- Matrix is denoted by a italic bold uppercase letter e.g. $\boldsymbol{X}$. $[\boldsymbol{X}]_{i,j}$ denotes the entry in ith row and jth column, where i and j are positive integers.

Chapter 2

Brain Computer Interfaces

A brain computer interface (BCI) is a communication system in which messages or commands that an individual sends to the external world do not pass through the brain's normal output pathway of peripheral nerves and muscles [2]. Like any communication or control system, a BCI has input (e.g. electrophysiological activity from the user), output (i.e. device commands), components that translate input into output, and a protocol that determines the onset, offset, and timing of operation. Figure 2.1 shows these elements and their principal interactions. In Fig. 2.1, the brain activities are measured by electroencephalogram (EEG). Sec. 2.1 shows other measure equipments for observing brain activity. Moreover, the BCI systems are sorted by the difference of the external stimuli or mental tasks [2,8] such as P300 [19], steady state visual evoked response (SSVEP) [13], and motor imagery (MI) [4] based BCI. In this study, we focus on MI-BCIs. In Sec. 2.2, we introduce the brain activities related to motor imagery.

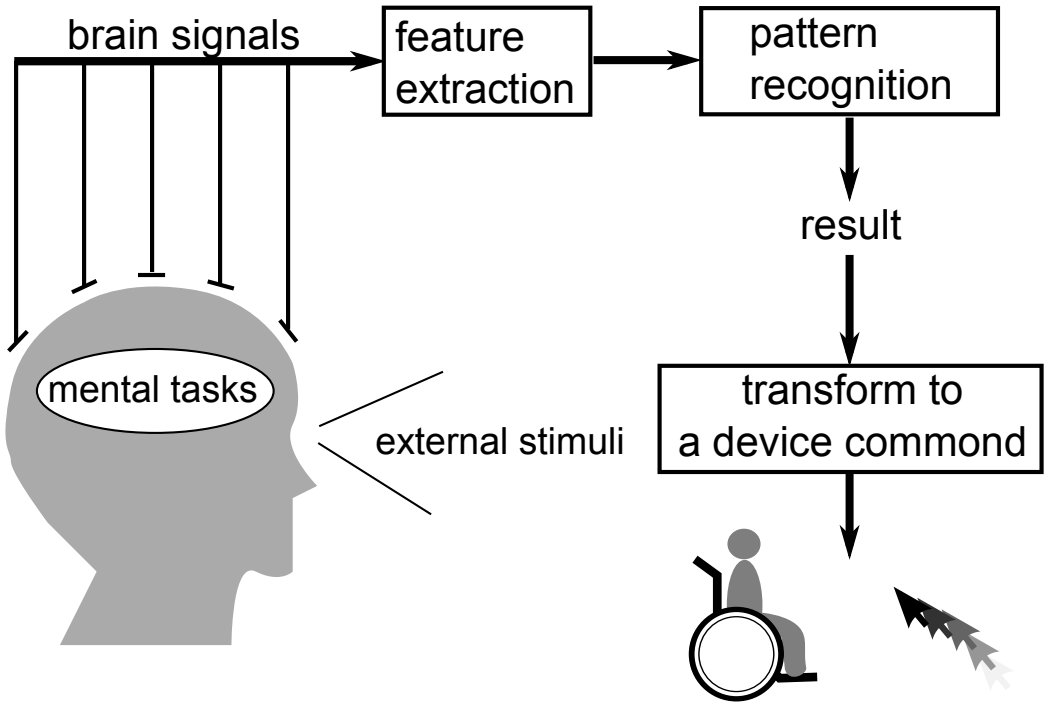


Figure 2.1: Basic design and operation of a BCI system.

2.1 Brain Activity Measures

A variety of methods for monitoring brain activity might serve as a BCI. These include, beside electroencephalography (EEG) and more invasive electrophysiological methods, magnetoencephlography (MEG), positron emission tomography (PET), functional magnetic resonance imaging (fMRI), and optical imaging. However, MEG, PET, fMRI, and optical imaging are still technically demanding and expensive. Furthermore, PET, fMRI, and optical imaging, which depend on blood flow, have long time constants and thus are less amenable to rapid communication. At present, only EEG and related methods, which have relatively short time constants, can function in most environments, and require relatively simple and inexpensive equipment, offer

the possibility of a new non-muscular communication and control channel, a practical BCI.

More recent EEG systems consist of a number of electrodes, a set of differential amplifiers (one for each channel). In an EEG measurement system, the electrodes and their proper function are crucial for acquiring high quality data. Different types of electrodes are often used in the EEG recording systems, such as:

- disposable (gel-less, and pre-gelled types);
- reusable disc electrodes; (made of gold, silver, stainless steel, or tin);
- headbands and electrode caps;
- saline-based electrodes;
- needle electrodes.

For multichannel recordings with a number of electrodes, the international 10/20 system has stood as the de-facto standard of electrode placement used in EEG [1, 8]. This system describes head surface locations via relative distances between cranial landmarks over the head surface. For higher density electrode settings, the extended international 10/20 system [9], the 10/10 system [10], and 10/5 system [11] have been used. Figure 2.2 shows the position of electrodes in international 10/20 system and 10/10 system.

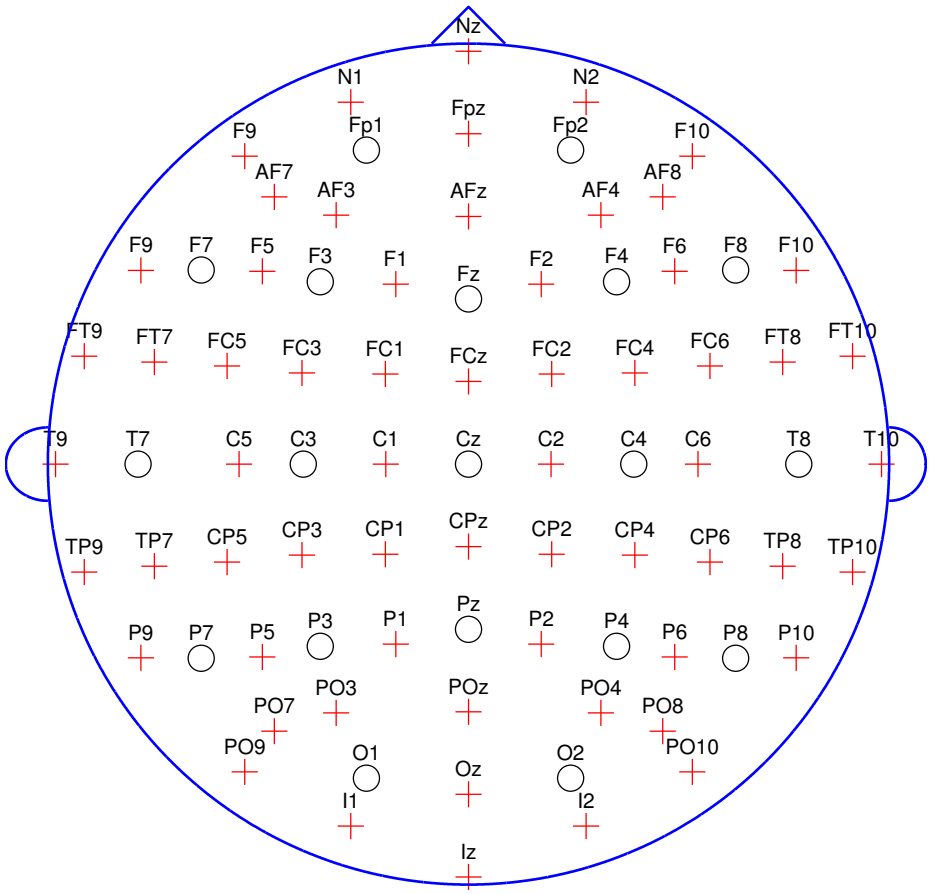


Figure 2.2: The electrodes placements of international 10/20 and 10/10 systems. The circles indicate the position of international 10/20 system. The circles and the cross marks indicate the position of international 10/10 system.

2.2 Motor-Imagery Based BCI

In awake people, primary sensory of motor cortical areas often display 8–12 Hz EEG activity when they are not engaged in processing sensory input or producing motor output. This idling activity, called mu rhythm when focused over somatosensory or motor cortex and visual alpha rhythm when focused over visual cortex, is thought to be produced by thalamocortical circuits. Unlike the visual alpha rhythm which is obvious in most normal people, the mu rhythm was until quite recently found only in minority. However, computer-based analyses reveal the mu rhythm in most adults. Such analyses also show that mu rhythm activity comprises a variety of different 8–12 Hz rhythms, distinguished from each other by location frequency, and/or relationship to concurrent sensory input or motor output. These mu rhythms are usually associated with 18–26 Hz beta rhythms. While some beta rhythms are harmonics of mu rhythms, some are separable from them by topography and/or timing, and thus are independent EEG features.

Several factors suggest that mu and/or beta rhythms could be good signal features for EEG-based communication. They are associated with those cortical areas most directly connected to the brain's normal motor output channels. Movement or preparation for movement is typically accompanied by a decrease in mu and beta rhythms, particularly contralateral to the movement. This decrease has been called event-related desynchronization (ERD) [12]. Its opposite, rhythm increase called event-related synchronization (ERS) occurs after movement and with relaxation. Furthermore, and most relevant for BCI use, ERD and ERS do not require actual movement,

they occur also with motor imagery (i.e. imagined movement).

Since the mid-1980s, several BCIs which use mu/beta rhythms feature have been developed. Wolpow et al. realized two-dimensional control with cursor movement on a computer screen to eight target [14]. In 2003, Neuper et al. (in cooperation with the Tübingen group) reported results of a patient with infantile cerebral paresis, who was trained over a period of several months with the MI-BCI [15]. The patient was trained with a tow-target task and was provided with feedback of ERD. Eventually, the targets were replaced by letters and the patient could spell with the system, using a so-called virtual keyboard [18]. The virtual keyboard allows patients to select letters in successive binary steps. A different set of letters is presented at the top and the bottom of the screen and the patient has to move the cursor into that half of the screen that contains the desired letter. The spelling rate varied between 0.2 and 2.5 letters per minute. Besides communication, the Graz group implemented MI-BCI mediated neuroprosthesis control in tow exemplary patients. First, a tetraplegic patient, whose residual muscle activity of the upper limbs was restricted to the left biceps after spinal cord injury, learned to open and close his hand with the aids of an orthosis that reacted upon changes in the ERD. The authors report an accuracy rate of almost 100%. In second study with the same patient, grasping movement was realized via MI-BCI-controlled functional electrical stimulation.

In summary, MI-BCIs have been successfully tested in amyotrophic lateral sclerosis (ALS), cerebral palsy, and spinal cord injury patients and may provide communication or restoration of lost motor function. Additionally, MI-BCIs have been tested with healthy people and many researches have

reported that mu and beta rhythms are good signal features for realizing BCI [2, 8, 14, 20, 21].

Chapter 3

CSP and Filter Design Methods

CSP is an effective method for the feature extraction and classification in two class MI-BCI. In this chapter, we review a basic CSP method [3, 4] in Sec. 3.1. Additionally, CSSP [5] and SPEC-CSP [7], which parameterize the bandpass filter, are reviewed in Sec. 3.2 and 3.3, respectively.

3.1 Common Spatial Pattern (CSP)

Let $\boldsymbol{X} \in \mathbb{R}^{M \times N}$ be an observed signal, where M is the number of channels and N is the number of samples. CSP finds a spatial weight vector, $\boldsymbol{w} \in \mathbb{R}^{M}$, in such a way that the variance of a signal extracted by linear combination of $\boldsymbol{X}$ is minimized in a class [3,4]. In BCI application, we do not directly use $\boldsymbol{X}$, but use the filtered signal described as $\hat{\boldsymbol{X}} = \mathcal{H}(\boldsymbol{X})$ in CSP, where $\mathcal{H}$ is a bandpass filter which passes the frequency band related to brain activity of

motor imagery. Denote the components of $\hat{\boldsymbol{X}}$ by $\hat{\boldsymbol{X}} = [\hat{\boldsymbol{x}}_1, \ldots, \hat{\boldsymbol{x}}_N]$, where $\hat{\boldsymbol{x}}_n \in \mathbb{R}^M$ and n is the time index. The time average of the observed signal is given by $\boldsymbol{\mu} = (1/N)\sum_{n=1}^{N} \hat{\boldsymbol{x}}_n$. Then, the time variance of the extracted signal of $\hat{\boldsymbol{X}}$ is given by

$$\sigma^2(\boldsymbol{X}, \boldsymbol{w}) = \frac{1}{N}\sum_{n=1}^{N} |\boldsymbol{w}^T(\hat{\boldsymbol{x}}_n - \boldsymbol{\mu})|^2, \tag{3.1}$$

where $\cdot^T$ denotes the transpose of a vector or matrix.

We assume that sets of the learning data, $\mathcal{C}_1$ and $\mathcal{C}_2$, where $\mathcal{C}_d$ contains the signals belonging to class d, $d \in \{1, 2\}$ is a class label, and $\mathcal{C}_1 \cap \mathcal{C}_2 = \phi$. CSP finds the weight vector that minimizes the intra-class variance in $\mathcal{C}_c$ under the normalization of samples, where c is a class label. More specifically, for class, c, fixed, CSP finds $\boldsymbol{w}_c$ by solving the following optimization problem [3, 4];

$$\begin{aligned} \min_{\boldsymbol{w}} \quad & E_{\boldsymbol{X}\in\mathcal{C}_c}[\sigma^2(\boldsymbol{X}, \boldsymbol{w})] \\ \text{subject to} \quad & E_{\boldsymbol{X}\in\mathcal{C}_1}[\sigma^2(\boldsymbol{X}, \boldsymbol{w})] + E_{\boldsymbol{X}\in\mathcal{C}_2}[\sigma^2(\boldsymbol{X}, \boldsymbol{w})] = 1, \end{aligned} \tag{3.2}$$

where $E_{\boldsymbol{X}\in\mathcal{C}_d}[\cdot]$ denotes the expectation over $\mathcal{C}_d$. Then, (3.2) can be rewritten as

$$\begin{aligned} \min_{\boldsymbol{w}} \quad & \boldsymbol{w}^T\boldsymbol{\Sigma}_c\boldsymbol{w}, \\ \text{subject to} \quad & \boldsymbol{w}^T(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)\boldsymbol{w} = 1, \end{aligned} \tag{3.3}$$

where $\boldsymbol{\Sigma}_d$, $d = 1, 2$, are defined as

$$\boldsymbol{\Sigma}_d = E_{\boldsymbol{X}\in\mathcal{C}_d}\left[\frac{1}{N}\sum_{n=1}^{N}(\hat{\boldsymbol{x}}_n - \boldsymbol{\mu})(\hat{\boldsymbol{x}}_n - \boldsymbol{\mu})^T\right]. \tag{3.4}$$

The solution of (3.3) is given by the generalized eigenvector corresponding to the minimum generalized eigenvalue of the generalized eigenvalue problem described as

$$\boldsymbol{\Sigma}_c \boldsymbol{w} = \lambda(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)\boldsymbol{w}. \tag{3.5}$$

Though the solution of (3.3) is given by the eigenvector corresponding to the minimum eigenvalue in (3.5), we can use the other eigenvectors for classification [20]. The M eigenvectors can be obtained by solving (3.5) as $\hat{\boldsymbol{w}}_1, \ldots, \hat{\boldsymbol{w}}_M$, where $\hat{\boldsymbol{w}}_i$ is the eigenvector corresponding to the ith maximum eigenvalue of (3.5). We assume that the $2r$ eigenvectors are used for classification of unlabeled data, $\boldsymbol{X}$. Then we obtain the feature vector, $\boldsymbol{y} \in \mathbb{R}^{2r}$, from $\boldsymbol{X}$ defined as

$$\begin{aligned}\boldsymbol{y} =&[\sigma^2(\boldsymbol{X}, \hat{\boldsymbol{w}}_1), \ldots, \sigma^2(\boldsymbol{X}, \hat{\boldsymbol{w}}_r),\\ &\sigma^2(\boldsymbol{X}, \hat{\boldsymbol{w}}_{M-r+1}), \ldots, \sigma^2(\boldsymbol{X}, \hat{\boldsymbol{w}}_M)]^T.\end{aligned} \tag{3.6}$$

For classification, $\boldsymbol{y}$ is input to a classifier such as linear discriminant analysis (LDA) [23] and support vector machine (SVM) [24].

3.2 Common Spatio-Spectral Pattern (CSSP)

CSSP is a method where a weight vector is obtained by applying the combination of observed signals with time-delayed signals to CSP [5]. Let $\boldsymbol{X} \in \mathbb{R}^{M \times N}$ be an observed signal with M channels and N samples. Let $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ be the subsignals included in $\boldsymbol{X}$. The components of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are defined as $[\boldsymbol{X}_1]_{m,n} = [\boldsymbol{X}]_{m,n}$ and $[\boldsymbol{X}_2]_{m,n} = [\boldsymbol{X}]_{m,n+\tau}$, respectively, where

$m = 1, \ldots, M$, $n = 1, \ldots, N - \tau$, τ is a delay sample, and $[\cdot]_{i,j}$ denotes the entry in ith row and jth column of a matrix. Then the τ-delay embedded signal is defined as

$$\boldsymbol{X}^{\tau} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix} \in \mathbb{R}^{2M \times (N-\tau)}. \tag{3.7}$$

CSSP uses $\boldsymbol{X}^{\tau}$ to seek the spatial weight vector.

For classification, CSSP can give the feature vector as follows. We obtain $2M$ eigenvectors, $\hat{\boldsymbol{w}}_i \in \mathbb{R}^{2M}, i = 1, \ldots, 2M$, from (3.5) using $\boldsymbol{X}^{\tau}$. We compose the feature vector, $\boldsymbol{y} \in \mathbb{R}^{2r}$, in a way similar to (3.6).

In CSSP, the weight vector, $\hat{\boldsymbol{w}}_i$, can be regarded as a set of FIR filters corresponding to channels in the following way [5]. Let w^0 and w^{τ} be the weight coefficients for the original signal and the delayed signal corresponding to the jth channel in the ith weight vector given by $w^0 = \hat{\boldsymbol{w}}_i[j]$ and $w^{\tau} = \hat{\boldsymbol{w}}_i[j+M]$, where $\boldsymbol{a}[i]$ denotes the entry of $\boldsymbol{a}$. Then, the set, $\{w^0, \underbrace{0, \ldots, 0}_{\tau-1}, w^{\tau}\}$, is regarded as the coefficients of the FIR filter.

3.3 Spectrally Weighted CSP (SPEC-CSP)

SPEC-CSP uses an iterative procedure to achieve optimization of both of spatially weights and filters [7]. A filtered signal of an observed signal, $\boldsymbol{X}$, can be written by $\hat{\boldsymbol{X}} = \boldsymbol{X}\boldsymbol{F}_N\boldsymbol{B}\boldsymbol{F}_N^{-1}$ with the discrete Fourier transform matrix, $\boldsymbol{F}_N \in \mathbb{C}^{N \times N}$, defined as $[\boldsymbol{F}_N]_{k,l} = (1/\sqrt{N})e^{-2\pi(k-1)(l-1)/N}$, $k, l = 1, \ldots, N$, and a filter in frequency domain, $\boldsymbol{B}$, represented by a diagonal matrix, $\boldsymbol{B} = \mathrm{diag}(b_1, \ldots, b_N)$. In SPEC-CSP, the feature value from $\boldsymbol{X}$ is

defined with the spatial weigh vector, $\boldsymbol{w}$, and the spectral weight, $\boldsymbol{B}$, as

$$v(\boldsymbol{X}, \boldsymbol{w}, \boldsymbol{\beta}) = \boldsymbol{w}^T \boldsymbol{X} \boldsymbol{F}_N \boldsymbol{B} \boldsymbol{B}^T \boldsymbol{F}_N \boldsymbol{X}^T \boldsymbol{w} = \boldsymbol{w}^T \left(\sum_{k=1}^{N} b_k^2 \boldsymbol{V}_k \right) \boldsymbol{w} \tag{3.8}$$

where we define $\boldsymbol{G} = \boldsymbol{X}\boldsymbol{F} = [\boldsymbol{g}_1, \ldots, \boldsymbol{g}_N]$, $\boldsymbol{V}_k = \boldsymbol{g}_k \boldsymbol{g}_k^T$, and $\boldsymbol{\beta} = [b_1^2, \ldots, b_N^2]^T$. SPEC-CSP decides $2r$ weight vectors, $\hat{\boldsymbol{w}}_i$, and $2r$ filters, $\boldsymbol{\beta}_i$, by alternately optimizing with CSP for $\hat{\boldsymbol{w}}_i$ and the optimization problem based on Fisher's criterion for $\boldsymbol{\beta}_i$, where $i = 1, \ldots, 2r$.

In optimization for $\hat{\boldsymbol{w}}_i$, covariance matrices for each $\boldsymbol{\beta}_i$ are defined as $\boldsymbol{\Sigma}_d^i = E_{\boldsymbol{X} \in \mathcal{C}_d} \left[\sum_{k=1}^{N} \beta_{k,i} \boldsymbol{V}_k \right]$, $d = 1, 2$, where the elements of $\boldsymbol{\beta}_i$ are denoted by $\boldsymbol{\beta}_i = [\beta_{1,i}, \ldots, \beta_{N,i}]^T$, and $\mathcal{C}_d$ is the set of the learning data belonging to class d. Then we solve the generalized eigenvalue problem;

$$\boldsymbol{\Sigma}_c^i \boldsymbol{w} = \lambda \boldsymbol{\Sigma}_{\bar{c}}^i \boldsymbol{w}, \tag{3.9}$$

where c is an optional class label such that $c \in \{1, 2\}$ and $\bar{c}$ is other class label. Let $\bar{\boldsymbol{w}}_1^i, \ldots, \bar{\boldsymbol{w}}_M^i$ be the generalized eigenvectors corresponding to generalized eigenvalues, $\lambda_1^i, \ldots, \lambda_M^i$, a and b are chosen by $a = \mathrm{argmin}_{i=1,\ldots,r} \lambda_1^i$ and $b = \mathrm{argmax}_{i=1,\ldots,r} \lambda_M^i$, and then, the $2r$ spatial weight vectors are defined as $\hat{\boldsymbol{w}}_j = \bar{\boldsymbol{w}}_j^a, j = 1, \ldots, r$ and $\hat{\boldsymbol{w}}_j = \bar{\boldsymbol{w}}_{M-2r+j}^b, j = r+1, \ldots, 2r$.

Next, SPEC-CSP formulates the optimization problem for $\boldsymbol{\beta}_i$ as

$$\begin{aligned} \max_{\boldsymbol{\beta}_i} \quad & \frac{E_{\boldsymbol{X} \in \mathcal{C}_c}[s_i] - E_{\boldsymbol{X} \in \mathcal{C}_{\bar{c}}}[s_i]}{\sqrt{\sum_{d=1,2} E_{\boldsymbol{X} \in \mathcal{C}_d}[|s_i - E_{\boldsymbol{X} \in \mathcal{C}_d}[s_i]|^2]}}, \\ \text{subject to} \quad & \beta_{k,i} \geq 0, \ \forall k = 1, \ldots, N, \end{aligned} \tag{3.10}$$

where s_i is defined as $s_i = \sum_{k=1}^{N} \beta_{k,i} \hat{\boldsymbol{w}}_i^T \boldsymbol{V}_k \hat{\boldsymbol{w}}_i$. Here, we define $\gamma_k^i = E_{\boldsymbol{X} \in \mathcal{C}_c}[\hat{\boldsymbol{w}}_i^T \boldsymbol{V}_k \hat{\boldsymbol{w}}_i] -$

$E_{\boldsymbol{X}\in\mathcal{C}_{\bar{c}}}[\hat{\boldsymbol{w}}_i^T\boldsymbol{V}_k\hat{\boldsymbol{w}}_i]$ and $\zeta_k^i = \sum_{d=1,2} E_{\boldsymbol{X}\in\mathcal{C}_d}[|\hat{\boldsymbol{w}}_i^T\boldsymbol{V}_k\hat{\boldsymbol{w}}_i - E_{\boldsymbol{X}\in\mathcal{C}_d}[\hat{\boldsymbol{w}}_i^T\boldsymbol{V}_k\hat{\boldsymbol{w}}_i]|^2]$. Because $\boldsymbol{s}_i$ is linear with respect to $\beta_{k,i}, k = 1, \ldots, N$, and we assume that the signal is a stationary Gaussian process where the frequency components are independent to each other, the solution of (3.10) is

$$\beta_{k,i} \propto \begin{cases} \gamma_k^i/\zeta_k^i & \gamma_k^i \geq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{3.11}$$

Because the norm of $\boldsymbol{\beta}_i$ cannot be determined by (3.11), we normalize $\boldsymbol{\beta}_i$ so that they sum to one.

SPEC-CSP optimizes $\hat{\boldsymbol{w}}_i$ and $\boldsymbol{\beta}_i$ by iteration of the optimization problems, (3.9) and (3.10). After $\hat{\boldsymbol{w}}_i$ and $\boldsymbol{\beta}_i$ are obtained, The feature vector, $\boldsymbol{y} \in \mathbb{R}^{2r}$ from an observed signal, $\boldsymbol{X}$, is defined as

$$\boldsymbol{y} = [v(\boldsymbol{X}, \hat{\boldsymbol{w}}_1, \boldsymbol{\beta}_1), \ldots, v(\boldsymbol{X}, \hat{\boldsymbol{w}}_{2r}, \boldsymbol{\beta}_{2r})]^T. \tag{3.12}$$

Chapter 4

Proposed Filter Design Methods

Though CSP is effective in MI-BCI, the classification accuracy depends on the choice of the bandpass filter $\mathcal{H}$ [3]. In this chapter, we present the methods for designing the filter. The proposed methods optimize the filter and the spatial weight by using the cost extended from the CSP cost. We introduce a design method by parameterizing the filter as weights for the spectrum of an EEG signal in Sec. 4.1. We also introduce a design method by parameterizing the filter with FIR in the time domain in Sec. 4.2.

4.1 Spectral Weight Design Method

4.1.1 Feature Extraction

Let $\boldsymbol{X} \in \mathbb{R}^{M \times N}$ be a matrix consisting of an observed signal with M channels and N samples, and denote the components of $\boldsymbol{X}$ by $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N]$,

$\boldsymbol{x}_n \in \mathbb{R}^M$. Without loss of generality, we assume that $\boldsymbol{x}_n$ is zero mean. This can be achieved by replacing $\boldsymbol{x}_n$ by $\boldsymbol{x}_n - (1/N)\sum_{n=1}^{N} \boldsymbol{x}_n$. Let $\tilde{\boldsymbol{X}}_l$, $l = 1, \ldots, L$, be subsignals or frames defined as follows. In component-wise, frame signal, $\tilde{\boldsymbol{X}}_l \in \mathbb{R}^{M \times 2N'}$, with the frame length of $2N'$ and the frame shift of D is defined as

$$[\tilde{\boldsymbol{X}}_l]_{m,n} = [\boldsymbol{X}]_{m,(l-1)D+n}. \tag{4.1}$$

Then, the number of frames is given by $L = \mathrm{fix}((N - 2N')/D + 1)$, where $\mathrm{fix}(\cdot)$ is the rounding operator. Moreover, we define $\tilde{\boldsymbol{Y}}_l \in \mathbb{R}^{M \times N'}$ as

$$[\tilde{\boldsymbol{Y}}_l]_{m,n} = |[\tilde{\boldsymbol{X}}_l \boldsymbol{F}_{2N'}]_{m,n}|, \tag{4.2}$$

where $m = 1, \ldots, M$ and $n = 1, \ldots, N'$. Because $\tilde{\boldsymbol{X}}_l$ is real-valued, rows of $|\tilde{\boldsymbol{X}}_l \boldsymbol{F}_{N'}|$ are symmetric. Hence, we can leave out duplicated components, that is, components corresponding to the negative frequency. We introduce $\boldsymbol{h} \in \mathbb{R}^{N'}$ to weight the spectra of the observed signal, and $\boldsymbol{h}$ acts as a filter. The filtered spectra, $\tilde{\boldsymbol{Y}}_l \boldsymbol{h}$, is used for finding spatial weights in the same manner as CSP. More specifically, as done in CSP, the feature value is defined as

$$z(\boldsymbol{X}, \boldsymbol{w}, \boldsymbol{h}) = \sum_{l=1}^{L} |\boldsymbol{w}^T \tilde{\boldsymbol{Y}}_l \boldsymbol{h}|^2, \tag{4.3}$$

where $\boldsymbol{w} \in \mathbb{R}^M$ represents the spatial weight vector.

4.1.2 Optimization

The underlying idea behind the proposed method is to find both $\boldsymbol{w}$ and $\boldsymbol{h}$ for class c that minimize z with respect to $\boldsymbol{X} \in \mathcal{C}_c$ under the normalization

of samples. For c fixed, $\boldsymbol{w}_c$ and $\boldsymbol{h}_c$ is obtained by solving the optimization problem defined as

$$\begin{aligned} \min_{\boldsymbol{w},\boldsymbol{h}} \quad & f(\boldsymbol{w},\boldsymbol{h}) = E_{\boldsymbol{X}\in\mathcal{C}_c}[z(\boldsymbol{X},\boldsymbol{w},\boldsymbol{h})] \\ \text{subject to} \quad & \nu(\boldsymbol{w},\boldsymbol{h}) = \sum_{d=1,2} E_{\boldsymbol{X}\in\mathcal{C}_d}[z(\boldsymbol{X},\boldsymbol{w},\boldsymbol{h})] = 1. \end{aligned} \tag{4.4}$$

Since we cannot seek for $\boldsymbol{w}$ and $\boldsymbol{h}$ simultaneously, we derive two subproblems that separately find $\boldsymbol{w}$ and $\boldsymbol{h}$.

The subproblem to optimize $\boldsymbol{w}$ while fixing $\boldsymbol{h}$ is as follows. z can be expanded to

$$z(\boldsymbol{X},\boldsymbol{w},\boldsymbol{h}) = \boldsymbol{w}^T \left(\sum_{l=1}^{L} \tilde{\boldsymbol{Y}}_l \boldsymbol{h}\boldsymbol{h}^T \tilde{\boldsymbol{Y}}_l^T \right) \boldsymbol{w}, \tag{4.5}$$

and we define

$$\boldsymbol{R}_d(\boldsymbol{h}) = E_{\boldsymbol{X}\in\mathcal{C}_d} \left[\sum_{l=1}^{L} \tilde{\boldsymbol{Y}}_l \boldsymbol{h}\boldsymbol{h}^T \tilde{\boldsymbol{Y}}_l^T \right], \quad d = 1, 2. \tag{4.6}$$

Then, (4.4) can be rewritten as

$$\begin{aligned} \min_{\boldsymbol{w}} \quad & f_1(\boldsymbol{w}|\boldsymbol{h}) = \boldsymbol{w}^T \boldsymbol{R}_c(\boldsymbol{h}) \boldsymbol{w}, \\ \text{subject to} \quad & \nu_1(\boldsymbol{w}|\boldsymbol{h}) = \boldsymbol{w}^T (\boldsymbol{R}_1(\boldsymbol{h}) + \boldsymbol{R}_2(\boldsymbol{h})) \boldsymbol{w} = 1. \end{aligned} \tag{4.7}$$

The solution of (4.7) is given by the generalized eigenvector corresponding to the minimal generalized eigenvalue of the generalized eigenvalue problem described as

$$\boldsymbol{R}_c(\boldsymbol{h}) \boldsymbol{w} = \lambda (\boldsymbol{R}_1(\boldsymbol{h}) + \boldsymbol{R}_2(\boldsymbol{h})) \boldsymbol{w}. \tag{4.8}$$

The subproblem to optimize $\boldsymbol{h}$ while fixing $\boldsymbol{w}$ is as follows. z can be

expanded to

$$z(\boldsymbol{X}, \boldsymbol{w}, \boldsymbol{h}) = \boldsymbol{h}^T \left(\sum_{l=1}^{L} \tilde{\boldsymbol{Y}}_l^T \boldsymbol{w} \boldsymbol{w}^T \tilde{\boldsymbol{Y}}_l \right) \boldsymbol{h}, \tag{4.9}$$

and we define

$$\boldsymbol{Q}_d(\boldsymbol{w}) = E_{\boldsymbol{X} \in \mathcal{C}_c} \left[\sum_{l=1}^{L} \tilde{\boldsymbol{Y}}_l^T \boldsymbol{w} \boldsymbol{w}^T \tilde{\boldsymbol{Y}}_l \right], \quad d = 1, 2. \tag{4.10}$$

Then (4.4) can be rewritten as

$$\begin{aligned} \min_{\boldsymbol{h}} \quad & f_2(\boldsymbol{h}|\boldsymbol{w}) = \boldsymbol{h}^T \boldsymbol{Q}_c(\boldsymbol{w}) \boldsymbol{h}, \\ \text{subject to} \quad & \nu_2(\boldsymbol{h}|\boldsymbol{w}) = \boldsymbol{h}^T (\boldsymbol{Q}_1(\boldsymbol{w}) + \boldsymbol{Q}_2(\boldsymbol{w})) \boldsymbol{h} = 1. \end{aligned} \tag{4.11}$$

The solution of (4.11) is given by the generalized eigenvector corresponding to the minimal generalized eigenvalue of the generalized eigenvalue problem described as

$$\boldsymbol{Q}_c(\boldsymbol{w}) \boldsymbol{h} = \lambda (\boldsymbol{Q}_1(\boldsymbol{w}) + \boldsymbol{Q}_2(\boldsymbol{w})) \boldsymbol{h}. \tag{4.12}$$

The iteration of the alternative update monotonically decreases $f(\boldsymbol{w}, \boldsymbol{h})$, and leads to a local minimum solution. In implementation, we initialize $\boldsymbol{h}$ as the N'-dimensional vector whose elements are all 1, and start the alternative update from updating $\boldsymbol{w}$. The iteration is stopped, when $f(\boldsymbol{w}, \boldsymbol{h})$ reaches a certain criterion. For example, we stop iteration when $f^{(k-1)} - f^{(k)} \leq \epsilon$, where $f^{(k)}$ is the value of $f(\boldsymbol{w}, \boldsymbol{h})$ in the kth iteration, and $\epsilon > 0$ is a minimum error of the cost between successive iterations.

4.1.3 Feature Vector Definition

The following method can be applied to classification of unlabeled data, $\boldsymbol{X}$. First, we choose a class label, g, in $\{1, 2\}$, and then we obtain $\boldsymbol{w}_g$ and $\boldsymbol{h}_g$ by (4.4) with the learning dataset. Next, we calculate $\boldsymbol{R}_d(\boldsymbol{h}_g)$, $d = 1, 2$ by (4.6), and solve (4.8). Then, we obtain the M eigenvectors as $\hat{\boldsymbol{w}}_1, \ldots, \hat{\boldsymbol{w}}_M$, where $\hat{\boldsymbol{w}}_i$ is the eigenvector corresponding to the ith minimum eigenvalue of (4.8). We assume that the $2r$ eigenvectors are used for classification. Then the feature vector $\boldsymbol{y} \in \mathbb{R}^{2r}$ is defined as

$$\begin{aligned}\boldsymbol{y} =&[z(\boldsymbol{X}, \hat{\boldsymbol{w}}_1, \boldsymbol{h}_g), \ldots, z(\boldsymbol{X}, \hat{\boldsymbol{w}}_r, \boldsymbol{h}_g),\\ &z(\boldsymbol{X}, \hat{\boldsymbol{w}}_{M-r+1}, \boldsymbol{h}_g)), \ldots, z(\boldsymbol{X}, \hat{\boldsymbol{w}}_M, \boldsymbol{h}_g)]^T. \end{aligned} \tag{4.13}$$

As well as the case of CSP, we input $\boldsymbol{y}$ to a classifier for classification.

4.2 FIR Filter Design Method

We establish the other approach for designing a filter when the filter is given as a standard FIR filter.

4.2.1 Feature Extraction

Let $\boldsymbol{X} \in \mathbb{R}^{M \times N}$ be a matrix consisting of an observed signal with M channels and N samples, and denote the components of $\boldsymbol{X}$ by $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N]$, $\boldsymbol{x}_n \in \mathbb{R}^M$. Let $\theta_p, p = 1, \ldots, P$ be the filter coefficients where P is the order of the filter, and the filtered signal of $\boldsymbol{X}$ denoted as $\hat{\boldsymbol{X}} = [\hat{\boldsymbol{x}}_1, \ldots, \hat{\boldsymbol{x}}_K]$, can

be defined for $n = 1, \ldots, K$, as

$$\hat{\boldsymbol{x}}_n = \sum_{p=1}^{P} \theta_p \boldsymbol{x}_{n+P-p}, \tag{4.14}$$

where $K = N - P + 1$. The variance of the extracted signal of CSP given in (3.1) is given with the spatial weight, $\boldsymbol{w}$, and the FIR filter, $\boldsymbol{\theta}$, by

$$\rho(\boldsymbol{X}, \boldsymbol{w}, \boldsymbol{\theta}) = \frac{1}{K} \sum_{n=1}^{K} \left| \boldsymbol{w}^T \sum_{p=1}^{P} \theta_k \boldsymbol{x}_{n+P-p} - \boldsymbol{\mu} \right|^2, \tag{4.15}$$

where $\boldsymbol{\theta}$ is the vector of the filter coefficients defined as $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_P]^T$, and $\boldsymbol{\mu}$ is the time average given by $\boldsymbol{\mu} = (1/K) \sum_{n=1}^{K} \boldsymbol{w}^T \sum_{p=1}^{P} \theta_k \boldsymbol{x}_{n+P-p}$. We define $\boldsymbol{A}_n, n = 1, \ldots, K$, whose elements are from $\boldsymbol{X}$ defined as

$$[\boldsymbol{A}_n]_{m,p} = [\boldsymbol{X}]_{m,n+P-p+1}, \tag{4.16}$$

where $m = 1, \ldots, M$, $p = 1, \ldots, P$. Therefore, (4.15) can be modified to

$$\begin{aligned} \rho(\boldsymbol{X}, \boldsymbol{w}, \boldsymbol{\theta}) &= \frac{1}{K} \sum_{n=1}^{K} \left| \boldsymbol{w}^T \boldsymbol{A}_n \boldsymbol{\theta} - \frac{1}{K} \sum_{m=1}^{K} \boldsymbol{w}^T \boldsymbol{A}_m \boldsymbol{\theta} \right|^2 \\ &= \frac{1}{K} \sum_{n=1}^{K} \left| \boldsymbol{w}^T \hat{\boldsymbol{A}}_n \boldsymbol{\theta} \right|^2, \end{aligned} \tag{4.17}$$

where $\hat{\boldsymbol{A}}_n$ is defined as $\hat{\boldsymbol{A}}_n = \boldsymbol{A}_n - (1/K) \sum_{m=1}^{K} \boldsymbol{A}_m$.

4.2.2 Optimization

As well as CSP, the underlying idea behind the proposed method is to find both $\boldsymbol{w}$ and $\boldsymbol{\theta}$ that minimize ρ with respect to $\boldsymbol{X} \in \mathcal{C}_c$ under the normal-

ization samples. For c fixed, $\boldsymbol{w}_c$ and $\boldsymbol{\theta}_c$ is obtained by solving the following optimization problem:

$$\begin{aligned} \min_{\boldsymbol{w},\boldsymbol{\theta}} \quad & g(\boldsymbol{w},\boldsymbol{\theta}) = E_{\boldsymbol{X}\in\mathcal{C}_c}[\rho(\boldsymbol{X},\boldsymbol{w},\boldsymbol{\theta})] \\ \text{subject to} \quad & \zeta(\boldsymbol{w},\boldsymbol{\theta}) = \sum_{d=1,2} E_{\boldsymbol{X}\in\mathcal{C}_d}[\rho(\boldsymbol{X},\boldsymbol{w},\boldsymbol{\theta})] = 1. \end{aligned} \tag{4.18}$$

As we have done in Sec. 4.1, we derive two subproblems that separately find $\boldsymbol{w}$ and $\boldsymbol{\theta}$.

The subproblem to optimize $\boldsymbol{w}$ while fixing $\boldsymbol{\theta}$ is as follows. $\rho(\boldsymbol{X},\boldsymbol{w},\boldsymbol{\theta})$ is can be written as

$$\rho(\boldsymbol{X},\boldsymbol{w},\boldsymbol{\theta}) = \boldsymbol{w}^T \left(\frac{1}{K} \sum_{n=1}^{K} \hat{\boldsymbol{A}}_n \boldsymbol{\theta}\boldsymbol{\theta}^T \hat{\boldsymbol{A}}_n^T \right) \boldsymbol{w}, \tag{4.19}$$

and we define

$$\boldsymbol{\Phi}_d(\boldsymbol{\theta}) = E_{\boldsymbol{X}\in C_d} \left[\frac{1}{K} \sum_{n=1}^{K} \hat{\boldsymbol{A}}_n \boldsymbol{\theta}\boldsymbol{\theta}^T \hat{\boldsymbol{A}}_n^T \right], \quad d = 1, 2. \tag{4.20}$$

Then (4.18) can be written as

$$\begin{aligned} \min_{\boldsymbol{w}} \quad & g_1(\boldsymbol{w}|\boldsymbol{\theta}) = \boldsymbol{w}^T \boldsymbol{\Phi}_c(\boldsymbol{\theta}) \boldsymbol{w} \\ \text{subject to} \quad & \xi_1(\boldsymbol{w}|\boldsymbol{\theta}) = \boldsymbol{w}^T (\boldsymbol{\Phi}_1(\boldsymbol{\theta}) + \boldsymbol{\Phi}_2(\boldsymbol{\theta})) \boldsymbol{w} = 1. \end{aligned} \tag{4.21}$$

The solution of (4.21) is given by the generalized eigenvector corresponding to the minimal generalized eigenvalue of the generalized eigenvalue problem described as

$$\boldsymbol{\Phi}_c(\boldsymbol{\theta}) \boldsymbol{w} = \lambda (\boldsymbol{\Phi}_1(\boldsymbol{\theta}) + \boldsymbol{\Phi}_2(\boldsymbol{\theta})) \boldsymbol{w}. \tag{4.22}$$

The subproblem to optimize $\boldsymbol{\theta}$ while fixing $\boldsymbol{w}$ is as follows. $\rho(\boldsymbol{X}, \boldsymbol{w}, \boldsymbol{\theta})$ is can be written as

$$\rho(\boldsymbol{X}, \boldsymbol{w}, \boldsymbol{\theta}) = \boldsymbol{\theta}^T \left(\frac{1}{K} \sum_{n=1}^{K} \hat{\boldsymbol{A}}_n^T \boldsymbol{w} \boldsymbol{w}^T \hat{\boldsymbol{A}}_n \right) \boldsymbol{\theta}, \tag{4.23}$$

and we define

$$\boldsymbol{\Psi}_d(\boldsymbol{w}) = E_{\boldsymbol{X} \in C_d} \left[\frac{1}{K} \sum_{n=1}^{K} \hat{\boldsymbol{A}}_n^T \boldsymbol{w} \boldsymbol{w}^T \hat{\boldsymbol{A}}_n \right], \quad d = 1, 2. \tag{4.24}$$

Then (4.18) can be written as

$$\begin{aligned} \min_{\boldsymbol{\theta}} \quad & g_2(\boldsymbol{\theta}|\boldsymbol{w}) = \boldsymbol{\theta}^T \boldsymbol{\Psi}_c(\boldsymbol{w}) \boldsymbol{\theta} \\ \text{subject to} \quad & \xi_2(\boldsymbol{\theta}|\boldsymbol{w}) = \boldsymbol{\theta}^T (\boldsymbol{\Psi}_1(\boldsymbol{w}) + \boldsymbol{\Psi}_2(\boldsymbol{w})) \boldsymbol{\theta} = 1, \end{aligned} \tag{4.25}$$

The solution of (4.25) is given by the generalized eigenvector corresponding to the minimal generalized eigenvalue of the generalized eigenvalue problem described as

$$\boldsymbol{\Psi}_c(\boldsymbol{w}) \boldsymbol{\theta} = \lambda (\boldsymbol{\Psi}_1(\boldsymbol{w}) + \boldsymbol{\Psi}_2(\boldsymbol{w})) \boldsymbol{\theta}. \tag{4.26}$$

As well as the method shown in Sec. 4.1, we alternately optimize $\boldsymbol{w}$ and $\boldsymbol{\theta}$ by solving the optimization problems of (4.21) and (4.25). In implementation, we initialize $\boldsymbol{\theta}$ as the P-dimensional vector as $\boldsymbol{\theta} = [1, 0, \ldots, 0]^T$, and start the alternative update from updating $\boldsymbol{w}$. The iteration is stopped, when $g(\boldsymbol{w}, \boldsymbol{\theta})$ reaches a certain criterion.

4.2.3 Feature Vector Definition

The method can extract the feature vector from an unlabeled data, $\boldsymbol{X}$, by the following way. First, we choose a class label, g, in $\{1, 2\}$, and then we obtain $\boldsymbol{w}_g$ and $\boldsymbol{\theta}_g$ by (4.18) with the learning dataset. Next, we calculate $\boldsymbol{\Phi}_d(\boldsymbol{\theta}_g)$, $d = 1, 2$ by (4.20), and solve (4.22). Then, we obtain the M eigenvectors as $\hat{\boldsymbol{w}}_1, \ldots, \hat{\boldsymbol{w}}_M$, where $\hat{\boldsymbol{w}}_i$ is the eigenvector corresponding to the ith minimum eigenvalue of (4.22). We assume that the $2r$ eigenvectors are used for classification. Then the feature vector, $\boldsymbol{y} \in \mathbb{R}^{2r}$ is defined as

$$\begin{aligned}\boldsymbol{y} =&[\rho(\boldsymbol{X}, \hat{\boldsymbol{w}}_1, \boldsymbol{\theta}_g), \ldots, \rho(\boldsymbol{X}, \hat{\boldsymbol{w}}_r, \boldsymbol{\theta}_g),\\ &\rho(\boldsymbol{X}, \hat{\boldsymbol{w}}_{M-r+1}, \boldsymbol{\theta}_g)), \ldots, \rho(\boldsymbol{X}, \hat{\boldsymbol{w}}_M, \boldsymbol{\theta}_g)]^T. \end{aligned} \tag{4.27}$$

As well as the case of CSP, we input $\boldsymbol{y}$ to a classifier.

4.3 Convergence of the Cost Function in the Optimization

The alternative optimizations of the proposed methods do not increase the cost functions and lead to local minimum solutions of (4.4) and (4.18). Here we give the guarantee of convergence of the cost in alternative iterations.

Proposition 1. *The cost function of* (4.4) *monotonically decreases or remains the same by iterations. That is,*

$$f(\boldsymbol{w}^{(k)}, \boldsymbol{h}^{(k)}) \geq f(\boldsymbol{w}^{(k+1)}, \boldsymbol{h}^{(k+1)}), \tag{4.28}$$

where $\boldsymbol{w}^{(k)}$ *and* $\boldsymbol{h}^{(k)}$ *are the parameters at* k*th iteration.*

Proof. Note that

$$f(\boldsymbol{w}^{(k)}, \boldsymbol{h}^{(k)}) = f_1(\boldsymbol{w}^{(k)}|\boldsymbol{h}^{(k)}) = f_2(\boldsymbol{w}^{(k)}|\boldsymbol{h}^{(k)}) \geq 0, \tag{4.29}$$

and

$$\nu(\boldsymbol{w}^{(k)}, \boldsymbol{h}^{(k)}) = \nu_1(\boldsymbol{w}^{(k)}|\boldsymbol{h}^{(k)}) = \nu_2(\boldsymbol{w}^{(k)}|\boldsymbol{h}^{(k)}) = 1. \tag{4.30}$$

Suppose $\boldsymbol{w}^{(k+1)}$ is given by minimizing (4.7) with $\boldsymbol{R}_d(\boldsymbol{h}^{(k)})$. Then the relationship;

$$f_1(\boldsymbol{w}^{(k)}|\boldsymbol{h}^{(k)}) \geq f_1(\boldsymbol{w}^{(k+1)}|\boldsymbol{h}^{(k)}), \tag{4.31}$$

is given, because $f_1(\boldsymbol{w}|\boldsymbol{h}^{(k)})$ with constraint of $\nu_1(\boldsymbol{w}|\boldsymbol{h}^{(k)}) = 1$ and fixing $\boldsymbol{h}^{(k)}$ achieves its minimum by $\boldsymbol{w}^{(k+1)}$. Moreover (4.31) leads

$$f(\boldsymbol{w}^{(k)}, \boldsymbol{h}^{(k)}) \geq f(\boldsymbol{w}^{(k+1)}, \boldsymbol{h}^{(k)}). \tag{4.32}$$

In a similar way, the relationship;

$$f(\boldsymbol{w}^{(k)}, \boldsymbol{h}^{(k)}) \geq f(\boldsymbol{w}^{(k)}, \boldsymbol{h}^{(k+1)}), \tag{4.33}$$

is given. This completes the proof. □

In a similar way, that (4.18) does not increase by the alternative optimization is also guaranteed.

Chapter 5

Experiment

We compare performance in classifying EEG signals during motor imagery using the proposed methods to that using conventional methods (CSP, CSSP, and SPEC-CSP).

5.1 Data Description

We used dataset IVa from BCI competition III [21], which was provided by Fraunhofer FIRST (Intelligent Data Analysis Group) and Campus Benjamin Franklin of the Charité - University Medicine Berlin (Department of Neurology, Neurophysics Group) [22]. This dataset consists of EEG signals during right hand and right foot motor-imageries. The EEG signals were recorded from five subjects labeled *aa*, *al*, *av*, *aw*, and *ay*. 118 EEG channels were measured at positions of the extended international 10/20-system. The measured signal was bandpass filtered with the passband of 0.05–200 Hz, and then digitized at 1000 Hz with 16 bit (0.1 μV). During each experiment, the

subject was given visual cues that were indicated for 3.5 seconds which of the three motor imagery should be performed: left hand, right hand, and right foot. The resting interval between two trials was randomized from 1.75–2.25 seconds. Only EEG trials for right hand and right foot were provided.

In this experiment, we furthermore applied to this data the lowpass filter whose the cutoff frequency is 50 Hz, and downsampled to 100 Hz. The dataset for each subject consisted of signals of 140 trials. A signal of one trial was measured for 3.5 seconds after a visual cue.

5.2 Result

First, we compare classification accuracy of the proposed methods with that of conventional methods in Sec. 5.2.1. The behavior of the cost function in the alternative optimizations are illustrated in Sec. 5.2.2. We discuss the amplitude characteristic of the filter and the spatial weights in each method in Sec. 5.2.3 and 5.2.4. Moreover, in Sec. 5.2.5, we consider suitable parameters in the proposed methods.

5.2.1 Classification Accuracy

In classification experiments, we extracted the feature vector from the EEG signal by each method, and classified the feature vector by linear support vector machine (linear SVM) [24]. The feature vector was extracted as follows.

- **CSP1:** A feature vector defined in (3.6) is used. We applied the Butterworth bandpass filter with the passband of 7–30 Hz, and minimized

the variance cost of the right hand class in (3.2).

- **CSSP:** We minimized the variance cost of the right hand class in (3.2). And, we optimized τ (τ was chosen in 1–15 samples) by 5×5 CV in the learning dataset for each CV. The signals are applied the bandpass filter between 7–30 Hz as preprocessing.

- **SPEC-CSP:** We assumed that the filter coefficients not corresponding to 7–30 Hz are 0. The iteration number in alternative optimization was 30.

- **Proposed1:** A feature vector defined in (4.27) is used, where the class corresponding to the right hand movement imagery was chosen as the class, g. The frame length, N, and the frame delay, D, were set to 30 and 5, respectively. In optimization, we stopped iteration, when error of the cost function between successive iterations becomes under 10^{-5}.

- **Proposed2:** A feature vector defined in (4.27) is used, where the class corresponding to the right hand movement imagery was chosen as the class, g. The filter order, P, was set to 20. In optimization, we stopped iteration, when error of the cost function between successive iterations becomes under 10^{-5}.

The extracted feature vector was classified by linear SVM with the soft margin parameter of 50 [24]. The linear SVM was implemented with SVM-Light [25]. Table 5.1 shows classification accuracy in each method which is calculated with 5×5 CV. For reference, we show in table 5.1 the accuracy given by the method using the filter which is searched manually. The

Table 5.1: Classification accuracy [%] given by 5×5 CV. The figures in the round brackets beside accuracies represent the number of dimensions of the feature vector used for classifications.

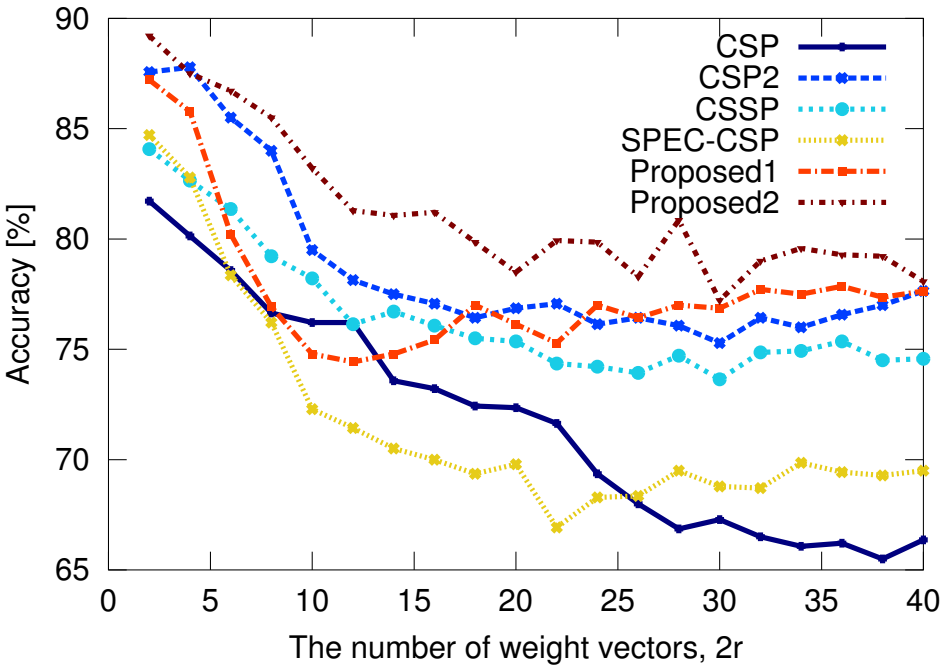
	Subject				
Method	*aa*	*al*	*av*	*aw*	*ay*
CSP1	81.7 (2)	94.6 (6)	68.3 (8)	95.9 (6)	89.6 (2)
CSP2	88.4 (4)	98.6 (2)	74.4 (10)	99.9 (6)	92.4 (4)
CSSP	84.1 (2)	95.4 (10)	69.9 (8)	96.9 (6)	90.5 (2)
SPEC-CSP	84.7 (10)	95.3 (6)	59.0 (20)	96.9 (6)	83.4 (4)
Proposed1	87.2 (2)	97.2 (2)	57.3 (18)	97.9 (2)	91.5 (2)
Proposed2	89.2 (2)	98.6 (4)	58.3 (6)	98.4 (16)	92.2 (4)

reference of classification accuracy labeled CSP2 is given as follows.

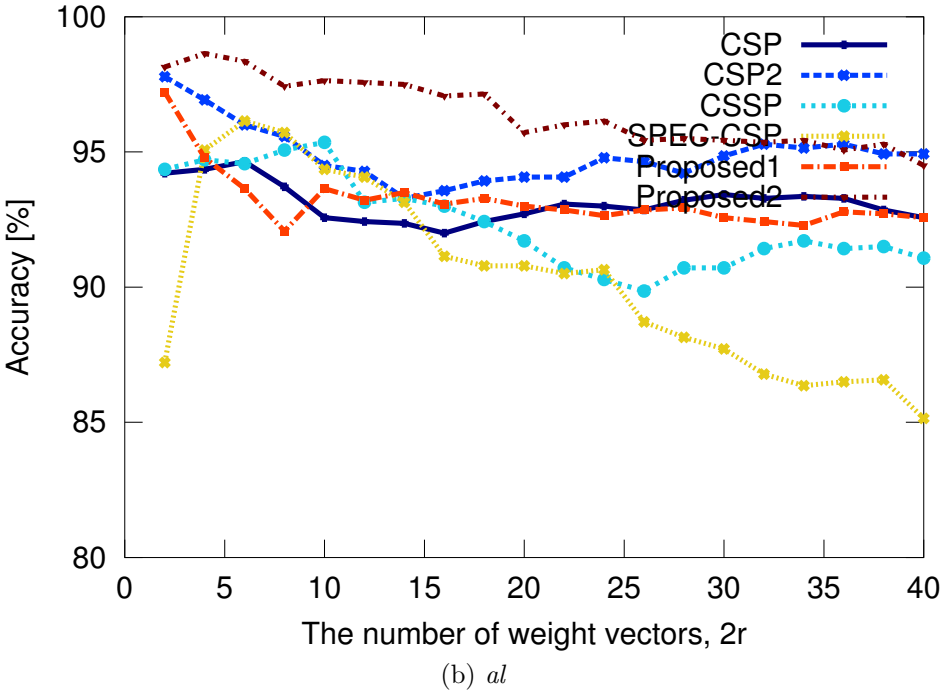
- **CSP2:** We show the best classification accuracy when several filters are applied to preprocessing in CSP1. The candidates of passbands are represented as f_l–f_u Hz for $f_l = 1, \ldots, 25$ and $f_u = f_l + 1, \ldots, 30$. That is, the number of the candidates is 266 for each subject. The passbands that provide the best accuracies are 11–16 Hz (*aa*), 12–16 Hz (*al*), 21–26 Hz (*av*), 11–18 Hz (*aw*), and 9–12 Hz (*ay*).

It is shown that the proposed methods outperform CSSP and SPEC-CSP in most of subjects (except to *av*) in classification accuracy.

Figure 5.1–5.3 shows the classification accuracy to the dimension of the feature vector in each method. We can observe from Fig. 5.1–5.3 the dependence on the dimension of the feature vector. This figure indicates that the dimension that provides the best classification accuracies in Proposed1 for most of subject is small and consistent, say 2, although the dimension that gives the best accuracy that CV result in strongly depends on subjects.

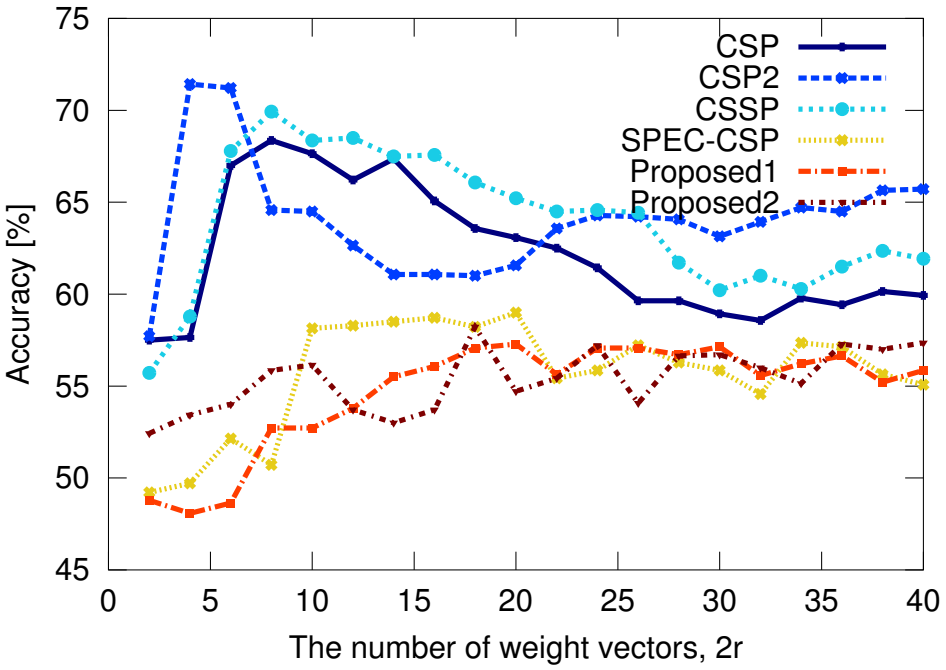


(a) *aa*

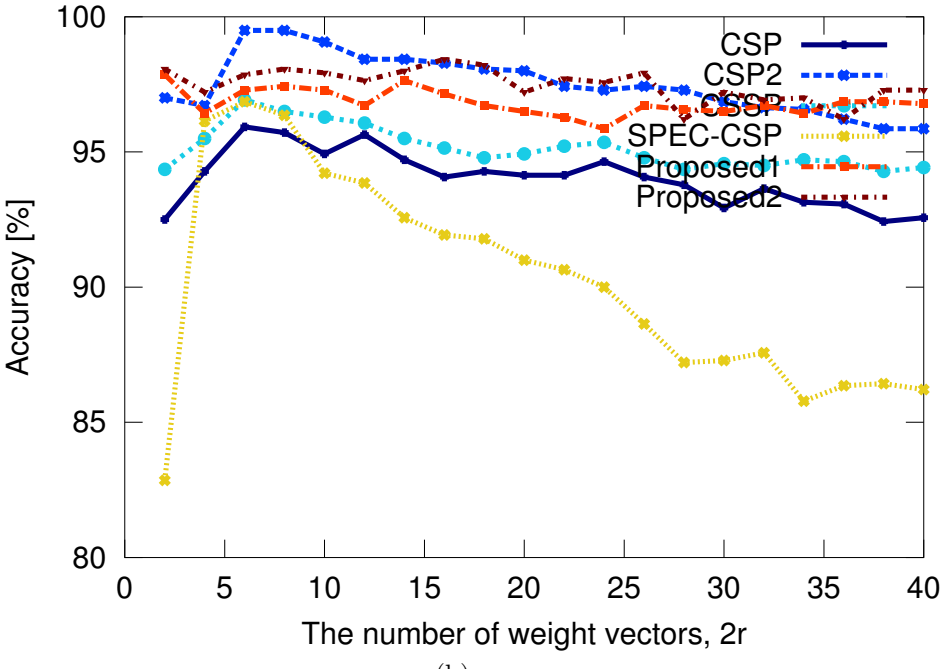


(b) *al*

Figure 5.1: Classification accuracy for the number of the weight vectors, $2r$, in subject *aa* and *al*

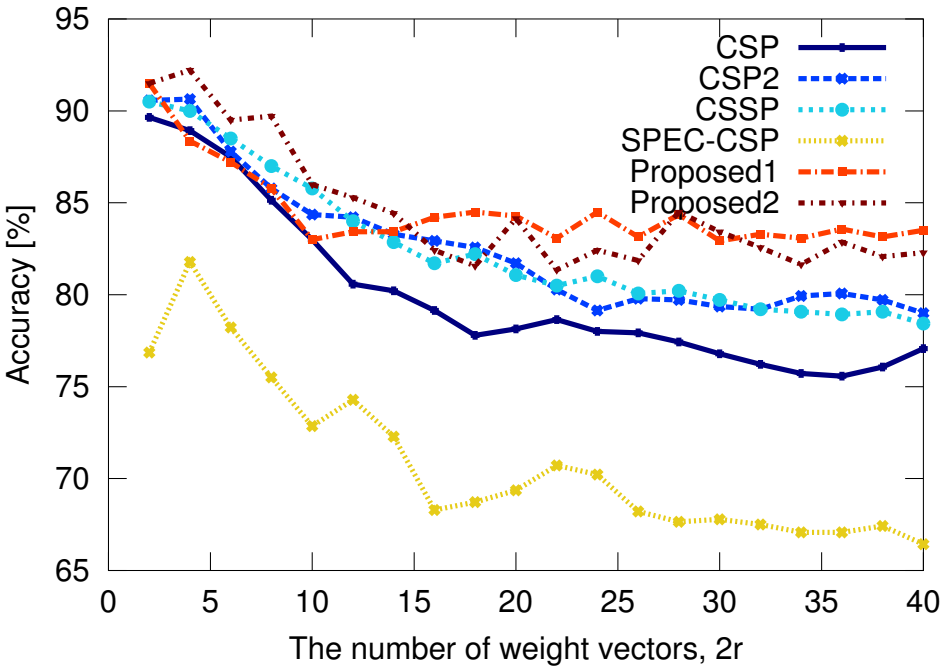


(a) *av*

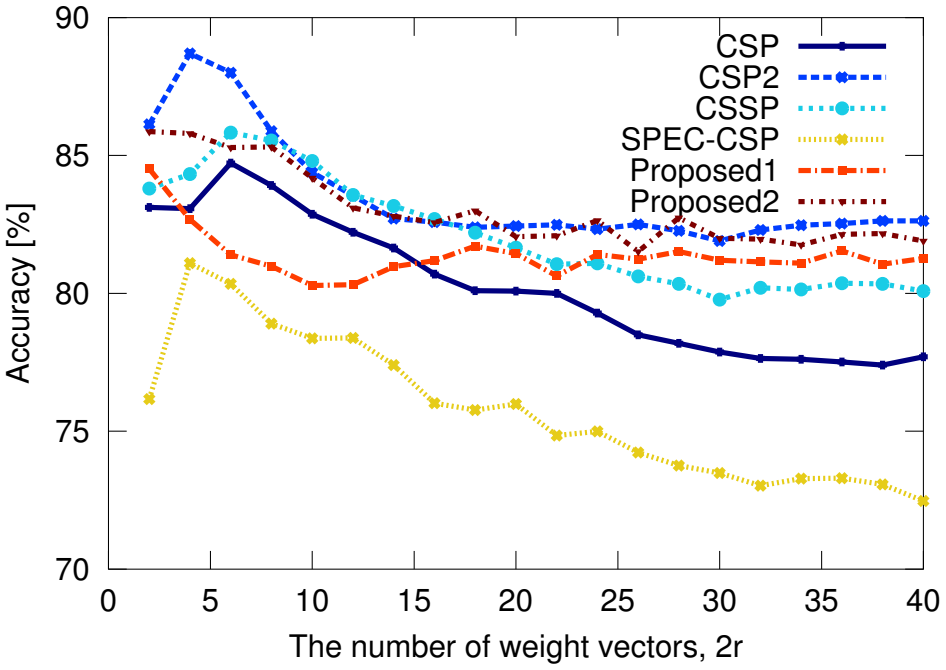


(b) *aw*

Figure 5.2: Classification accuracy for the number of the weight vectors, $2r$, in subject *av* and *aw*



(a) *ay*



(b) Average

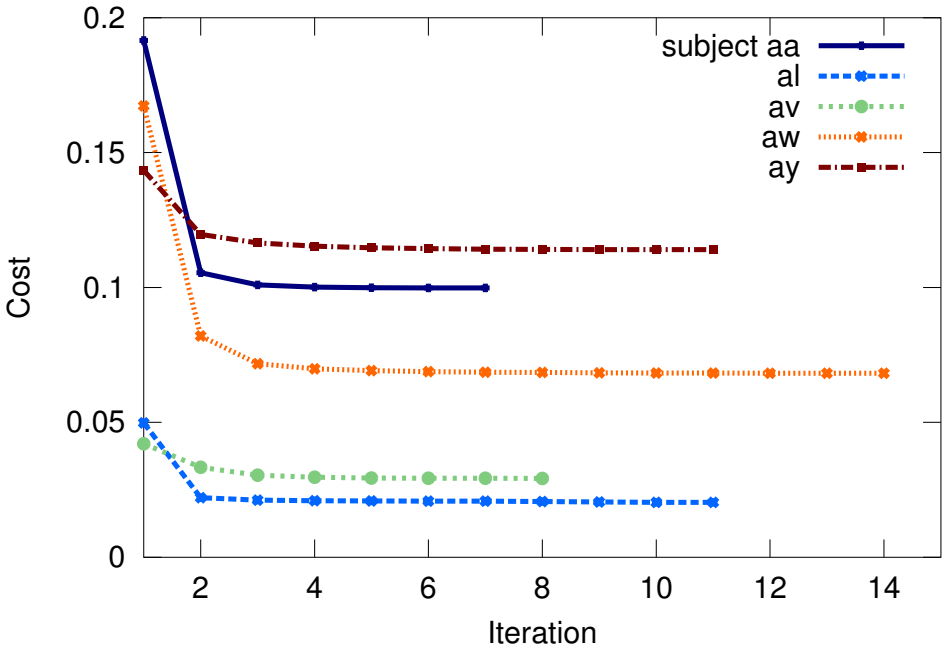
Figure 5.3: Classification accuracy for the number of the weight vectors, $2r$, in subject *av* and average over all subjects.

5.2.2 Convergence of the Proposed Methods

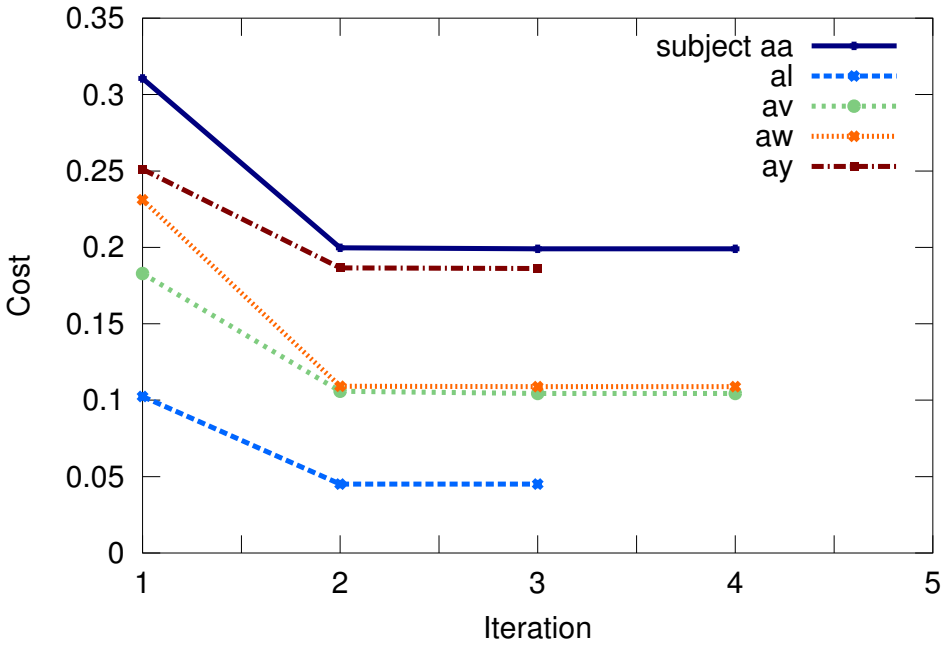
We illustrate the behavior of the cost function in the alternative optimizations in Fig. 5.4. The optimization parameters are optimized under the parameters shown in Sec. 5.2.1. We can observe the monotonically decreasing and fast convergence cost.

5.2.3 Amplitude Characteristic of the Filter

Figure 5.5 and 5.6 shows the amplitude characteristic of the filter used/obtained in each method. Each method was applied to datasets under the parameters shown in Sec. 5.2.1, and the amplitude characteristic in each method was obtained as follows. In CSP2, we show the amplitude response of the Butterworth filter used as preprocessing. In CSSP, we choose the coefficients for the channel which has largest coefficients as w^0 and w^τ. We show the amplitude response of the FIR filter given by w^0 and w^τ. In SPEC-CSP, since the filter is given as the weight vector for the spectrum, we show elements of the spectral weight vector, $\boldsymbol{\beta}_1$ shown in (3.8). In Proposed1, we show elements of $|\boldsymbol{h}_g|$ shown in (4.13). In Proposed2, we show the amplitude characteristic of the FIR filter whose its coefficients is $\boldsymbol{\theta}_g$ shown in (4.27). In all methods, the amplitude characteristics are normalized the maximum to 1. From Fig. 5.5–5.6, we can observe that the frequency bands specified by SPEC-CSP, Proposed1, and Proposed2 have large weights in the passband of CSP2. Moreover, SPEC-CSP and Proposed1 also have large weights in band higher than the passband of CSP2.

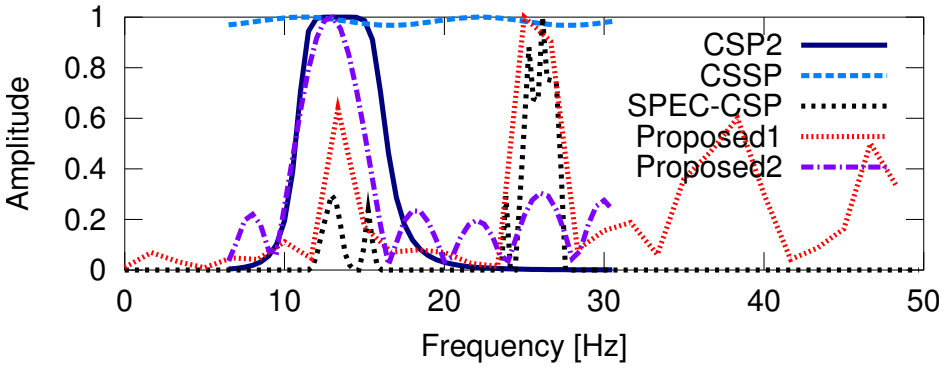


(a) Proposed1

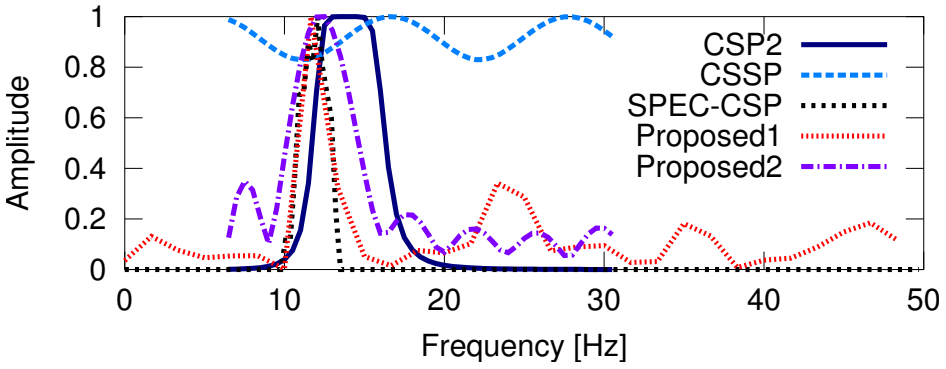


(b) Proposed2

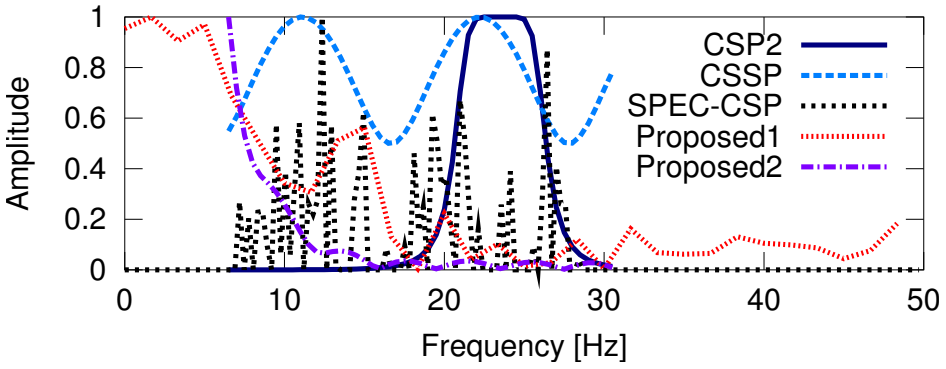
Figure 5.4: Cost functions for iteration in the optimizations in the proposed methods.



(a) Subject *aa*

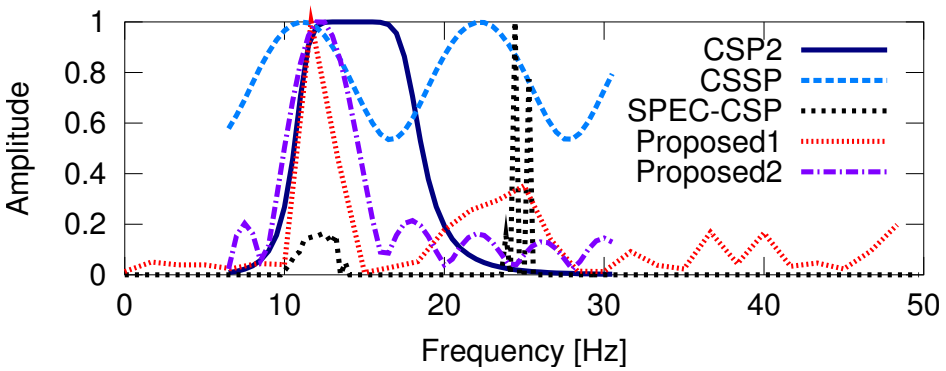


(b) Subject *al*

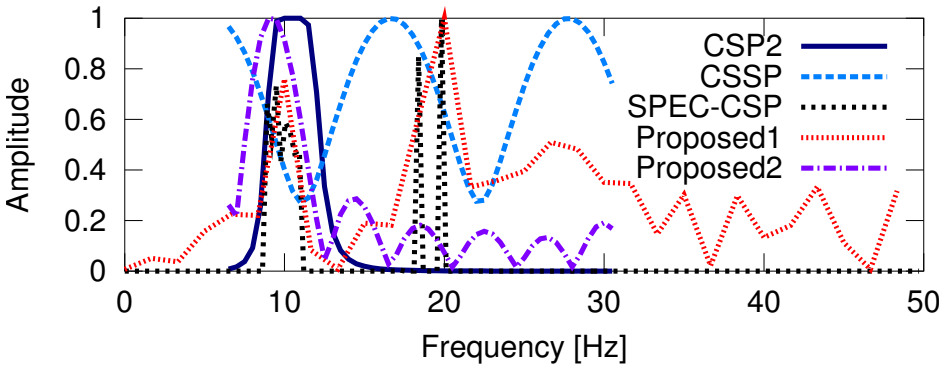


(c) Subject *av*

Figure 5.5: Amplitude characteristic of the filter given by each method in subject *aa*, *al*, and *av*. Note that the characteristics in 7–30 Hz are not plotted in the methods which used the bandpass filter with passband of 7–30 Hz (CSSP, Proposed2).



(a) Subject *aw*



(b) Subject *ay*

Figure 5.6: Amplitude characteristic of the filter given by each method in subject *aw* and *ay*. Note that the characteristics in 7–30 Hz are not plotted in the methods which used the bandpass filter with passband of 7–30 Hz (CSSP, Proposed2).

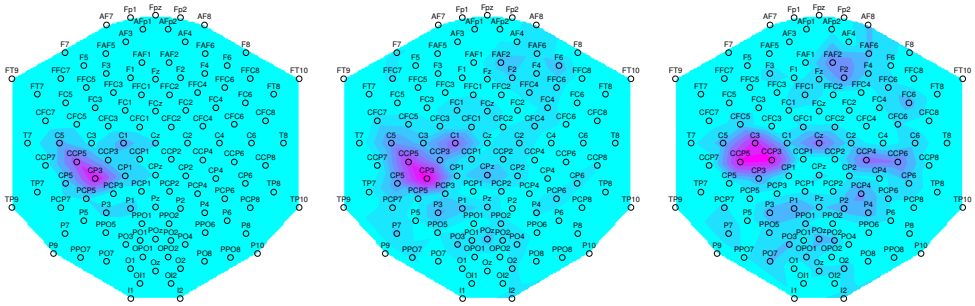
5.2.4 Spatial Weight

In Fig. 5.7–5.11, we topographically show amplitudes of the spatial weights obtained by each method in subject *aa* and *ay*. Each method was applied to datasets under the parameters shown in Sec. 5.2.1. These spatial weights minimize the feature value of an observed signal during right hand movement imagery. In Fig. 5.7, all of the spatial weights are almost same. However, we can observe that the different weight vectors are obtained between the each method in Fig. 5.11. Additionally, from Fig. 5.11a and Fig. 5.11b, we can observed that the spatial weight is changed by the passband of a bandpass filter.

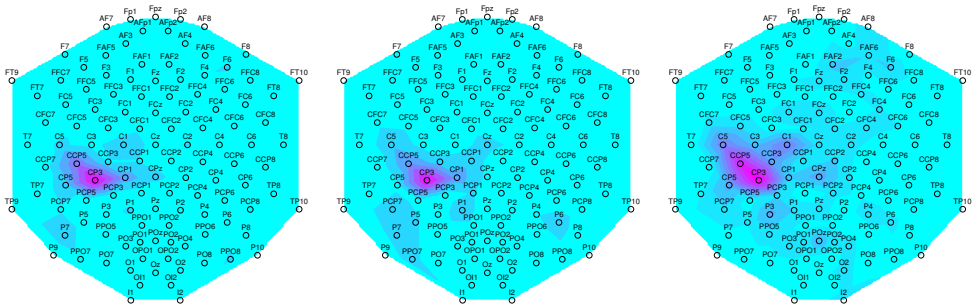
5.2.5 Parameters in the Proposed Method

Figure 5.12 shows classification accuracy averaged over all subjects for the parameters of Proposed1. The classification accuracy is given by 5×5 CV. In Fig. 5.12, accuracy decreases in $D > 50$ (0.5 secs) with almost N', and according as increasing N' decreases accuracy when D is small.

Figure 5.13 shows the classification accuracy for the parameter of Proposed2 in each subject. This result suggests that the classification accuracy decreases when the filter order is smaller than 15, and is not sensitive for the order of the filter when the order is higher than 15.

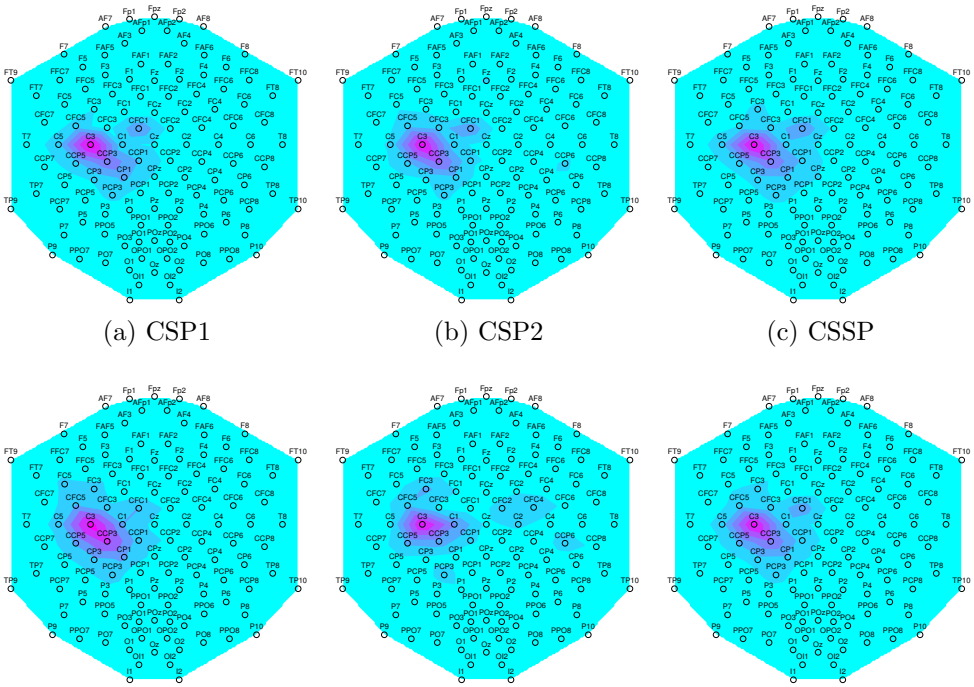


(a) CSP1 (b) CSP2 (c) CSSP



(d) SPEC-CSP (e) Proposed1 (f) Proposed2

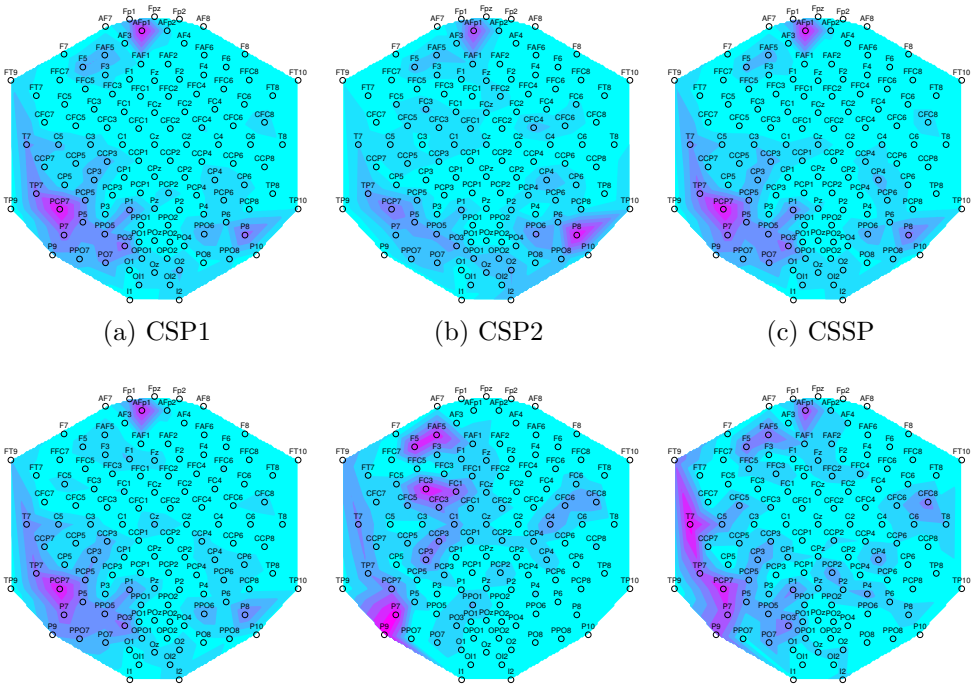
Figure 5.7: The coefficients of the spatial weights given by CSP and the proposed method in subject *aa*. The color level represents the amplitude of coefficients. Upper and lower part of the topography corresponds front of the head and the back, respectively.



(a) CSP1 (b) CSP2 (c) CSSP

(d) SPEC-CSP (e) Proposed1 (f) Proposed2

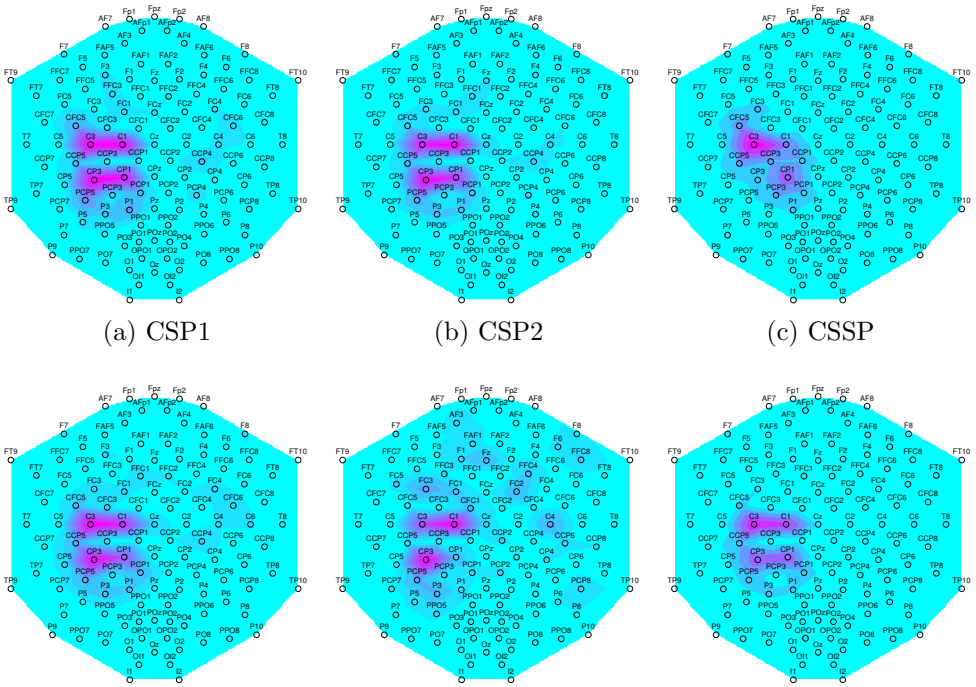
Figure 5.8: The coefficients of the spatial weights given by CSP and the proposed method in subject *al* as well as Fig. 5.7.



(a) CSP1 (b) CSP2 (c) CSSP

(d) SPEC-CSP (e) Proposed1 (f) Proposed2

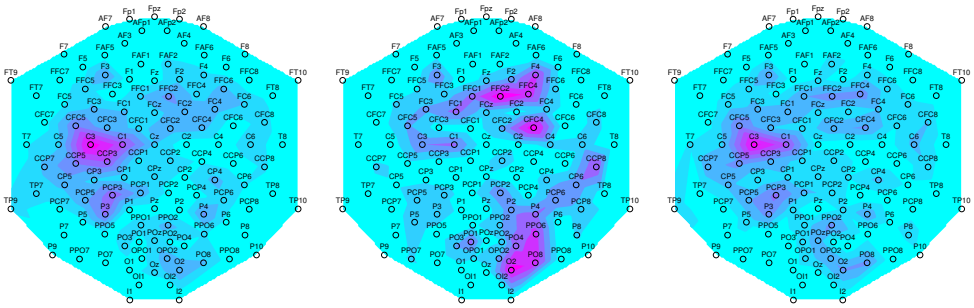
Figure 5.9: The coefficients of the spatial weights given by CSP and the proposed method in subject *av* as well as Fig. 5.7.



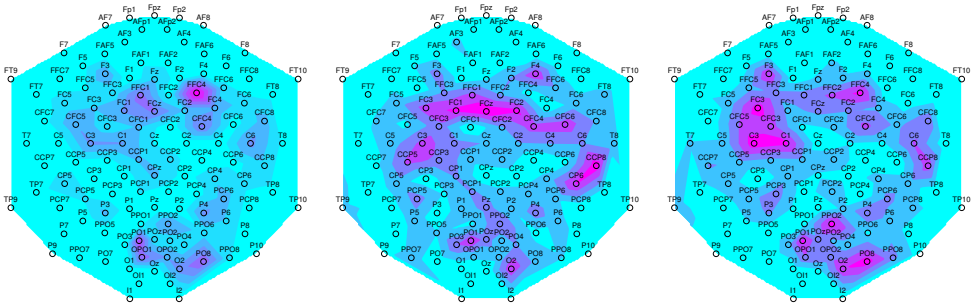
(a) CSP1 (b) CSP2 (c) CSSP

(d) SPEC-CSP (e) Proposed1 (f) Proposed2

Figure 5.10: The coefficients of the spatial weights given by CSP and the proposed method in subject *aw* as well as Fig. 5.7.



(a) CSP1 (b) CSP2 (c) CSSP



(d) SPEC-CSP (e) Proposed1 (f) Proposed2

Figure 5.11: The coefficients of the spatial weights given by CSP and the proposed method in subject *ay* as well as Fig. 5.7.

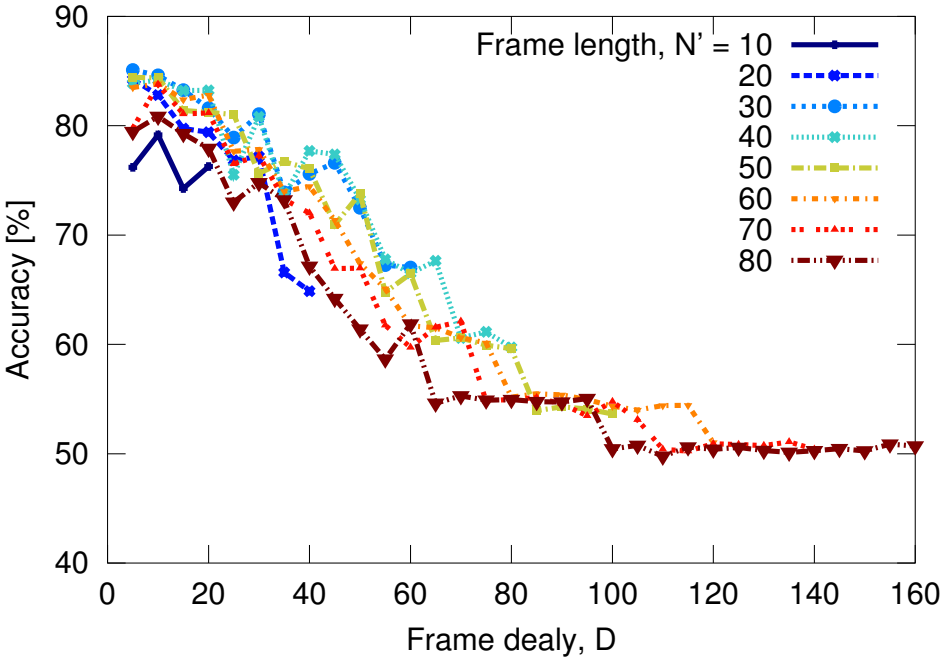


Figure 5.12: Classification accuracy averaged over all subjects for parameters of the proposed method. Accuracy rates are given by 5×5 CV.

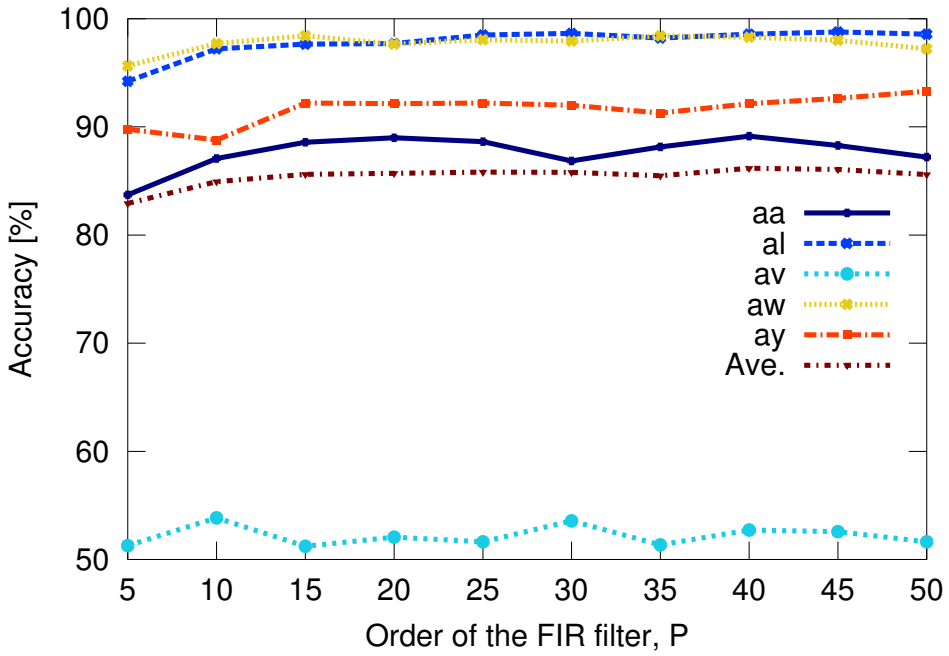


Figure 5.13: Classification accuracy for the filter order, P of Proposed2. Accuracy rates are given by 5×5 CV

Chapter 6

Conclusions

We have proposed a novel method for classification of EEG signals during motor imagery. Unlike CSP, parameters to be found are not only spatial weights but also frequency filters. The cost function to determine the parameters is defined as a natural extension of CSP. The optimal parameters are found by alternately solving the generalized eigenvalue problems. By experiment, we have demonstrated that the proposed methods achieve high classification accuracy. Our experimental results suggest also that the proposed methods can specify the frequency band same as the band searched by manually.

Because the classification accuracy by the proposed method can be variable under the influences of parameters, we will clarify this causes and solve the problem of the parameter tuning. Moreover, we will develop the algorithm for online application.

Acknowledgement

I would like to thank my adviser Dr. Toshihisa Tanaka who is Associate Professor at Tokyo University of Agriculture and Technology for his valuable support. Also, I would like to thank the members of Toshihisa Tanaka's laboratory and Laboratory for Advanced Brain Signal Processing, Brain Science Institute, RIKEN.

Finally, I want to express gratitude for all my teacher, friends, and family for helping and supporting me.

Bibliography

[1] S. Sanei and J. Chambers, *EEG Signal Processing.* Hoboken, NJ: John Wiley & Sons, 2007.

[2] J. R. Wolpaw, N. Bribaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clinical Neurophysiology*, vol. 113, pp. 767–791, 2002.

[3] J. Müller-Gerling, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," *Clinical Neurophysiology*, vol. 110, pp. 787–789, 1999.

[4] H. Ramoser, J. Müller-Gerling, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 4, pp. 441–447, 2000.

[5] S. Lemm, B. Blankertz, G. Curio, and K. R. Müller, "Spatio-spectral filters for improved classification of single trial EEG," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 9, pp. 441–446, 2005.

[6] 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.

[7] R. Tomioka, G. Dornhege, G. Nolte, K. Aihara, and K. R. Müller, "Optimizing spectral filters for single trial EEG classification," *Lecture Notes in Computer Science*, vol. 4174, pp. 414–423, 2006.

[8] G. Dornhege, J. R Millán, T. Hinterberger, D. J. McFarland, and K. R. Müller, *Toward Brain-Computer Interfacing*. Cambridge, MA: MIT Press, 2007.

[9] M. R. Nuwer, "Recording electrode site nomenclature," *Journal of Clinical Neurophysiology*, vol. 4, pp. 121–133, 1987.

[10] E. Suarez, M. D. Viegas, M. Adjouadi, A. Barreto, "Relating induced changes in EEG signals to orientation of visual stimuli using the ESI-256 machine," *Biomedical Science Instrumentation*, vol. 36, no. 33–38, 2000.

[11] R. Oostenveld, P. Praamstra, "The five percent electrode system for high-resolution EEG and ERP measurements," *Clinical Neurophysiology*, vol. 112, pp. 713–719, 2001.

[12] G. Pfurtscheller, C. Neuper, A. Schögl, and K. Lugger, "Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameteres," *IEEE Transactions on Rehabilitation Engineering*, vol. 6, no. 3, pp. 316–325 2002.

[13] M. Middendorf, G. McMillan, G. Calhoun, and K. S. Jones, "Brain-computer interfaces based on the steady-state visual-evoked response," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 2, pp. 211–214, 2002.

[14] J. R. Wolpaw and D. J. McFarland, "Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans," in *proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 51, pp. 17849, 2004.

[15] C. Neuper, G. R. Müller, A. Kübler, N. Birbaumer, and G. Pfurtscheller, "Clinical application of an EEG-based brain-computer interface: a case study in a patient with severe motor impairment," *Clinical Neurophysiology*, vol. 114, pp. 399–409, 2003.

[16] G. Pfurtscheller, C. Guger, G. Müller, G. Krausz, and C. Neuper, "Brain oscillations control hand orthosis in a tetraplegic," *Neuroscience Letters*, vol. 292, no. 3, pp. 211–214, 2000.

[17] G. Pfurtscheller, G. R. Müller, J. Pfurtscheller, H. J. Gerner, and R. Rupp, "Thought-control of functional electrical stimulation to restore hand grasp in a patient with tetraplegia," *Neuroscience Letters*, vol. 351, no. 1, pp. 33–36, 2003.

[18] B. Obermaier, G. R. Müller, and G. Pfurtscheller, "Virtual keyboard controlled by spontaneous EEG activity," *IEEE Transactions on Neural System and Rehabilitation Engineering*, vol. 11, no. 4, pp. 422–426, 2003.

[19] L. A. Farwell and E. Donchin, "Taking off the top of your head: toward a mental prosthesis utilizing evnet-related brain potentials," *Electroencephalography and Clinical Neurophysiology*, vol. 70, no. 6, pp. 510–523, 1988.

[20] B. Blankertz, G. Dornhege, S. Lemm, M. Krauledat, G. Curio, and K. R. Müller, "The Berlin Brain-Computer Interface: Machine learning based detection of user specific brain states," *Journal of Universal Computer Science*, vol. 12, no. 6, pp. 581–607, 2006.

[21] B. Blankertz, K. R. Müller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlogl, G. Pfurtscheller, J. R. Millan, M. Schroder, and N. Birdaumer, "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.

[22] 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.

[23] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[24] V. N. Vapnik, *Statistical Learning Theory*. New York, Wiley, 1998.

[25] T. Joachims, "SVM-Light," `http://www.cs.cornell.edu/People/tj/svm_light/`.

List of Publications

International Conference Papers (refereed)

1. H. Higashi and T. Tanaka, "Classification by weighting for spatio-frequency components of EEG signal during motor imagery," in *Proceedings of 2011 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Prague, Czech Republic, May 2011 (accepted).

2. Y. Washizawa, H. Higashi, T. Rutkowski, T. Tanaka, and A. Cichocki, "Tensor based simultaneous feature extraction and sample weighting for EEG classification," *Lecture Notes in Computer Science*, vol. 6444, pp. 26–33, Sydney, Australia, Nov. 2010.

3. H. Higashi, T. Tanaka, and Y. Mitsukura, "Rhythmic component extraction considering phase alignment and the application to motor imagery-based brain computer interfacing," in *Proceedings of 2010 International Joint Conference on Neural Networks*, pp. 3508–3513, Barcelona, Spain, Jul. 2010.

4. T. Tanaka, H. Higashi, and Y. Saito, "Rhythmic component extraction for EEG signals with reduced computational complexity," in *Proceed-*

ings of 2009 IEEE International Symposium on Biomedical Engineering, no. 1054, Bangkok, Thailand, Dec. 2009.

5. H. Higashi, T. Rutkowski, Y. Washizawa, T. Tanaka, and A. Cichocki, "Imagery movement paradigm user adaption improvement with quasi-movement phenomenon," *Advances in Cognitive Neurodynamics (II)*, pp. 683–688 Hangzshou, China, Nov. 2009.

6. H. Higashi, T. Tanaka, and A. Funase, "Classification of single trial EEG during imagined hand movement by rhythmic component extraction," in *Proceedings of 31st Annual International Conference of the IEEE EMBS*, pp. 2482–2485, Minneapolis, USA, Sept. 2009.

7. Y. Saito, T. Tanaka, and H. Higashi, "Adaptive rhythmic component extraction with regularization for EEG data analysis," in *Proceedings of 2009 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 353–356, 2009.

Domestic Conference Papers (not refereed)

1. 東 広志, 田中聡久, "運動想像時脳波識別のためのFIRフィルタ設計法," 電子情報通信学会総合大会, 2011年3月（予定）.

2. 東 広志, 鷲沢嘉一, T. Rutkowski, 田中聡久, A. Cichocki, "定常的聴覚誘発電位を用いた脳コンピュータインターフェイス," 信学技報, vol. 110, No. 368, SIP2010-106, pp. 221–226, 2011年1月.

3. 東 広志, 田中聡久, "運動想像時脳波の空間-周波数成分の重み付けによる

識別," 電子情報通信学会第 25 回信号処理シンポジウム論文集, pp. 121–126, 奈良, 2010 年 11 月.

4. 東 広志, 田中聡久, "脳コンピュータインターフェイスのための位相を用いた律動成分抽出法," 電子情報通信学会第 24 回信号処理シンポジウム論文集, pp. 402–407, 鹿児島, 2009 年 11 月.

5. 木村陽介, 東 広志, 田中聡久, "BCI のための律動成分抽出を用いた定常的視覚誘発電位の観測法," 信学技報, NC2009-54, pp. 23–28, 宮城, 2009 年 11 月.

6. 田中聡久, 斎藤祐樹, 東 広志, "律動成分分析の適応高速アルゴリズムに関する一検討," 信学技報, SIP2009-27, pp. 49–53, 北海道, 2009 年 6 月.

7. 斎藤祐樹, 田中聡久, 東 広志, "適応律動成分抽出法と脳波解析への応用," 電子情報通信学会第 23 回信号処理シンポジウム論文集, pp. 58–63, 宮城, 2008 年 11 月.