Study of Feature Extraction for Motor-Imagery EEG Signals

Hiroshi Higashi

DISSERTATION

Submitted to The Graduate School of Engineering of

Tokyo University of Agriculture and Technology

in Electronic and Information Engineering

for The Degree of

DOCTOR OF PHILOSOPHY

TOKYO UNIVERSITY OF AGRICULTURE AND TECHNOLOGY

December 2013

Abstract

This study proposes signal processing methods to presume or classify brain state of a human by using signals observed by electrodes installed on a scalp (EEG; Electroencephalogram). The issue of classifying the brain states from EEG signals is important for realization of brain computer/machine interface (BCI/BMI) and its applications (e.g. rehabilitation). Target states of the brain to be classified in the study are brain states when the human is imaging movements of he/she own hands and feet. The task performing the imaginations is called motor-imagery (MI) task. The goal in this classification is to classify EEG signals into classes corresponding to motor-imagery parts of body (e.g. right/left hand, feet)

As features in the EEG signals associated with the MI tasks, the variation in energy of certain bands observed around motor cortex is known. For extracting the features, we can use a spatial filter that weights with different coefficients to each electrode, a frequency filter that extracts the certain bands, and a time window that removes periods not including feature components. In many cases, the parameters such as the coefficients of the filters and the window are empirically determined based on knowledge of neuroscience. For instance, the electrodes are installed around motor cortex, the passband of the frequency filter is set to 7 to 30 Hz including the band of the so-called mu rhythm the energy of which decrease by MI tasks, and the time window extracts signals observed from a start cue having

a subject perform a task to an end cue. However, the optimal parameters for the filters highly depend on measurement environments and individuals. Therefore, a decision method for the parameters adapted for data is needed.

This study proposes methods to determine the parameters by learning observed samples. The proposed methods expand the concept of the common spatial patterns (CSP) method that is well-known for designing effective spatial filters for classification in 2-class MI-BCI. The proposed methods extract a signal from a raw observed signal by the following way. First, we adopt an FIR filter as the frequency filter. Second, we use linear combination of a multi-channel signal and weight coefficients as the spatial filter. Finally, we apply a time window with binary coefficients to a signal after the start cue. To determine the parameters composed of three elements (frequency, space, time), we take advantage of EEG signals with class labels as a learning dataset. In the learning, the parameters are determined in such a way that the ratio of the energy of the extracted signals between two classes is maximized. We propose an approximate solution using a method that maximizes alternately parameters of each element. Moreover, we introduce an orthogonal constraint over the coefficient vectors of the multiple FIR filters. We propose a sequentially optimizing method for the optimization problem including the orthogonal constraint. In this way, the proposed method designs a set of the filters (filterbank) that have different characteristics in each other. Even in the situation that some components are associated with the MI task, the proposed method is able to extract separately the components by the use of the filterbank.

Moreover, collecting EEG signals takes a long time and small sample problem in the parameter learning can happen. It causes overfittings and ill-posed problems. To address the problem, we propose regularizations for optimization

problems of the spatial filters. At first, we make an assumption that the electrodes that are located close to each other observe an electric activity of the same neurons. Under the assumption, the proposed regularizations work in such a way that the weight coefficients or the weighted signals for the nearby sensors take similar values in the optimization problems. By adding spatial information of an electrode arrangement as a prior information, the robust coefficients of the spatial filter can be found.

Experimental analysis for the proposed methods is conducted with artificial signals. Moreover, we show the EEG classification experiments of the MI tasks. The results show the proposed methods improve accuracy of the classification.

Acknowledgements

This work would not have been possible without the guidance of my academic supervisor, Dr. Tohihisa Tanaka. I am very grateful for his advice and support that have opened new gates of opportunities. My deepest gratitude also goes to the doctoral committee members, Dr. Hitoshi Kitazawa, Dr. Masatoshi Sekine, Dr. Akinobu Shimizu, and Dr. Toshiyuki Kondo. Their questions, comments, and suggestions have greatly helped in the improvement of this work.

I would like to extend my grateful acknowledgement to the Japan Society for the Promotion of Science (JSPS) and RIKEN. They have helped me with financial support throughout my academic life.

I would like to thank to Dr. Cichocki Andrzej, the head of the Laboratory for Advanced Brain Signal Processing, Brain Science Institute, RIKEN, and members of his laboratory. He has given me opportunities for collaborations with his laboratory. Especially, I am grateful to ex-researchers in the laboratory, Dr. Tomasz M. Rutkowski who is now an assistant professor, The University of Tsukuba, and Dr. Yoshikazu Washizawa, who is now an assistant professor, The University of Electro-Communications.

I would like to represent gratitude to the Machine Learning Group, Berlin Institute of Technology. Dr. Klaus-Robert Müller, the head of the group, and Dr. Benjamin Blankertz accepted my request to stay in their laboratory for three

months. The members of the laboratory gave me useful advices and showed some advanced researches that encourage me to study hard. The experience in Berlin made me grow up as a researcher.

I am grateful for professors and fellows who have given me valuable advices at conferences. To the members of Toshihisa Tanaka's laboratory, I would like to thank for a good time with you and your valuable supports.

Last, thank my family. I am very grateful to your financial and moral supports to me.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

Decording brain activity from brain signal is an important and challenging technology. The applications of the technology are detection of diseases [1, 2], design of an interface [3] and research for brain function [4–7], etc. An interface using brain signal is called a brain machine interface (BMI). BMI is an interface connecting a human brain and an external device. Especially, BMIs to send commands to the device is called output-type BMIs. A brain activity that is evoked by certain tasks is allocated to an output of the interface. A user of the interface performs the allocated tasks to generate the output. The tasks inducing the brain activities are not limited to just tasks with muscular movements. Certain mental tasks such as turning attention to external stimuli and imagination of something can be used as the tasks for BMIs. Therefore BMIs realize non-muscular communication and control channel for conveying messages and commands to the external world [3, 8, 9].

BMI is a challenging technology of nueroscience, signal processing, and pat-

tern recognition. From the neuroscience point of view, they explore mental tasks and responses to stimuli that can be observed in brain signals. People who are in signal processing and pattern recognition societies are interested in how to extract the features evoked by the tasks from the observed signals and how to classify the extracted features to the task. Above all, this study mainly tackles problems of how to extract features associated to the tasks.

For acquisition of the brain signals in BMI, there are invasive and noninvasive methods [10]. Electrocorticogram (ECoG), electrosubcorticogram (ESCoG), and electroventriculogram (EVG) are typical invasive measurement methods. They need surgeries for installing electrodes on a cortex or a cerebral ventricle and measure electrical activities of brain neurons. The invasive methods can measure the brain activities with noise much less than that in the noninvasive methods. BMIs with the invasive measurements have realized without muscle movements a control of a lever by a rat [11], a control of a robot arm by a monkey [12], and a control of a computer cursor by a patient suffering from motor disorders [13]. On the other hand, the noninvasive methods do not require such medical surgeries. The noninvasive methods are considered that they impose a less load on subject and are more practical methods for realizing BMIs than the invasive methods [14]. Noninvasively measured data such as electroencephalogram (EEG) [15], magnetoencephalogram (MEG) [16], and functional magnetic resonance imaging (fMRI) [17], and near-infrared spectroscopy [18] are widely used to the BMI research. The details of these noninvasive methods are shown in Chapter 2.2. Among them, because of its simplicity and low cost, EEG is practical for use in engineering applications [19, 20]. Moreover, EEG can achieve higher temporal resolution than the other invasive methods [8]. In this study, we use EEG as the

measurement method for BMI. Although EEG is considered to be practical for BMI, decrease of desired components, high noise, and low spatial resolution are significant problems. The BMI systems with EEG (EEG-based BMI) such as an input of letters [21–23] and controls of an object in a monitor [24, 25], a wheel chair [26], and a robot [19, 27] have been developed.

BMIs can be categorized by the tasks which a subject performs. The frameworks composed of the tasks and feedbacks to the subject are often called paradigms. Typical BMI paradigms are perception of random-displayed stimuli, gazing at flicker stimuli, and imagination of muscle movements. The details of these paradigms are shown in Chapter 2.3. This paradigm of imagination of muscle movements is called the motor-imagery (MI) paradigm and the BMI using MI is called MI-based BMI (MI-BMI) [3, 9]. In MI-BMI, an observed brain signal is classified into classes corresponding to a body part which a subject imagines movement of. For example, the right hand, the left hand, feet, and tongue of the MI tasks are widely used for realizing BMI [28, 29].

This study addresses problems for feature extraction in the BMI using the paradigm of imagination of muscle movements. The reasons why we focus on the MI paradigm are as follows. As merits as an interface, unlike the paradigms that make use of responses to some stimuli, MI-BMIs do not need stimuli because a subject performs the MI tasks voluntarily. Moreover, the subject needs at least a perception out of visual, auditory, tactile, olfactory, and gustatory perceptions in order to respond to the stimuli. On the other hand, the MI paradigm can be used by patients suffering from sense disorders and even for locked-in patients. Recently, some researches explore a new field of application of MI-BMI in medical rehabilitation area [30, 31]. The rehabilitation of the application aims to recover

motor function in patients who have damages in brain regions that play a role in the motor function. Typically, the rehabilitation asks the patient to try to move its paralyzed limb. Then, the motor function can be gradually recovered as the patient tries it repeatedly. The recovery makes use of neuroplasticity that is a property of changes in neural pathways and synapses which are resulted by stimuli such as behavior, environment and neural processes [32]. Thanks to neuroplasticity, the disabled motor function is recovered in the different region from the damaged region. In the process of trying the movement, it has been suggested that physical feedbacks to the paralyzed part promote neuroplasticity [33]. For instance as the feedbacks, a robot gives force to the paralyzed part and moves the part. Some researches suggest that the use of BMI for the control of the feedbacks is effective for promotion of neuroplasticity [31,34,35]. BMI is used for detecting patterns of MI or an intention of movements in a brain signal and makes the feedbacks when it the patterns is detected. The feature patterns in the brain signal according to MI and the intention of movement are thought to be similar [36]. As we mentioned, the application areas of MI-BMI are widespread, and classification of brain states of MIs is an important issue in these applications.

Different MI tasks (e.g. imaginations of movements in a left hand, a right hand, feet, and a tongue) evoke different feature patterns in EEG [3, 37]. Especially, the change in energy of a certain frequency band called mu band [8] is well known as a feature related to MI. The mu band is the name of the frequency band of 10–15 Hz in EEG and MEG signals and there are some bands named such as alpha, beta, and theta bands (waves) in brain science [38]. The energy in the mu band observed in a motor cortex decreases while a subject performs the MI tasks. This decrease of the energy is called event related desynchronization (ERD) and

the location observing ERD depends on the body part of which a subject imagines the movement [8, 39–41]. Extraction of these changes from the measured EEG signals enables us to classify the EEG signal associated with the different MI tasks.

1.2 Problems

For classification of some MI tasks, we have to extract the features in the presence of measurement noise and spontaneous components which are related to the other brain activities because the features, noise and the other components are mixed in the observed EEG signals. For the extraction of the features, signal processing techniques such as bandpass filtering and spatial weighting are used [3]. For the processing, finding the parameters such as coefficients of the filters and weights that accurately extract the related components is a crucial issue. The knowledges in neuroscience have been often used to design the parameters. However, the optimum parameters in classification are highly dependent on users and measurement environments [42]. In order to determine the parameters, data-driven techniques that exploit observed data are widely used [3, 8]. The observed data essentially include class labels corresponding to the tasks. The dataset of the data with the class labels refers to a learning dataset in machine learning literacy. The data-driven techniques try to find the parameters that extract discriminative features as much as possible. To this end, it would be simple to apply cross validation (CV) for searching these parameters that give the best classification accuracy for the learning dataset among several candidates of the parameters. However, the classification performance depends on the choice of the candidates and the selected filter is not always "optimal." Moreover, a large size of the candiate set leads to

high computational cost.

As the data-driven methods, common spatial pattern (CSP) method is widely used for design of the spatial weights in MI-BMIs. The method finds the spatial weights by using the observed signals [3, 42, 43] in such a way that the variances of the signal 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 [44–49], and methods to select the CSP features extracted with various parameters [50, 51]. The feature extraction procedures in these methods can be modeled by a model applying spatial weights, frequency filters, and time windows to EEG signals.

The study focuses on three problems regarding to the design of the parameters in the model;

1) simultaneous design of spatial weights, frequency filters, and time windows is impossible,

2) extraction of multiple feature components modeled with different spatial patterns, frequency patterns, and time patterns for each other is impossible,

3) overfitting of the designed parameters for the spatial weights often happens.

1.3 Solutions

We propose a method to solve the problems 1) and 2) in Chapter 4 and a method to solve the problem 3) in Chapter 5.

Chapter 4 proposes common spatio-time-frequency patterns (CSTFP). This method is one of the extensions of CSP. The different points from the existing

methods, which design the spatial weights and frequency filters, are that CSTFP can simultaneously design sets of the spatial filter, the frequency filter, and the time window. The reasons of the importance of design of the sets of the filters are described in the following two paragraphs.

The first reason relates to the problem 1). A kind of BMIs is implemented based on cues which a user follows. In this 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, when the brain activity associated to 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 milliseconds after the cues are assumed not to be used to extract the features in previous works [46, 50, 52, 53], which heuristically determined the time window. Contrary to these works, this study 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 [54]. The reaction time can be related to the time to response to the cues. Therefore, the time window should also be designed by using the observed signals or data-driven.

The second reason relates to the problem 2). It is suggested that the feature components evoked by performing the MI tasks are observed in the mu and beta bands [37]. Therefore, a bandpass filter with a passband of 7–30 Hz including the mu and beta bands is widely used in feature extraction in MI-BMI [42, 43, 55]. These two components in the mu and beta bands are considered to have different frequency bands, different spatial patterns, and different time patterns [3,

9]. Because of the suggestion about the mu and beta bands, we suggest that the features associated to the MI task are in various patterns. Therefore, multiple filter sets that separately extract each patterns are needed.

Moreover, the CSTFP method does not have the following problems that the existing methods have. Common spatio-spectral pattern (CSSP) [44] and common sparse spectral spatial pattern (CSSSP) [45] 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. Unlike CSSP that provides different spectral patterns for each channel, CSSSP can provide a common spectral pattern for all channels. However, CSSSP needs the computationally expensive optimization because the optimization problem for the filter involves a sparsity cost and an extensive parameter tuning is needed. Spectrally weighted CSP (SPEC-CSP) [46] and iterative spatio-spectral patterns learning (ISSPL) [56] use iterative procedures for optimizing spatial weights and a filter, 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 other criteria. These two optimization problems are alternately solved, however this alternating iteration is not guaranteed to be converged, because the cost functions for these two problems are different. Recently proposed sub-band CSP (SBCSP) [57] and filter bank CSP (FBCSP) [50] are based on subband decomposition with multiple filters that have different passbands. As reported in BCI competition III [29] and IV, FBCSP achieves a high classification accuracy. In these methods, EEG signals are decomposed into multiple frequency components and CSP is separately applied for each frequency component. By using multiple filters such as these methods, fea-

tures associated to brain activities such as mu and beta rhythms can be separately extracted. The result of FBCSP suggests that feature extraction using subband decomposition is an effective method to increase classification accuracy of MI-BCI. The methods using frequency decomposition need selection of feature values to achieve accurate classification. In order to select the feature values, the methods based on mutual information have been proposed [50, 51]. The other methods for selecting efficient filters also proposed in [58] and [59]. In these methods, the filters in the filter bank have to be designed in advance and the fixed filter bank is independent of dataset. Although the parameters for the filters such as the passbands that accurately extract the feature components can change by users and measurement environments, the method using the fixed filter bank cannot use the filters optimized for a specific user.

Chapter 5 proposes a regularization method to prevent overfitting in the optimization of the spatial weights. The data-driven methods for design of the spatial weights like CSP need a large amount of the learning data. Collecting enough amount of the learning data takes long time and a subject will feel fatigue. Therefore, the optimization can be ill-posed and the optimized parameters can be overfitting because of limited numbers of the learning data. To prevent overfitting or to solve an ill-posed problem in signal processing and machine learning for learning parameters, regularization is widely used [60, 61]. The regularization for an optimization problem is to add to an original cost function a penalty term which represents additional information such as smoothness or bounds of the vector norm of parameters to be optimized. In this way, the regularization can help design more robust spatial weights against ill-posed problems [53]. We propose a regularization to add information of an EEG electrode arrangement to an opti-

mization problem. The regularization is proposed based on the assumption: the signals measured by the electrodes that are located near each other (the nearby sensors) are similar and also the observed components are similar. To describe the assumption, consider a measurement device of EEG where electrodes installed on scalp observe faint electrical difference. The EEG signal reflects the summation of the synchronous activity of thousands or millions of neurons [9,62]. Therefore, the nearby sensors likely observe activities which are induced from the same neurons. For the reason, the spatial filters such as the Laplacian filter that averages the signals observed in the nearby sensors are often used for improving SNR in EEG signal processing [8]. Based on the assumption, we propose the regularization that works in the optimization problems such that the signals weighted by each coefficient of the spatial weights are similar to each other if these signals are observed in the nearby sensors.

The CSTFP method and the regularization are evaluated in their performances by experiments using artificial signals and BMI datasets in each chapter.

1.4 Organization of Thesis

This thesis is divided into six chapters. In Chapter 1, the background regarding this work, the problems, the ideas of the proposed method is discussed. Chapter 2 introduces the foundation of BMIs. Chapter 3 describes feature components in EEG associated with the MI tasks and some methods to extract these features. We introduce the CSTFP method and its experimental results in Chapter 4. The regularization with information of sensor arrangement is introduced Chapter 5. The study is concluded in Chapter 6.

1.5 List of Symbols

- $\mathbb{R}$: the set of all real numbers
- $\mathbb{C}$: the set of all complex numbers
- $\mathbb{Z}$: the set of all integers
- An italic letter: a scalar, e.g. x and X
- An italic bold lower-case letter: a scalar, e.g. $\boldsymbol{x}$
- $[\cdot]_i$: the ith entry of a vector
- An italic bold upper-case letter: a matrix. e.g. $\boldsymbol{X}$
- $[\cdot]_{i,j}$: the ith row and the $j$the column of a matrix
- $\cdot^T$: transposes for a vector and a matrix, e.g. $\boldsymbol{x}^T$ and $\boldsymbol{X}^T$
- $|\cdot|$: the absolute value of a scalar, e.g. $|x|$
- $\|\cdot\|_p$: the p-norm of a vector defined as $\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{N} |[x]_i|^p\right)^{1/p}$
- $\|\cdot\|$: the l_2-norm of a vector
- Span$\{\cdots\}$: the subspace spanned by vectors
- $\oplus$: an operator to give the direct sum of two subspaces
- $\cdot^{\perp}$: the ortognal complement of a subspace
- δ_{ij}: the Kronecker delta defined as 1 for $i = j$ and 0 otherwise
- a mod b: an operator to take a residue of dividing a by b.

- $\Re$: an operator to take a real part of a complex number.
- $\mathcal{N}(m, \sigma^2)$: a Gaussian distribution with a mean, m, and a variance, σ^2
- $\mathcal{U}(a, b)$: a uniform distribution whose minimum and maximum values are denoted by a and b, respectively

Chapter 2

EEG-Based Brain Machine Interfaces

In this chapter, we describe EEG-based brain machine interfaces. We describe fundamental structures of BMI in Sec. 2.1. Section 2.2 lists some methods for measuring brain activities of human. Section 2.3 shows principal paradigms for EEG-based BMI.

2.1 Fundamentals of Brain Machine Interfaces

BMIs are interfaces directly connecting brains and external worlds without muscle movements of hands, feet, and a tongue. Figure 2.1 shows a fundamental structure of BMI [9, 14]. As shown in the figure, a procedure of BMI system can be divided into the following four steps.

Step 1 A subject performs certain tasks to induce brain activity in its brain. The certain tasks are mental tasks such as imagination of something or turning attention to a visual/auditory/tactile stimulus.

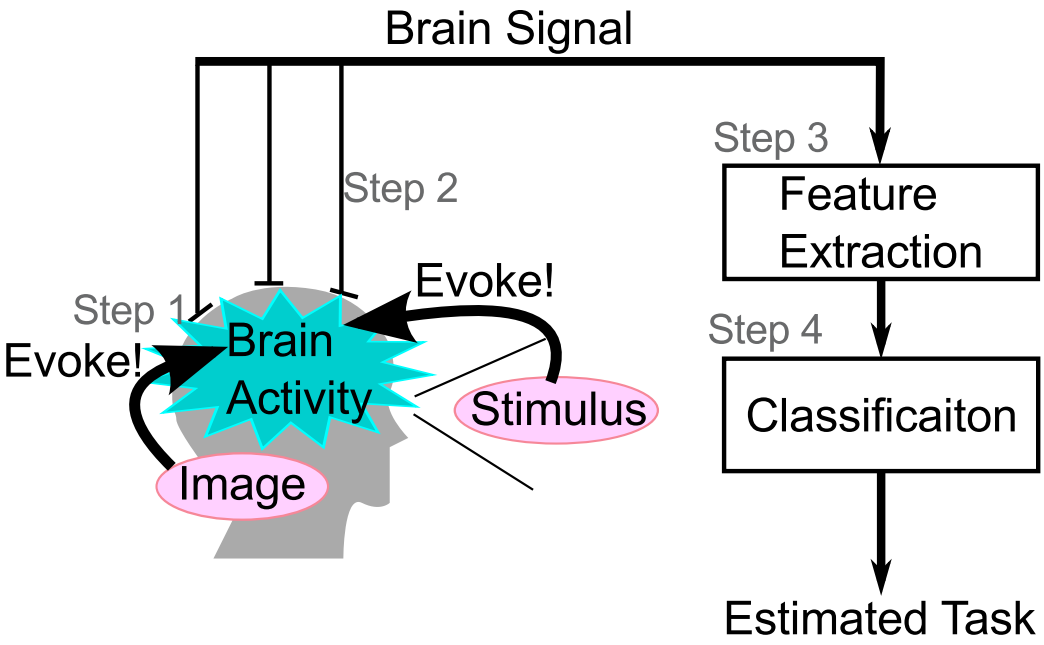


Figure 2.1: A fundamental structure of BMI

Step 2 Brain signal is acquired by measurement systems. Because subsequent signal processing is applied to the digitized signal, the acquired signals are converted to the digital signal by an A/D converter after amplification and filtering are applied.

Step 3 Feature components are extracted from the signal. The signal mixed with components associated to various brain activities and noise are observed. In this step, we remove the noise and extract the components that are associated with the task and are important for the estimation of the tasks.

Both of classical and statistical signal processing techniques (e.g. discrete Fourier transform, FIR/IIR filtering, averaging, principal component analysis (PCA), regression analysis, and independent component analysis (ICA)) [63, 64] are used for extracting the feature components. Depending on the tasks,

the feature components to be extracted are different. Therefore, we should adopt suitable techniques for this step [8, 65].

Step 4 The task that the subject performs is estimated by recognition of the extracted feature components. The feature components extracted in Step 3 are given as a vector. Hence we can use pattern recognition techniques that are used in the research areas such as speech recognitions and image recognition. Classifiers with machine learning (e.g. linear discriminant analysis, support vector machine, artificial neural network, and Bayesian classifier) [61, 66–68] are widely used for recognizing the tasks with the feature value (vector).

Through these steps, we obtain the task that the subject is performing. The recognized task is used for analysing the subject's emotions and brain state or is converted to a command to control the device. The BMI system used for inputting the device commands is called *an output type of BMI*. Because the output type of BMI does not need any muscle movements to send the commands to the devices, the output type of BMI is expected as a communication tool for patients suffering from motor disorders caused by amyotrophic lateral sclerosis, brain stroke, brian or spinal cord injury, cerebral palsy, muscular dystrophies, multiple sclerosis, and numerous other diseases [9]. BMI is also expected as an interface for virtual reality systems and video games [69]. The application of BMIs spreads to the area of rehabilitation for patients suffering from motor disorders caused by a stroke [30, 31, 33, 70].

2.2 Methods for Measuring Brain Activity

Neurons composing a brain generate spikes (change of electrical potential at short-term) when they accept input stimulus and convey information to the other neurons [71]. We capture this kind of changes of electrical potentials and/or blood flows by measurement devices. The information of these changes is used for estimation of brain state. In this section, we introduce some of the measurement devices. In this study, we focus on EEG with scalp electrodes. This section also shows reasons why we focus on this devices.

2.2.1 Functional Magnetic Resonance Imaging

Functional magnetic resonance imaging (fMRI) is one of neuroimaging procedures using magnetic resonance imaging (MRI) [72]. When neuron activates, oxygen is consumed and the blood flow increases in the capillary around the neuron [73]. FMRI measures brain activity in the brain or spinal cord of humans or other animals by detecting associated changes in the blood flow.

FMRI is able to obtain high spatial resolution images. However, the change of blood flow is slower than neuron activations and the temporal resolution is poor.

2.2.2 Near-Infrared Spectroscopy

Near-infrared spectroscopy (NIRS) is a spectroscopic method [74]. NIRS uses the near-infrared spectrum (about 800 nm) that reaches inside of a skull through a scalp and a skull. By measuring the reflected light, we can observe the change of amount of hemoglobin and information of exchange of oxygen in the brain.

2.2.3 Magnetoencephalography

Magnetoencephalography (MEG) is a imaging technique that uses superconducting quantum interference device (SQUID) [75]. SQUID measures magnetic field generated by electrical activities of brain neurons.

2.2.4 Electroencephalography

Electroencephalography (EEG) is a measurement device observing electrical activity of neuron by using electrodes [76]. The electrical activity that is observed by an electrode is the summation of activities of the group of neurons around the electrode. The electrodes are located on scalp, on cerebral cortex, under cerebral cortex, and in ventricle and the measurement methods are called scalp EEG, electrocorticography (ECoG), electrosubcorticography (ESCoG), and Electroventriculography (EVG), respectively. The scalp EEG is an noninvasive measurement method. ECoG, ESCoG, and EVG are invasive measurement methods. This study focuses on the scalp EEG and addresses problems in signal processing for the signal observed by the scalp EEG systems. In the remaining of the thesis, we call the scalp EEG simply EEG unless otherwise specified.

In the EEG measurement, the electrodes are installed on the scalp by using conductivity pastes and special caps. For the BMI application, the EEG signal is usually observed by multiple electrodes [9, 14, 37]. In the case of the multi-electrode measurement, International 10-20 [62], Extended 10-20 [77], International 10-10 [78], and 10-5 [79, 80] methods have stood as the de-fact standard of the electrode arrangement. In these systems, locations on a head surface are described by relative distances between cranial landmarks over the head surface. For an example, we show how to decide the locations of the electrodes according

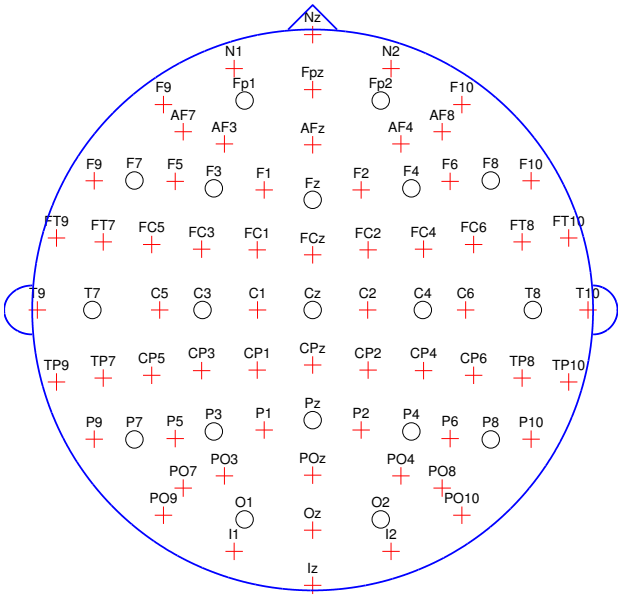


Figure 2.2: The electrode arrangement of the International 10-20 method and the International 10-10 method. The circles show the electrodes defined by the International 10-20 method. The circles and the crosses show the electrodes defined by the International 10-10 method.

to the International 10-20 method. The landmarks are a nasal point located in the middle point of eyes, the height of eyes, and a inion which is the most prominent point at the back of the head [62]. We drew the line connecting these landmarks along the head surface and moreover make marks that divide the line into short lines of 10% and 20% length of the line. Each electrode point is defined at the intersection point of the lines connecting these marks along the head surface. An examples of the electrode arrangement are shown in Fig. 2.2.

The EEG systems record the electrical potential difference between a two electrodes [62]. Referential recording is a method recording the electrical potential difference between a target electrode and a reference electrode and the reference

electrodes are common to all electrodes. Earlobes that are considered to be electrically inactive are widely used as the reference electrode. In contrast, bipolar recording records the electrical potential difference between electrodes located on head surface.

The EEG system can be more compact than the system of fMRI, NIRS, and MRI. Temporal resolution of EEG is higher than fMRI and NIRS [3]. However, it is difficult to make the electrodes for EEG smaller and spatial resolution of EEG is very low comparing with other devices. Moreover, the high frequency component of electrical activity of neuron decreases in the observed signal because the barriers such as a skull work as a lowpass filter. Additionally, noise caused by poor contacting of the electrode, muscle movements (EMG; electromyography) and eye movements (EOG; electrooculography) contaminates the signal.

The tasks that induce discriminative features in the scalp EEG signal and are available for BMI are limited. The typical tasks for BMI are;

- concentration and meditation,
- perception of random-displayed stimuli,
- gazing at flickering stimulus,
- imagination of muscle movement.

The details of these BMIs are shown in Sec. 2.3.

2.3 Variety of EEG-based Brain Machine Interfaces

2.3.1 Perception of Random-Displayed Stimuli

BMI using perception induces a brain activity called event-related potential (ERP) in a brain by perception to a stimulus. This type of BMIs is widely studied for letter input [3]. The BMI has achieved the input speed of around 1.5 letters per minute [23, 81, 82].

ERPs are electrical changes in EEG signal and they occur when a human perceives something. P300 is one of the ERPs. P300 is observed as a positive change of potential after 300 ms from the perception of stimulus [83]. Visual, auditory, tactile, olfactory, and gustatory stimuli are available for inducing P300. An example of the waveform of P300 is shown in Fig. 2.3. We use the dataset, Dataset II that is an open data in BCI Competition III [28, 84, 85]. The signal labeled "Targets" is the averaged signal observed in the period 0–1 secs after displaying a stimulus when the subject perceived the stimulus. The signal labeled "NonTargets" is the averaged signal observed in the period 0–1 sec after displaying a stimulus when the subject does not perceived the stimulus. We can observe the difference of the potentials between "Targets" and "NonTargets" in the period of 0.2–0.4 sec.

In particular, the task called oddball paradigm can induce high potential of P300 [86, 87]. In the oddball paradigm, the subject is asked to react either by counting or by button pressing incidences of target stimuli that are hidden as rare occurrences amongst a series of more common stimuli, that often require no response. As BMI with the oddball paradigm, BMIs making input of letters have been proposed [3, 88–90]. This kind of BMIs is called P300-spellers. As an

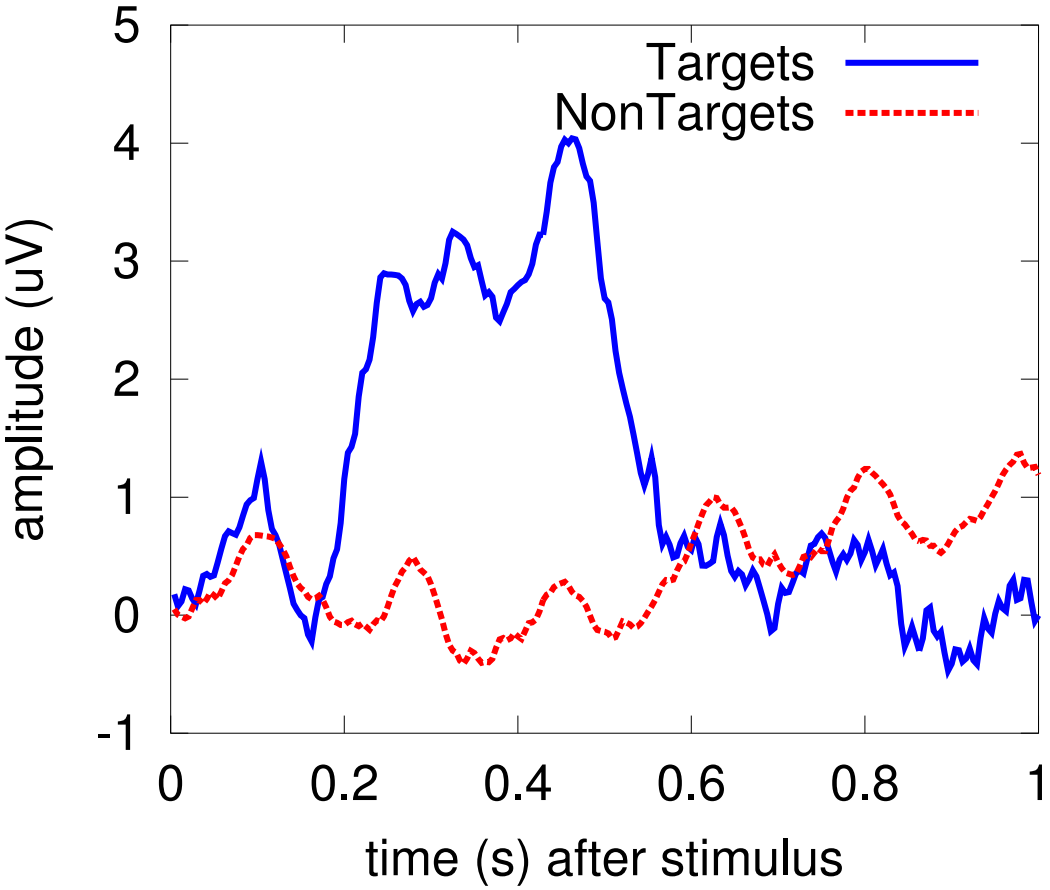


Figure 2.3: The waveform of P300. The signals averaged over 85 trials. "Targets" is the observed signal with the subject's perception after displaying a stimulus. "NonTargets" is the observed signal without the subject's perception.

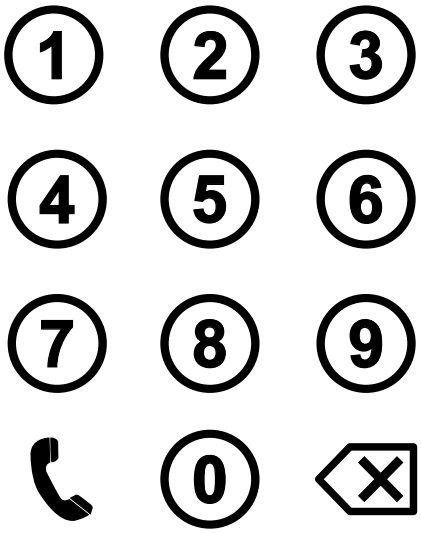


Figure 2.4: An Interface of a BMI system making a telephone call.

example for a BMI using the oddball paradigm, we show an interface of a BMI system making a telephone call. The interface to be presented to a user by a monitor is Fig. 2.4. Each row and line flashes during 50–500 ms in random order and the interval between successive flashes is 500 to 1000 ms [89]. The procedure for displaying flash stimuli is illustrated in Fig. 2.5. The user gazes at a target symbol in the interface and counts incidences of the flash of the target symbol. Because P300 happens after 300 ms of the flash of the target, BMI detects whether P300 happens or not for each flash. The BMI decides which symbol the user wants to input.

For detection of P300, recording some trials to input one symbol is often needed due to low signal to noise ratio [88]. Averaging over some trials, low-pass filtering, stepwise linear discriminant analysis (SWLDA) [23, 91, 92] based on linear discriminant analysis (LDA) [61, 93] are used as signal processing for this type of the BMIs.

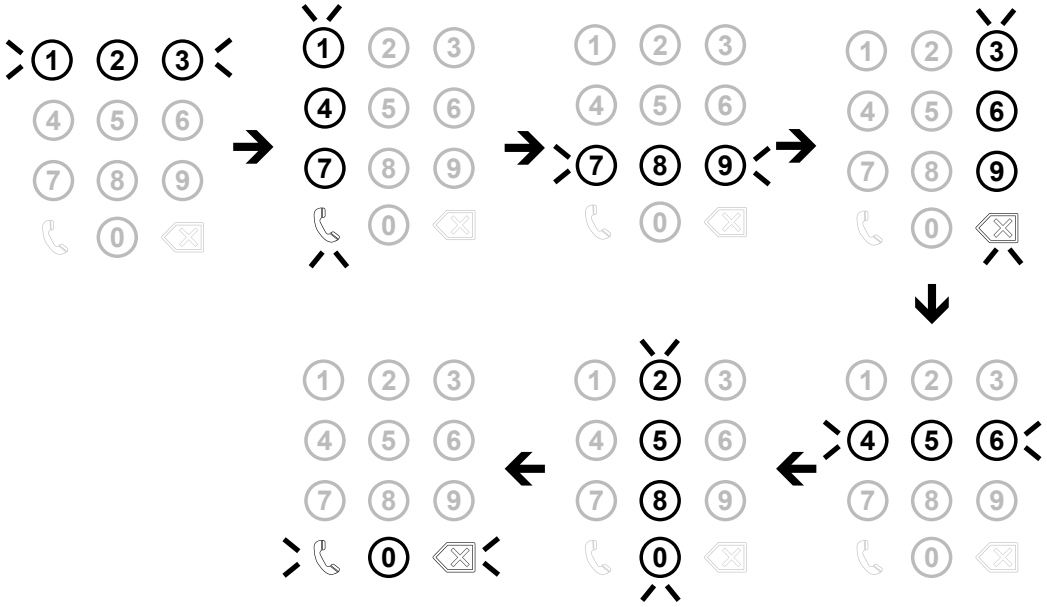


Figure 2.5: The procedure for the stimuli in the telephone call BMI. According to the order indicated by the arrows, each row and line flashes.

2.3.2 Gazing at Flicker Stimuli

BMIs using flicker stimuli is an interface where a user inputs a command by gazing at a stimulus flickering at flash frequency of 3–70 Hz [94–96]. This type of BMI does not realize many number of the commands, but achieves higher input speed [3, 14]. However, the flicker stimuli can cause attacks of epilepsy and you should be careful for use [97, 98].

The BMI systems use light emitting diodes (LEDs) [99] or monitors for displaying the flicker stimuli [100]. When the subject gazes at the stimuli flickering at 3–70 Hz, steady-state visually evoked potential (SSVEP) is observed in an EEG signal recorded at a visual cortex [98]. Figure 2.6 shows the power spectra of the signals observed when the subject had gazed at the stimuli flickering at 8 Hz for

15 seconds and when the subject had not gazed at it. When the subject gazes at the stimulus, the power spectrum has a peak at 8 Hz. Additionally, there are the other steady state evoked potentials evoked by turning attention to the repeat of a short-term stimulus such as a flicker. The repeat of tactile [101] and auditory [102, 103] stimuli can evoke such potentials.

BMI paradigms using SSVEP are as follows. As flicker stimuli, we arrange some checkerboards as in Fig. 2.7 on a monitor. We make monochrome inversion (Fig. 2.8) in the checkerboards at different frequencies for each checkerboard. When the user gazes at the left-top stimulus in Fig. 2.7, SSVEP of F_1 Hz happens in a recorded EEG signal. By estimating the frequency of SSVEP, BMI decides the stimulus which the user gazes at. Taking a simple way, we can assign the commands to each flicker stimulus and design BMI having six commands with an interface of Fig. 2.7.

The SSVEP-based BMIs have following problems. First, the power of SSVEP is very weak when the flicker frequency is more than 25 Hz [104]. Second, SSVEP of harmonics frequency of a gazed stimulus are also evoked. Therefore, the use of harmonics frequency of a stimulus as the other stimulus frequency can let classification accuracy decrease [100, 105]. Third, in the case of the use of a monitor for displaying the stimuli, available flicker frequencies depend on the refresh rate of the monitor. Due to the above reasons, it is difficult to realize many commands.

To detect SSVEP frequency, discrete Fourier transform can be used. And, canonical correlation analysis (CCA) [106, 107] is also widely used for processing for multi-channel EEG signal [108].

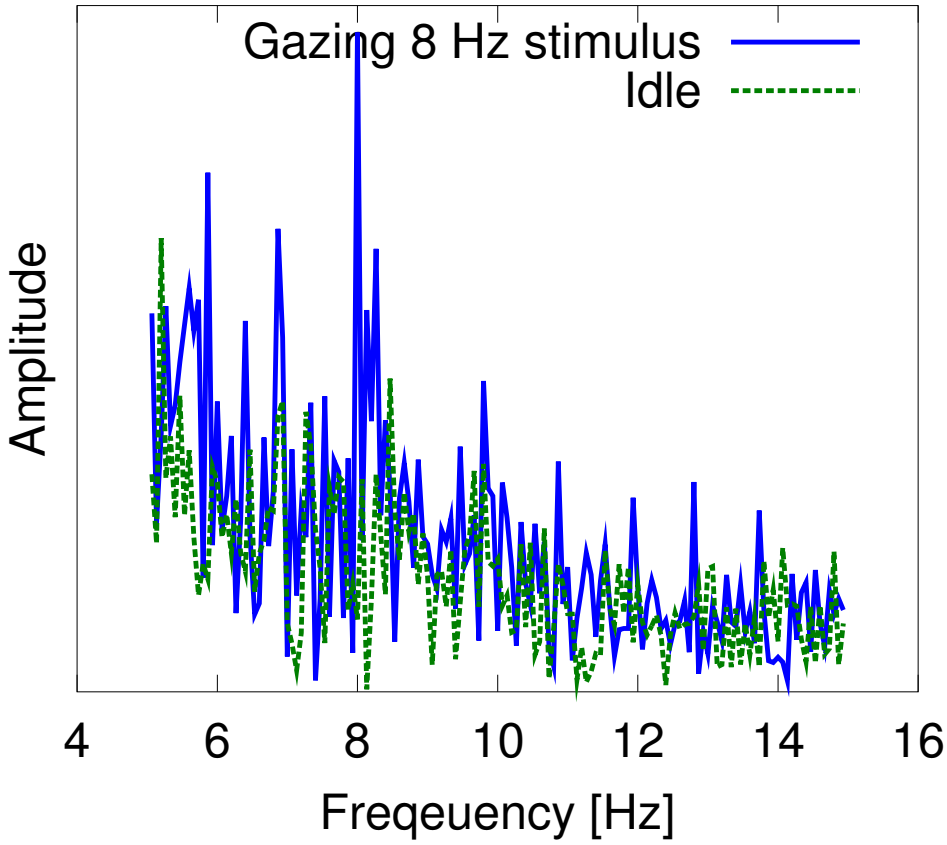


Figure 2.6: The power spectrum of the signals observed when the subject gazes at a flicker stimulus for 15 seconds.

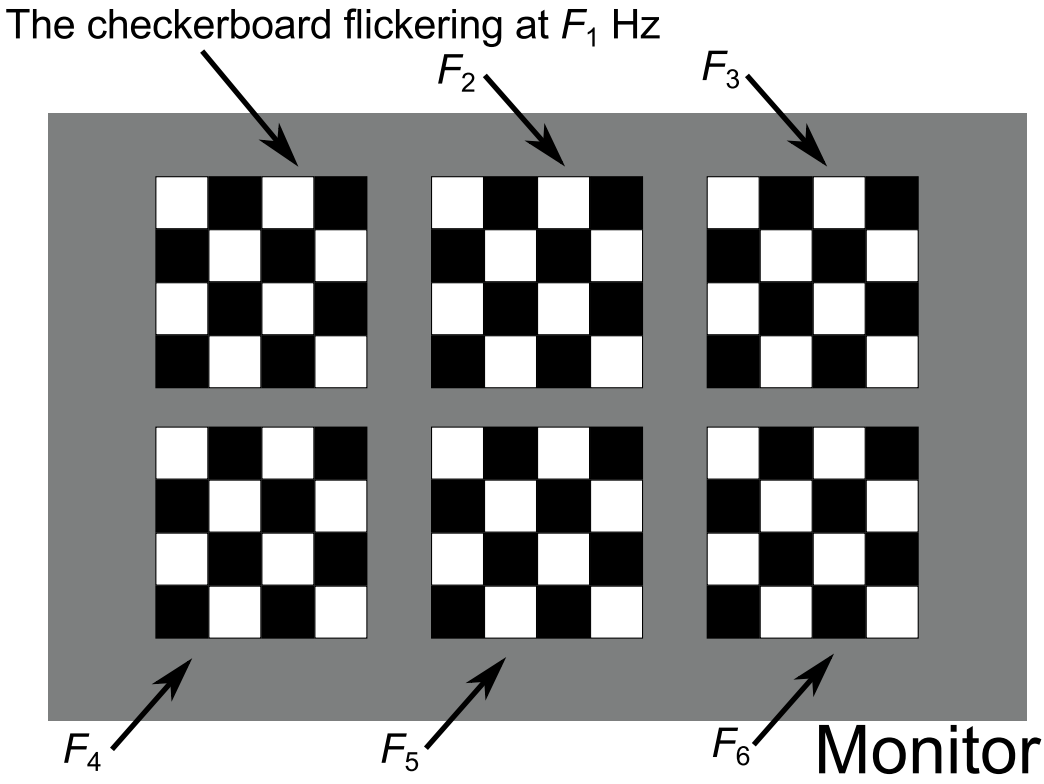


Figure 2.7: An interface for SSVEP-based BMI. Each checkerboard makes monochrome inversion at different frequency.

2.3.3 Imagination of Muscle Movements

BMIs using imagination of muscle movements is an interface where a user inputs a command by imaging movement (motor imagery) of body parts such as hands, feet, and tongue. [3]. This type of BMIs is called the motor-imagery based BMI (MI-BMI). As an example of the MI-BMIs, a task of the motor imagery of right hand and a task of the imagination of the motor imagery of left hand are the tasks to be classified from brain signals.

The tasks of the motor imagery decrease the energy in the certain bands called mu (8–15 Hz) and beta (10–30 Hz) bands in EEG signals observed by the elec-

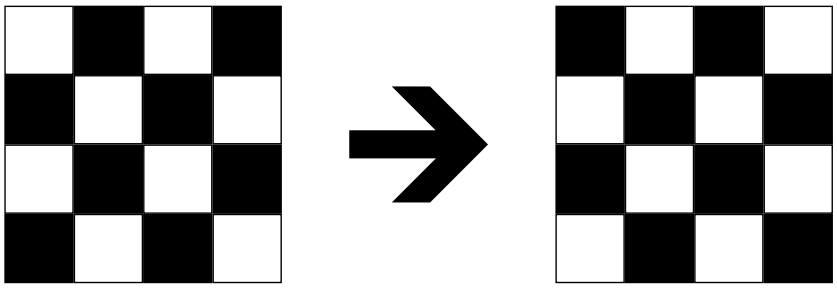


Figure 2.8: Monochrome inversion in a checkerboard.

trodes located on (sensory) motor motor cortices. The decrease of the energy is called event related desynchronization (ERD) [8, 39]. Moreover, the location observing ERD depends on the body part of which a subject imagines the movement [39–41]. Therefore, we can estimate which body part the subject imagines the movement of from EEG signals by detecting the location where ERD happens. ERD can be observed in not even healthy subject and paralyzed patients. ERD is induced by motor imagery tasks performed by healthy subjects and also paralyzed patients [109]. An example of ERD we can observe in EEG signals is shown in Fig. 2.9. The EEG signals were recorded while a subject performed the tasks of motor imagery of left and right hands [110]. After the recording, the EEG signals that is bandpass filtered with 8–30 Hz. Then these squared signals shown in Fig. 2.9 are averaged over the 100 trials for each motor imagery task of the left or right hand and normalized by a base energy which is averaged over squared signals observed before performing motor imagery tasks. The symbol 0 on the horizontal axis is the time when the subject started performing the tasks. We can observe a decrease in the energy while the subject performs the tasks.

One of the merits of MI-BMI against the BMIs using perception or flicker stimuli is that an instrument for displaying stimuli is unnecessary. Moreover, it

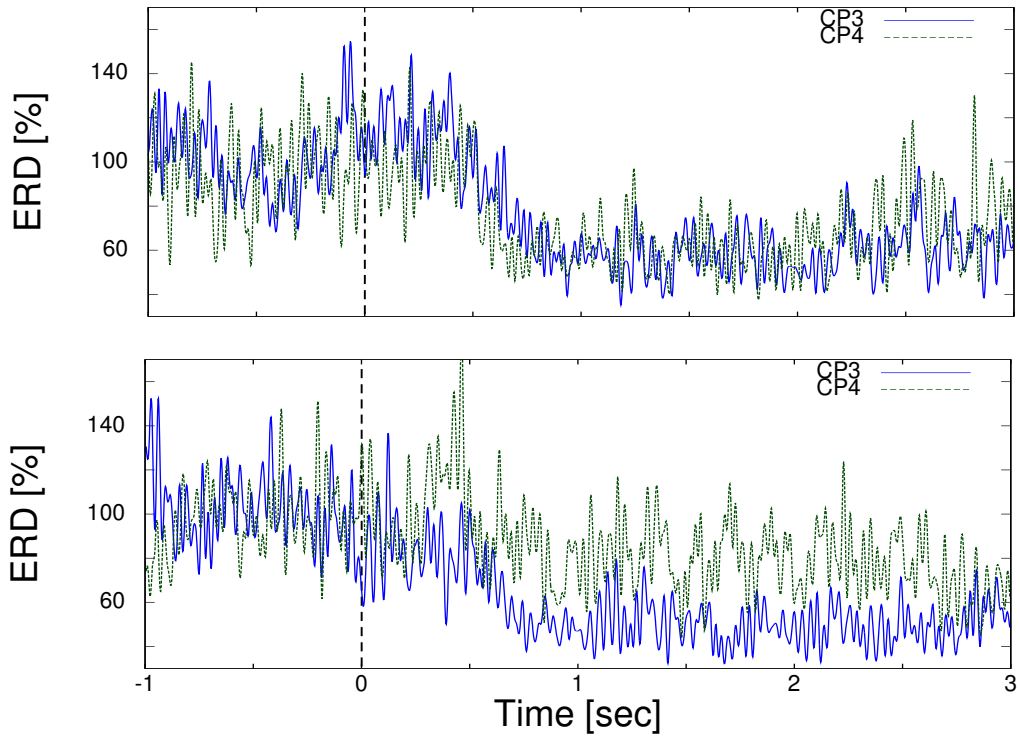


Figure 2.9: ERD by the motor imagery tasks of left (top) and right (bottom) hands.

has been reported recently that the detection of the motor imagery tasks and a feedback is useful for rehabilitation for patients who suffer from motor disorders caused by brain damages [30, 31, 33, 70]. MI-BMI will be utilized in the rehabilitation to recover of the motor functions as follows. One of the rehabilitation procedures for the recovery of the motor functions is to make a subject perform movements of a disabled body part by a cue and give a visual or physical feedback [30]. The rehabilitation takes advantage of the plasticity of the brain. The plasticity of the brain is an intrinsic property. The plasticity enables the nervous system to adapt the environmental pressures, physiologic changes, and experiences [111]. The coincident events of the intention of the movement that the subject has and the corresponding feedback to the subject are supposed to promote the plasticity of the brain. Because of the plasticity of the brain, the other parts of the brain with no damage take over the functions disabled by the damages. In the rehabilitation, to generate the feedback coincidentally with intention is considered to be significant. In the general procedure of the rehabilitation illustrated in above, the cues control generating the intention and the feedback. Using the MI-BMIs for the rehabilitation, the feedback generation is controlled by BMI. MI-BMI enable the rehabilitation system to detect the intention of the movements and generate the feedback at an appropriate time. The some researches have suggested that the rehabilitation with the MI-BMIs can promote the plasticity of the brain more efficiently than the conventional system based on the cue [30, 31, 33, 70].

Chapter 3

Feature Extraction for Motor-Imagery EEG

In this chapter, we illustrate feature patterns observed in EEG associated with the MI tasks in more detail than Sec. 2.3.3. Next, we show some of methods to extract these features. Especially, we show the CSP method [42, 43] which is a well-known method for the feature extraction and classification in two class MI-BMI and its variants [44–46, 50].

3.1 Features Associated with Motor-Imagery Task

As we show in Sec. 2.3.3, the features of the MI tasks are observed as the energy changes of certain frequency bands. Moreover, spatial patterns of the changes depend on the kind of the MI tasks [8]. For example, it is said that the MI of a left hand evokes energy changes in the right side of a motor cortex. The MI of a right hand evokes energy changes in the left side of a motor cortex.

The differences of the spatial pattern of the features are caused by differences

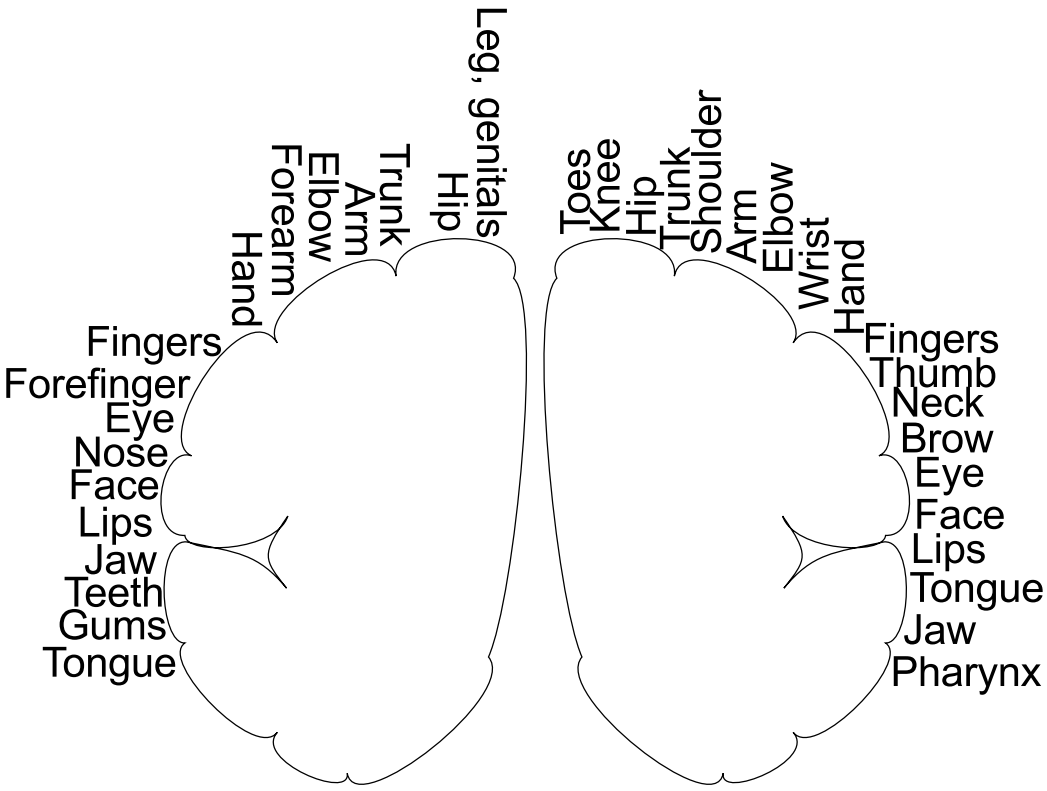


Figure 3.1: Sensory and motor cortex and associated body regions

of brain regions that work for each muscle. Figure 3.1 illustrates which region works for a body region [112].

As shown in Fig. 2.9, the differences in ERD between the different MI tasks are observed. By finding in spatial patterns of ERD, we can associate an observed EEG signal with the MI tasks which a subject performs.

3.2 Discrete Fourier Transform

The features of the MI are observed in frequency domain. Therefore, the signal transformed to the frequency domain can be used as useful features. Discrete Fourier transform (DFT) is the simplest way to obtain the signal in frequency

domain. Let $\boldsymbol{x} \in \mathbb{R}^N$ be a signal observed by an electrode, where N is the number of discrete samples. DFT of $\boldsymbol{x}$ is represented as

$$[\boldsymbol{f}]_j = \sum_{k=1}^{N} [\boldsymbol{x}]_k \exp\left(\frac{-2\pi i(j-1)(k-1)}{N}\right), \quad j = 1, \ldots, N, \tag{3.1}$$

where i denotes an imaginary unit defined as $i^2 = -1$.

For BMI use, multichannel EEG measurement systems in which many electrodes located in different places are used are widely used for improving SNR. Therefore, in case of forming a feature vector with some DFTs, the dimension of the feature vector is large and it causes some problems such as overfitting and an ill-posed problem in classification process.

3.3 Common Spatial Pattern

CSP is a spatial filter. In the definition in the study, the spatial filter works as

$$y = \sum_{m=1}^{M} w_m x_m, \tag{3.2}$$

where, x is a sample observed in the mth channel at a discrete time, w_m is a weight coefficient for the signal observed in the mth channel, y is the spatial-filtered signal at the time. The CSP is denoted the vector as $\boldsymbol{w} = [w_1, w_2, \ldots, w_M]^T$.

CSP is found the spatial filter, $\boldsymbol{w} \in \mathbb{R}^M$, in such a way that the variance of a signal extracted by linear combination of $\boldsymbol{X}$ is minimized in a class [42, 43]. 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 components 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

$$\boldsymbol{\mu} = \frac{1}{N}\sum_{n=1}^{N}\hat{\boldsymbol{x}}_n. \tag{3.3}$$

Then, the time variance of the extracted signal of $\hat{\boldsymbol{X}}$ is given by

$$\sigma^2(\boldsymbol{X}, \boldsymbol{w}) = \frac{1}{N}\sum_{n=1}^{N}|\boldsymbol{w}^T(\hat{\boldsymbol{x}}_n - \boldsymbol{\mu})|^2. \tag{3.4}$$

We assume sets of the learning data, C_1 and C_2, where C_d contains the signals belonging to class d, $d \in \{1, 2\}$ is a class label, and $C_1 \cap C_2 = \emptyset$. CSP finds the weight vector that minimizes the intra-class variance in C_c under the normalization of samples, where c is a class label. More specifically, for c fixed, CSP finds $\boldsymbol{w}_c$ by solving the following optimization problem [42, 43];

$$\begin{aligned} \min_{\boldsymbol{w}} \quad & E_{\boldsymbol{X}\in C_c}[\sigma^2(\boldsymbol{X}, \boldsymbol{w})], \\ \text{subject to} \quad & \sum_{d=1,2} E_{\boldsymbol{X}\in C_d}[\sigma^2(\boldsymbol{X}, \boldsymbol{w})] = 1, \end{aligned} \tag{3.5}$$

where $E_{\boldsymbol{X}\in C_d}[\cdot]$ denotes the expectation over C_d. Then, (3.5) can be rewritten as

$$\begin{aligned} \min_{\boldsymbol{w}} \quad & \boldsymbol{w}^T\boldsymbol{\Sigma}_c\boldsymbol{w}, \\ \text{subject to} \quad & \boldsymbol{w}^T(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)\boldsymbol{w} = 1, \end{aligned} \tag{3.6}$$

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

$$\boldsymbol{\Sigma}_d = E_{\boldsymbol{X}\in C_d}\left[\frac{1}{N}\sum_{n=1}^{N}(\hat{\boldsymbol{x}}_n - \boldsymbol{\mu})(\hat{\boldsymbol{x}}_n - \boldsymbol{\mu})^T\right]. \tag{3.7}$$

The solution of (3.6) is given by the generalized eigenvector corresponding to the smallest generalized eigenvalue of the generalized eigenvalue problem described as

$$\boldsymbol{\Sigma}_c\boldsymbol{w} = \lambda(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)\boldsymbol{w}. \tag{3.8}$$

Although the solution of (3.6) is given by the eigenvector corresponding to the smallest eigenvalue in (3.8), we can use the other eigenvectors for classification [55]. The M eigenvectors can be obtained by solving (3.8) as $\hat{\boldsymbol{w}}_1, \ldots, \hat{\boldsymbol{w}}_M$, where $\hat{\boldsymbol{w}}_i$ is the eigenvector corresponding to the ith smallest eigenvalue of (3.8). 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

$$\begin{aligned}\boldsymbol{y} =&[\sigma^2(\boldsymbol{X}, \hat{\boldsymbol{w}}_1), \ldots, \sigma^2(\boldsymbol{X}, \hat{\boldsymbol{w}}_r),\\ &\sigma^2(\boldsymbol{X}, \hat{\boldsymbol{w}}_{M-r+1}), \ldots, \sigma^2(\boldsymbol{X}, \hat{\boldsymbol{w}}_M)]^T. \end{aligned} \tag{3.9}$$

For classification, $\boldsymbol{y}$ is an input to a classifier such as linear discriminant analysis (LDA) [61].

3.4 Common Spatio-Spectral Pattern

CSSP is a method where a weight vector is obtained by applying the combination of observed signals with time-delayed signals to CSP [44]. 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 sample delay, 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

$$\boldsymbol{X}^\tau = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix} \in \mathbb{R}^{2M\times(N-\tau)}. \tag{3.10}$$

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 (3.8) using $\boldsymbol{X}^\tau$. We compose the feature vector, $\boldsymbol{y} \in \mathbb{R}^{2r}$, in a way similar to (3.9).

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 [44]. 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.

3.5 Common Sparse Spectral Spatial Pattern

The CSSSP [45] method obtains a weight vector and an FIR filter by using delayed signals like CSSP. The main difference between both methods is that CSSSP achieves design of an FIR filter with order more than 2. When CSSSP designs an FIR filter with order T, the optimization problem of CSSSP has the optimization parameters consisting of M spatial weight coefficients and $T - 1$ coefficients for the FIR filter. Let $b_1, \ldots, b_T$ be the coefficients of the FIR filter that we want to design. The optimization problem of CSSSP is

$$\begin{aligned} \max_{b_2,\ldots,b_T} \max_{\boldsymbol{w}} \quad & \boldsymbol{w}^T \left(\sum_{\tau=0}^{T-1} \left(\sum_{j=1}^{T-\tau} b_j b_{j+\tau} \right) \boldsymbol{\Sigma}_c^\tau \right) \boldsymbol{w} + \frac{C}{T} \|\boldsymbol{b}\|_1, \\ \text{subject to} \quad & \boldsymbol{w}^T \left[\sum_{\tau=0}^{T-1} \left\{ \sum_{j=1}^{T-\tau} b_j b_{j+\tau} \right\} (\boldsymbol{\Sigma}_1^\tau + \boldsymbol{\Sigma}_2^\tau) \right] \boldsymbol{w} = 1, \end{aligned} \tag{3.11}$$

where b_1 is fixed to 1, $\boldsymbol{b}$ is a coefficient vector defined as $\boldsymbol{b} = [b_1, \ldots, b_T]^T$, $\boldsymbol{\Sigma}_c^\tau$ is a correlation matrix between the signal, $\boldsymbol{X}$, and the τ delayed signal, $\boldsymbol{X}_\tau$, defined as

$$\boldsymbol{\Sigma}_c^\tau = E_{\boldsymbol{X} \in C_c} \left[\boldsymbol{X} \boldsymbol{X}_\tau^T + \boldsymbol{X}_\tau \boldsymbol{X}^T \right], \quad \tau > 0, \tag{3.12}$$

and

$$\boldsymbol{\Sigma}_c^0 = E_{\boldsymbol{X}\in C_c}\left[\boldsymbol{X}\boldsymbol{X}^T\right], \tag{3.13}$$

and C is a non-negative regularization coefficient that affects sparsity of $\boldsymbol{b}$. Optimization techniques like gradient or line-search methods can be used for solving (3.11) if T is not too large [45].

3.6 Spectrally Weighted CSP

SPEC-CSP uses an iterative procedure to achieve optimization of both of spatially weights and filters [46]. A filtered signal of an observed signal, $\boldsymbol{X}$, can be written by

$$\hat{\boldsymbol{X}} = \boldsymbol{X}\boldsymbol{F}_N\boldsymbol{B}\boldsymbol{F}_N^{-1} \tag{3.14}$$

with the discrete Fourier transform matrix, $\boldsymbol{F}_N \in \mathbb{C}^{N\times N}$, defined as

$$[\boldsymbol{F}_N]_{k,l} = \frac{1}{\sqrt{N}}e^{-2\pi(k-1)(l-1)/N}, \quad k,l = 1,\ldots,N, \tag{3.15}$$

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

$$v(\boldsymbol{X},\boldsymbol{w},\boldsymbol{\beta}) = \boldsymbol{w}^T\boldsymbol{X}\boldsymbol{F}_N\boldsymbol{B}\boldsymbol{B}^T\boldsymbol{F}_N\boldsymbol{X}^T\boldsymbol{w} = \boldsymbol{w}^T\left(\sum_{k=1}^{N} b_k^2\boldsymbol{V}_k\right)\boldsymbol{w} \tag{3.16}$$

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

$$\boldsymbol{\Sigma}_d^i = E_{\boldsymbol{X}\in C_d}\left[\sum_{k=1}^{N}\beta_{k,i}\boldsymbol{V}_k\right], \quad d = 1,2, \tag{3.17}$$

where the elements of $\boldsymbol{\beta}_i$ are denoted by $\boldsymbol{\beta}_i = [\beta_{1,i}, \ldots, \beta_{N,i}]^T$, and C_d is the set of the learning data belonging to class d. Then we solve the generalized eigenvalue problem;

$$\boldsymbol{\Sigma}_c^i \boldsymbol{w} = \lambda \boldsymbol{\Sigma}_{\bar{c}}^i \boldsymbol{w}, \tag{3.18}$$

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 as

$$a = \underset{i=1,\ldots,r}{\operatorname{argmin}} \ \lambda_1^i \tag{3.19}$$

and

$$b = \underset{i=1,\ldots,r}{\operatorname{argmax}} \ \lambda_M^i, \tag{3.20}$$

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

$$\begin{aligned} \max_{\boldsymbol{\beta}_i} \quad & \frac{E_{\boldsymbol{X} \in C_c}[s_i] - E_{\boldsymbol{X} \in C_{\bar{c}}}[s_i]}{\sqrt{\sum_{d=1,2} E_{\boldsymbol{X} \in C_d}[|s_i - E_{\boldsymbol{X} \in C_d}[s_i]|^2]}}, \\ \text{subject to} \quad & \beta_{k,i} \geq 0, \ \forall k = 1, \ldots, N, \end{aligned} \tag{3.21}$$

where s_i is defined as

$$s_i = \sum_{k=1}^{N} \beta_{k,i} \hat{\boldsymbol{w}}_i^T \boldsymbol{V}_k \hat{\boldsymbol{w}}_i. \tag{3.22}$$

Here, we define

$$\gamma_k^i = E_{\boldsymbol{X} \in C_c}[\hat{\boldsymbol{w}}_i^T \boldsymbol{V}_k \hat{\boldsymbol{w}}_i] - E_{\boldsymbol{X} \in C_{\bar{c}}}[\hat{\boldsymbol{w}}_i^T \boldsymbol{V}_k \hat{\boldsymbol{w}}_i] \tag{3.23}$$

and

$$\zeta_k^i = \sum_{d=1,2} E_{\boldsymbol{X} \in C_d}[|\hat{\boldsymbol{w}}_i^T \boldsymbol{V}_k \hat{\boldsymbol{w}}_i - E_{\boldsymbol{X} \in C_d}[\hat{\boldsymbol{w}}_i^T \boldsymbol{V}_k \hat{\boldsymbol{w}}_i]|^2]. \tag{3.24}$$

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 (3.21) is

$$\beta_{k,i} \propto \begin{cases} \gamma_k^i / \zeta_k^i & \gamma_k^i \geq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{3.25}$$

Because the norm of $\boldsymbol{\beta}_i$ cannot be determined by (3.25), 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, (3.18) and (3.21). 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

$$\boldsymbol{y} = [v(\boldsymbol{X}, \hat{\boldsymbol{w}}_1, \boldsymbol{\beta}_1), \ldots, v(\boldsymbol{X}, \hat{\boldsymbol{w}}_{2r}, \boldsymbol{\beta}_{2r})]^T. \tag{3.26}$$

3.7 Filter Bank Common Spatial Patterns

FBCSP selects some bandpass filters from a filter bank [50]. It comprises four stages: multiple bandpass filtering, spatial filtering by CSP, feature selection of the CSP features, and classification.

FBCSP first defines a bank of filters denoted by $\{\mathcal{H}_1, \ldots, \mathcal{H}_{N_F}\}$. Let $\hat{\boldsymbol{X}}_j$ be the filtered signal by $\hat{\boldsymbol{X}}_j = \mathcal{H}_j(\boldsymbol{X})$. CSP is applied for each $\hat{\boldsymbol{X}}_j$ and spatial weights, $\boldsymbol{w}_j^i$, $i = 1, \ldots, N_F$, $j = 1, \ldots, M$, are found by solving eigenvalue problems of (3.8) as done for CSP, where $\boldsymbol{w}_j^i$ is the eigenvector corresponding to the jth largest eigenvalue in the eigenvalue problem in the filtered signals by $\mathcal{H}_i$. Features are selected out of a set of $\{\boldsymbol{w}_j^i | i = 1, \ldots, N_F, j = 1, \ldots, N_M, M - N_M + 1, \ldots, M\}$ by a method based on mutual information between extracted features and its class labels, where N_M is the number of the spatial filter for a filter in the set. In [50], mu-

tual information based best individual feature (MIBIF) and mutual information-based naïve Bayesian Parzen window (MINBPW) are introduced for selecting features. After selecting features, the selected r sets of indexes of the spatial weight and the filter are obtained as $\{\kappa_1, \upsilon_1\}, \ldots, \{\kappa_r, \upsilon_r\}$, where $\kappa_i \in \{1, \ldots, N_M, M - N_M + 1, \ldots, M\}$ and $\upsilon_i \in \{1, \ldots, N_F\}$. We denote the feature vector extracted from $\boldsymbol{X}$ by

$$\boldsymbol{y} = [\sigma^2(\hat{\boldsymbol{X}}_{\upsilon_1}, \boldsymbol{w}_{\kappa_1}^{\upsilon_1}) \ldots, \sigma^2(\hat{\boldsymbol{X}}_{\upsilon_r}, \boldsymbol{w}_{\kappa_r}^{\upsilon_r})]^T. \tag{3.27}$$

Chapter 4

Common Spatio-Time-Frequency Patterns

The chapter proposes the CSTFP method. This method enables us to simultaneously design the parameters for the sets of the spatial weights, the frequency filters, and the time windows. We first show a signal extraction model with the sets of the filters in multichannel EEG observations in Sec. 4.1. In Sec. 4.2, an optimization problem is defined for design of the parameters that make the extracted signal have discriminative features. For solving the optimization problem, we propose an optimization method by using sequential optimization procedures and alternating optimization procedures. Section 4.4 proves the convergence in the iterative updating for solving the problem. Section 4.5 gives some examples for the constraints in the optimization problems. We evaluate performances of the CSTFP method and compare it with existing method in experiments of artificial signals and BMI datasets in Secs 4.6 and 4.7, respectively.

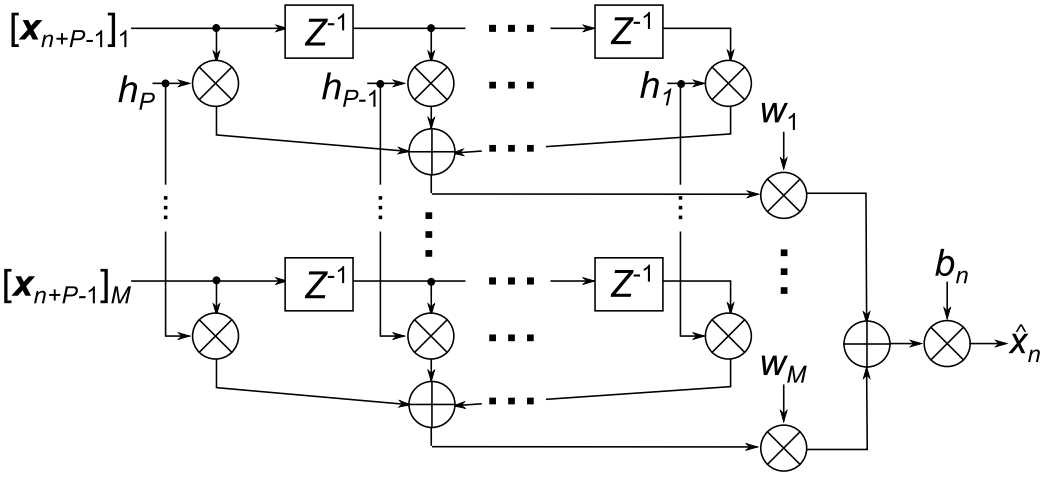


Figure 4.1: Signal extraction with a temporal filter $\boldsymbol{h}$, spatial weights, $\boldsymbol{w}$, and a time window, $\boldsymbol{b}$. Z^{-n} is an operator of n samples delay.

4.1 Signal Extraction Model

The target signal and signal extraction procedure is formulated in this section. The filtered signal of a target signal, $\boldsymbol{X}$, denoted as $\hat{\boldsymbol{x}} = [\hat{x}_1, \ldots, \hat{x}_K]^T$, is defined as

$$\hat{x}_n = b_n \sum_{m=1}^{M} w_m \sum_{p=1}^{P} h_p [\boldsymbol{x}_{n+P-p}]_m, \tag{4.1}$$

for $n = 1, \ldots, K$, $K = N - P + 1$, where $[\cdot]_i$ is the ith entry of a vector, P is the filter order of an FIR filter of which the coefficients are denoted by $h_1, \ldots, h_P$, w_m is a spatial weight for mth channel, b_n is a time window for nth sample that takes a binary value of either 0 or 1. The structure of temporally filtering, spatial weighting, and windowing for $\boldsymbol{X}$ is illustrated in Fig. 4.1. In this model, w_m regards as the spatial pattern, h_p regards as the frequency pattern, and b_n regards as the time pattern of the extracted signal.

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

$$\alpha_{\boldsymbol{X}}(\boldsymbol{w}, \boldsymbol{h}, \boldsymbol{b}) = \frac{1}{\|\boldsymbol{b}\|} \sum_{n=1}^{K} \left| \hat{x}_n - \frac{1}{K} \sum_{i=1}^{K} \hat{x}_i \right|^2, \tag{4.2}$$

where $\boldsymbol{w}$ is defined as $\boldsymbol{w} = [w_1, \ldots, w_M]^T$, $\boldsymbol{h}$ is defined as $\boldsymbol{h} = [h_1, \ldots, h_P]^T$, and $\boldsymbol{b}$ is defined as $\boldsymbol{b} = [b_1, \ldots, b_N]^T$.

The variance defined in (4.2) can be transformed to matrix-vector form as follows. We define $\boldsymbol{A}_n$, $n = 1, \ldots, K$, whose elements are from $\boldsymbol{X}$ as

$$[\boldsymbol{A}_n]_{m,p} = [\boldsymbol{X}]_{m,n+P-p}, \tag{4.3}$$

for $m = 1, \ldots, M$, $p = 1, \ldots, P$, where $[\cdot]_{i,j}$ is the element at the ith row and the jth column of a matrix. Then, (4.2) can be modified to

$$\alpha_{\boldsymbol{X}}(\boldsymbol{w}, \boldsymbol{h}, \boldsymbol{b}) = \frac{1}{\|\boldsymbol{b}\|} \sum_{n=1}^{K} b_n \left| \boldsymbol{w}^T \hat{\boldsymbol{A}}_n \boldsymbol{h} \right|^2, \tag{4.4}$$

where $\hat{\boldsymbol{A}}_n$ is defined as

$$\hat{\boldsymbol{A}}_n = \boldsymbol{A}_n - \frac{1}{\|\boldsymbol{b}\|} \sum_{m=1}^{K} b_m \boldsymbol{A}_m. \tag{4.5}$$

4.2 Optimization for Sets of Parameters

We consider the problem for design of F sets of the FIR filter, the spatial weights, and the time window represented by $\{\boldsymbol{w}_i, \boldsymbol{h}_i, \boldsymbol{b}_i\}_{i=1}^{F}$. These sets of the parameters are designed in such a way that $\boldsymbol{w}_i$, $\boldsymbol{h}_i$, and $\boldsymbol{b}_i$ maximize expectation of $\alpha_{\boldsymbol{X}}(\boldsymbol{w}_i, \boldsymbol{h}_i, \boldsymbol{b}_i)$ with respect to $\boldsymbol{X} \in C_c$ samples under the normalization of an expectation of $\alpha_{\boldsymbol{X}}(\boldsymbol{w}_i, \boldsymbol{h}_i, \boldsymbol{b}_i)$ over all of the observation. Additionally, we impose the orthonormality on $\boldsymbol{h}_i$, $i = 1, \ldots, F$ to avoid the trivial solution. Moreover, the

time windows are chosen from given candidates for efficient optimization. Therefore, we formulate the following maximization problem;

$$\begin{aligned} \max_{\mathcal{P}_i, i=1,\ldots,F} \quad & \sum_{i=1}^{F} \hat{J}(\mathcal{P}_i) + \frac{\epsilon}{K}\|\boldsymbol{b}_i\|, \\ \text{subject to} \quad & \frac{\boldsymbol{h}_i^T \boldsymbol{h}_j}{\|\boldsymbol{h}_i\| \, \|\boldsymbol{h}_j\|} = \delta_{ij}, \quad i, j = 1, \ldots, F, \\ & \boldsymbol{h}_k \in \mathcal{S}_k^{\perp}, \quad k = 1, \ldots, F, \\ & \boldsymbol{b}_l \in \mathcal{B}, \quad l = 1, \ldots, F, \end{aligned} \tag{4.6}$$

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

$$\hat{J}(\mathcal{P}_i) = \frac{E_{\boldsymbol{X} \in C_c}[\alpha_{\boldsymbol{X}}(\boldsymbol{w}_i, \boldsymbol{h}_i, \boldsymbol{b}_i)]}{\sum_{d=1,2} E_{\boldsymbol{X} \in C_d}[\alpha_{\boldsymbol{X}}(\boldsymbol{w}_i, \boldsymbol{h}_i, \boldsymbol{b}_i)]}, \tag{4.7}$$

$\mathcal{S}_i$ is any subspace in $\mathbb{R}^P$, $\mathcal{B}$ is a candidate set for the time windows, defined as $\mathcal{B} = \{\boldsymbol{b}_l\}_{l=1}^{L}$, 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 $\boldsymbol{h}_1$. This sequential optimization is represented with respect to each i as

$$\begin{aligned} \max_{\mathcal{P}_i} \quad & \hat{J}(\mathcal{P}_i) + \frac{\epsilon}{K}\|\boldsymbol{b}_i\|, \\ \text{subject to} \quad & \boldsymbol{h}_i \in \hat{\mathcal{S}}_i^{\perp}, \quad \boldsymbol{b} \in \mathcal{B}, \end{aligned} \tag{4.8}$$

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

$$\hat{\mathcal{S}}_i = \text{Span}\{\boldsymbol{h}_1, \ldots, \boldsymbol{h}_{i-1}\} \oplus \mathcal{S}_i \in \mathbb{R}^P, \tag{4.9}$$

where $\text{Span}(\cdots)$ represents a subspace spanned by vectors and the operator denoted by $\oplus$ gives the direct sum of two subspaces. Methods for choosing $\hat{\mathcal{S}}_i$ will

be discussed in Sec. 4.5. In (4.8), to optimize the parameters indexed with i, we adopt alternating optimization procedure based on alternating least square (ALS). In the optimization, we separate the problem of (4.8) into three subproblems for $\boldsymbol{w}_i$, $\boldsymbol{h}_i$, and $\boldsymbol{b}_i$, 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 $\boldsymbol{w}_i$. While fixing $\boldsymbol{h}_i$ and $\boldsymbol{b}_i$, $\boldsymbol{w}_i$ maximizing (4.8) is found as the generalized eigenvector corresponding to the largest generalized eigenvalue of the generalized eigenvalue problem [48] described as

$$\boldsymbol{R}_c \boldsymbol{w}_i = \lambda(\boldsymbol{R}_1 + \boldsymbol{R}_2)\boldsymbol{w}_i, \tag{4.10}$$

where

$$\boldsymbol{R}_d = E_{\boldsymbol{X} \in C_d}\left[\frac{1}{\|\boldsymbol{b}_i\|}\sum_{n=1}^{K}[\boldsymbol{b}]_n \hat{\boldsymbol{A}}_n \boldsymbol{h}_i \boldsymbol{h}_i^T \hat{\boldsymbol{A}}_n^T\right], \tag{4.11}$$

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

The second subproblem is to optimize $\boldsymbol{h}_i$. The solution to the problem optimizing $\boldsymbol{h}$ while fixing $\boldsymbol{w}_i$ and $\boldsymbol{b}$ with the orthogonal constraint is given by the following theorem.

Theorem 1. *When matrices $\boldsymbol{Q}_c$ and $\boldsymbol{Q}_1 + \boldsymbol{Q}_2$ are nonsingular, the solution is given by the unit-length generalized eigenvector corresponding to the largest generalized eigenvalue of the generalized eigenvalue problem described as*

$$\boldsymbol{G}\boldsymbol{Q}_c \boldsymbol{h}_i = \zeta(\boldsymbol{Q}_1 + \boldsymbol{Q}_2)\boldsymbol{h}_i, \tag{4.12}$$

where

$$\boldsymbol{Q}_d = E_{\boldsymbol{X} \in C_d}\left[\frac{1}{\|\boldsymbol{b}_i\|}\sum_{n=1}^{K}[\boldsymbol{b}_i]_n \hat{\boldsymbol{A}}_n^T \boldsymbol{w}_i \boldsymbol{w}_i^T \hat{\boldsymbol{A}}_n\right], \quad d = 1, 2, \tag{4.13}$$

$$\boldsymbol{G} = \boldsymbol{I}_P - \boldsymbol{V}(\boldsymbol{V}^T(\boldsymbol{Q}_1 + \boldsymbol{Q}_2)^{-1}\boldsymbol{V})^{-1}\boldsymbol{V}^T(\boldsymbol{Q}_1 + \boldsymbol{Q}_2)^{-1}, \tag{4.14}$$

$\boldsymbol{V}$ is the matrix defined as

$$\boldsymbol{V} = [\boldsymbol{v}_{i1}, \ldots, \boldsymbol{v}_{iD_i}] \in \mathbb{R}^{P \times D_i}, \tag{4.15}$$

$\boldsymbol{v}_{i1}, \ldots, \boldsymbol{v}_{iD_i}$ are vectors spanning $\hat{S}_i$ described as

$$\text{Span}\{\boldsymbol{v}_{i1}, \ldots, \boldsymbol{v}_{iD_i}\} = \hat{\mathcal{S}}_i, \tag{4.16}$$

and $\boldsymbol{I}_P$ is the $P \times P$ identify matrix, and ζ is an eigenvalue.

Proof. See Appendix A.1. □

The third subproblem is to optimize $\boldsymbol{b}_i$ while fixing $\boldsymbol{w}_i$ and $\boldsymbol{h}_i$. The cost of (4.8) can be reduced to

$$J_3(\boldsymbol{b}_i|\boldsymbol{w}_i, \boldsymbol{h}_i) = \frac{g_1(\boldsymbol{b}_i)}{g_1(\boldsymbol{b}_i) + g_2(\boldsymbol{b}_i)} + \frac{\epsilon}{K}\|\boldsymbol{b}_i\|. \tag{4.17}$$

Then

$$\begin{aligned} g_d(\boldsymbol{b}) &= E_{\boldsymbol{X} \in C_d}\left[\frac{1}{\|\boldsymbol{b}\|}\sum_{n=1}^{K}\left(b_n\tilde{x}_n - \frac{1}{\|\boldsymbol{b}\|}\sum_{k=1}^{K} b_k\tilde{x}_k\right)^2\right] \\ &= \frac{1}{\|\boldsymbol{b}\|}E_{\boldsymbol{X} \in C_d}\left[\boldsymbol{b}^T(\boldsymbol{\xi} + 2\mu\tilde{\boldsymbol{x}}) + \mu\right], \end{aligned} \tag{4.18}$$

where $\tilde{x}_n = \boldsymbol{w}_i^T \boldsymbol{A}_n \boldsymbol{h}_i$, $\tilde{\boldsymbol{x}} = [\tilde{x}_1, \ldots, \tilde{x}_K]^T$, $\boldsymbol{\xi} = [\tilde{x}_1^2, \ldots, \tilde{x}_K^2]^T$, and $\mu = \|b\|^{-1}\boldsymbol{b}^T\tilde{\boldsymbol{x}}$. Because $\boldsymbol{b}_i$ are chosen out of $\mathcal{B}$, we calculate the values $J_3(\hat{\boldsymbol{b}}_i)$ for all candidates in $\mathcal{B}$ and the optimal $\boldsymbol{b}_i$ can be chosen as the candidates that maximizing J_3. This formulates that

$$\boldsymbol{b}_i = \underset{\boldsymbol{b} \in \mathcal{B}}{\text{argmax}}\ J_3(\boldsymbol{b}). \tag{4.19}$$

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

Algorithm 1 Design of the FIR filters, the spatial weights, and the time windows

Input: C_1 and C_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: $\{\boldsymbol{w}_i\}_{i=1}^F$: the spatial weights, $\{\boldsymbol{h}_i\}_{i=1}^F$: the vectors of FIR filter coefficients, $\{\boldsymbol{b}_i\}_{i=1}^F$: the time window.

for $i = 1, \ldots, F$ **do**

Initialize $\boldsymbol{h}_i$ and $\boldsymbol{b}_i$.

Set the index of iteration as $k = 0$.

repeat

$k \leftarrow k + 1$.

Update $\boldsymbol{w}_i$ by solving (4.10).

Update $\boldsymbol{h}_i$ by solving (4.12).

Update $\boldsymbol{b}_i$ by solving (4.19).

Calculate cost, C_k from the cost function, $\hat{J}(\mathcal{P}_i)$.

until $C_k - C_{k-1}$ is sufficiently small.

end for

4.3 Feature Vector Definition

With designed $\boldsymbol{w}_i$, $\boldsymbol{h}_i$, and $\boldsymbol{b}_i$, the feature vector of an EEG signal, $\boldsymbol{X}$, for classification is defined as

$$\begin{aligned}\boldsymbol{y} =&[\alpha_{\boldsymbol{X}}(\hat{\boldsymbol{w}}_1^{(1)}, \boldsymbol{h}_1, \boldsymbol{b}_1), \ldots, \alpha_{\boldsymbol{X}}(\hat{\boldsymbol{w}}_1^{(2r)}, \boldsymbol{h}_1, \boldsymbol{b}_1), \ldots, \\ &\alpha_{\boldsymbol{X}}(\hat{\boldsymbol{w}}_F^{(1)}, \boldsymbol{h}_F, \boldsymbol{b}_F), \ldots, \alpha_{\boldsymbol{X}}(\hat{\boldsymbol{w}}_F^{(2r)}, \boldsymbol{h}_F, \boldsymbol{b}_F)]^T,\end{aligned} \tag{4.20}$$

where the set of $\{\hat{\boldsymbol{w}}_i^{(1)}, \ldots, \hat{\boldsymbol{w}}_i^{(2r)}\}$ is the CSPs corresponding to $\boldsymbol{h}_i$ and $\boldsymbol{b}_i$. $\hat{\boldsymbol{w}}_i^{(m)}$, $i = 1, \ldots, F$, $m = 1, \ldots, 2r$, are decided as follows. By solving (4.10) with $\boldsymbol{h}_i$ and $\boldsymbol{b}_i$, we obtain MF spatial patterns as $\tilde{\boldsymbol{w}}_i^{(m)}$ for $i = 1, \ldots, F$ and $m = 1, \ldots, M$, where $\tilde{\boldsymbol{w}}_i^{(m)}$ is the unit-length eigenvector corresponding to the m-th largest eigenvalue of (4.10). Then $\hat{\boldsymbol{w}}_i^{(m)}$ are defined as $\hat{\boldsymbol{w}}_i^{(j)} = \tilde{\boldsymbol{w}}_i^{(j)}$ and $\hat{\boldsymbol{w}}_i^{(r+j)} = \tilde{\boldsymbol{w}}_i^{(M-j+1)}$ for $i = 1, \ldots, F$, $j = 1, \ldots, r$.

4.4 Convergence of Cost Function in Optimization

The alternating optimization of DFBCSP does not increase the cost functions and leads to a local maximum solution of (4.8). Here we give the guarantee of convergence of the cost in the alternating iteration.

Proposition 1. *The cost function of* (4.8) *monotonically increases or remains the same by iterations. That is,*

$$0 \leq \hat{J}(\mathcal{P}_i^{(k)}) \leq \hat{J}(\mathcal{P}_i^{(k+1)}) \leq 1, \tag{4.21}$$

where $\mathcal{P}_i^{(k)}$ are the set of the parameters to be optimized at kth iterations.

Proof. See Appendix A.2. □

4.5 Search Space for FIR Filter Coefficients

The algorithm of the sequential update defined in (4.8) seeks $\boldsymbol{h}_i$ in the orthogonal complement of $\hat{\mathcal{S}}_i$. In this section, we introduce two ways to determine $\hat{\mathcal{S}}_i$.

4.5.1 Subspace Spanned by Filter Coefficients Vectors

This way uses as $\hat{\mathcal{S}}_i$ a subspace:

$$\hat{\mathcal{S}}_i = \mathrm{Span}\{\boldsymbol{h}_1, \ldots, \boldsymbol{h}_{i-1}\}. \tag{4.22}$$

Because $\boldsymbol{h}_1, \ldots, \boldsymbol{h}_{i-1}$ are orthogonal from each other, they are a set of an orthogonal basis of $\hat{\mathcal{S}}_i$.

4.5.2 Subspace Spanned by Filter Coefficients Vectors and Its Shifted Vectors

The subspace given as in (4.22) can yield similar bandpass filters in terms of the amplitude characteristics in the frequency domain. To avoid this redundancy, time-shifted filters of $\boldsymbol{h}_1, \ldots, \boldsymbol{h}_{i-1}$ can be used for forming the subspace;

$$\hat{\mathcal{S}}_i = \mathrm{Span}\{\boldsymbol{h}_1, \ldots, \boldsymbol{h}_{i-1}, \boldsymbol{h}'_1, \ldots, \boldsymbol{h}'_{i-1}\}, \tag{4.23}$$

where $\boldsymbol{h}'_j$, $j = 1, \ldots, i-1$ are defined as

$$[\boldsymbol{h}'_j]_k = [\boldsymbol{h}_j]_{((k+1) \bmod P)}, \quad k = 1, \ldots, P. \tag{4.24}$$

This formulation is justified as follows. Taking into account that a filter vector can be decomposed into the superposition of sinusoids, we consider a simple filter given by a sinusoidal impulse response as

$$[\boldsymbol{h}_a]_n = \sin \omega_p (n-1), \tag{4.25}$$

for $n = 1, \ldots, P$. Suppose $\hat{\mathcal{S}}_b$ as the search space for $\boldsymbol{h}_b$ in the optimization problem of (4.8) and $\hat{\mathcal{S}}_b \ni \boldsymbol{h}_a$. The solution, $\boldsymbol{h}_b$, can be a time-shifted filter of $\boldsymbol{h}_a$ such as

$$[\boldsymbol{h}_b]_n = \cos \omega_p (n-1), \tag{4.26}$$

for $n = 1, \ldots, P$, because of the orthogonality:

$$\langle \boldsymbol{h}_a, \boldsymbol{h}_b \rangle = \sum_{n=0}^{N-1} \sin \omega_p n \cos \omega_p n = 0. \tag{4.27}$$

This implies that given a filter, the orthogonal complement to $\boldsymbol{h}_a$ can include a solution of filters that have the amplitude characteristics similar to the given filter.

The time-shifted solution such $\boldsymbol{h}_b$ can be avoided by including $\cos \omega_p n$ or any time-shifted filter of $\boldsymbol{h}_a$ in $\mathcal{S}_a$, since

$$\cos \omega_p n = c_1 \sin \omega_p n + c_2 \sin \omega_p (n+m), \quad \forall n, \tag{4.28}$$

or

$$\boldsymbol{h}_b = c_1 \boldsymbol{h}_a + c_2 \boldsymbol{h}'_a, \tag{4.29}$$

where c_1 and c_2 are appropriate coefficients and $m \neq l\pi/\omega_p$, $m, l \in \mathbb{Z}$. Therefore, after obtaining $\boldsymbol{h}_a$, its shifted vector such as $\boldsymbol{h}_b$ is not in $\hat{\mathcal{S}}_{a+1}^{\perp}$ defined in (4.23).

4.6 Simulation by 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 observed EEG signals are modeled by a mixture of narrow-band signals (see Fig. 4.2). In this model, a trial signal belonging to class d is given by

$$\boldsymbol{x}[n] = \sum_{i=1}^{N_s} s_i[n] \boldsymbol{a}_i^{(d)} + \boldsymbol{\eta}, \quad n = 0, \ldots, N-1, \tag{4.30}$$

where d is a class label taking either 1 or 2, $\boldsymbol{x}[n] \in \mathbb{R}^M$ 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 ith source signal of feature components, N_s is the number of the source signals, $\boldsymbol{a}_i^{(d)} \in \mathbb{R}^M$ is a vector defined as $\boldsymbol{a}_i^{(d)} = [a_{i1}^{(d)}, \ldots, a_{iM}^{(d)}]^T$, $a_{im}^{(d)} \in \mathbb{R}$ is the amplitude of $s_i[n]$ in the mth channel for class d, and $\boldsymbol{\eta} \in \mathbb{R}^M$ is a stochastic noise. The source signals, $s_i[n]$, are generated as

$$s_i[n] = t_i[n]\Re\left[\sum_{k=0}^{N-1} S_i[k]e^{j\theta}e^{j\frac{2\pi k}{N}n}\right], \tag{4.31}$$

where $S_i[k] \in \mathbb{R}$ represents a discrete spectrum, $t_i[n]$ represents the time window that decides the period when the ith 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.

Specifically, 200 artificial signals we used in the experiment were generated with the conditions shown in Table 4.1 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.

We applied the CSTFP method to the artificial signals as follows. The class label represented by c in (4.7) 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

$$\boldsymbol{d}_j = [\underbrace{0, \ldots, 0}_{D(j-1)}, \underbrace{1, \ldots, 1}_{D}, 0, \ldots, 0]^T, \quad j = 1, \ldots, 10, \tag{4.32}$$

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

$$\hat{\boldsymbol{b}}_l = p_1\boldsymbol{d}_1 + \ldots + p_{10}\boldsymbol{d}_{10}, \tag{4.33}$$

Table 4.1: The conditions for generating the artificial signals.

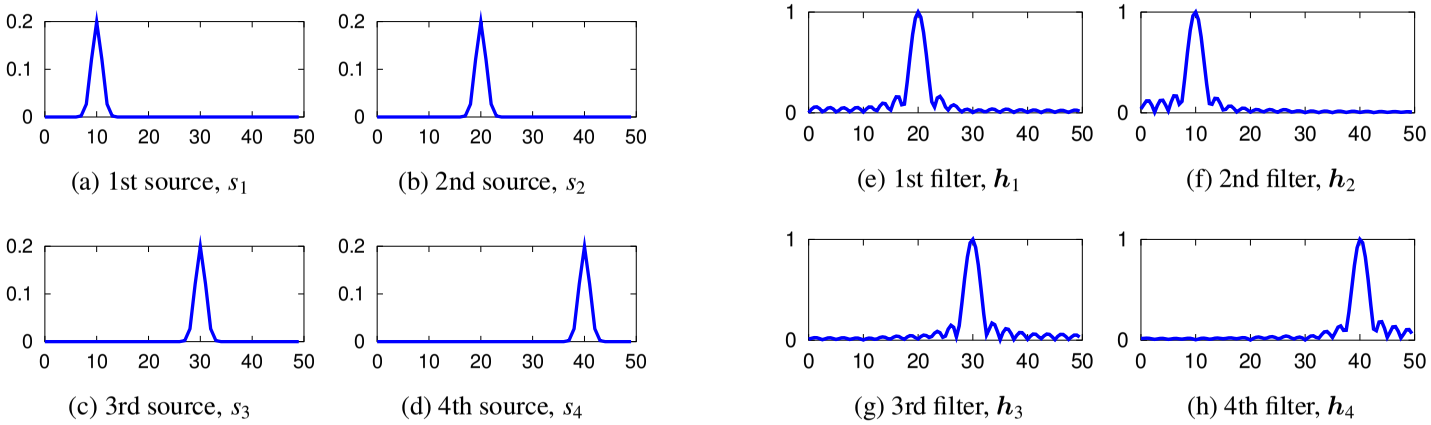
Parameter	Value or distributions
Number of channel, M	10
Number of samples, N	100
Number of trials for each class	100
Sampling frequency	100 Hz
Number of sources, N_s	4
Spectra of sources, $\lvert S_i[k]\rvert$	Figs. 4.2a–4.2d
Stochastic phase, θ	$\mathcal{U}(0, 2\pi)$
Time windows, $t_i[n]$	Figs. 4.3a–4.3d
Amplitudes, $\boldsymbol{a}_i^{(d)}$	Figs. 4.4a–4.4d
Stochastic noise, $[\boldsymbol{\eta}]_m$	$\mathcal{N}(0, 0.1)$

and $p_j \in \{0, 1\}$, $j = 1, \ldots, 10$, as the given candidate set. Therefore, the number of the candidates, L, was 1023. Moreover, for $\hat{\mathcal{S}}_i$ that determines the search space for $\boldsymbol{h}_i$, we used that

$$\hat{\mathcal{S}}_i = \text{Span}\{\boldsymbol{h}_1, \ldots, \boldsymbol{h}_{i-1}, \boldsymbol{h}'_{i-1}, \ldots, \boldsymbol{h}'_{i-1}\}. \tag{4.34}$$

The optimization resulted in Figs. 4.2e–4.2h for the amplitude characteristics of the FIR filters, Figs. 4.3e–4.3h for the time windows, and Figs. 4.4e–4.4h for the normalized spatial weights.

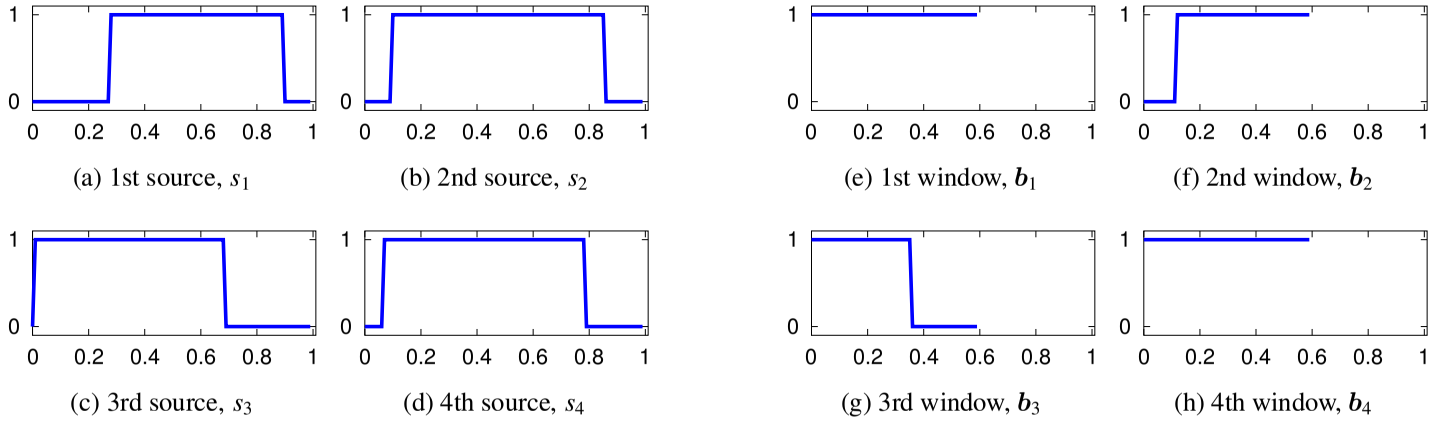
The centers of passbands of the filters shown in Figs. 4.2e–4.2h coincide with the centers of the source signals shown in Figs. 4.2a–4.2d. Moreover, in the spectrum of the source and the amplitude characteristic of the designed FIR filter that have the similar center frequencies, the spatial amplitude corresponding to the source and the spatial weight vector corresponding to the FIR filter are simi-



(a) 1st source, s_1 (b) 2nd source, s_2 (e) 1st filter, $\boldsymbol{h}_1$ (f) 2nd filter, $\boldsymbol{h}_2$

(c) 3rd source, s_3 (d) 4th source, s_4 (g) 3rd filter, $\boldsymbol{h}_3$ (h) 4th filter, $\boldsymbol{h}_4$

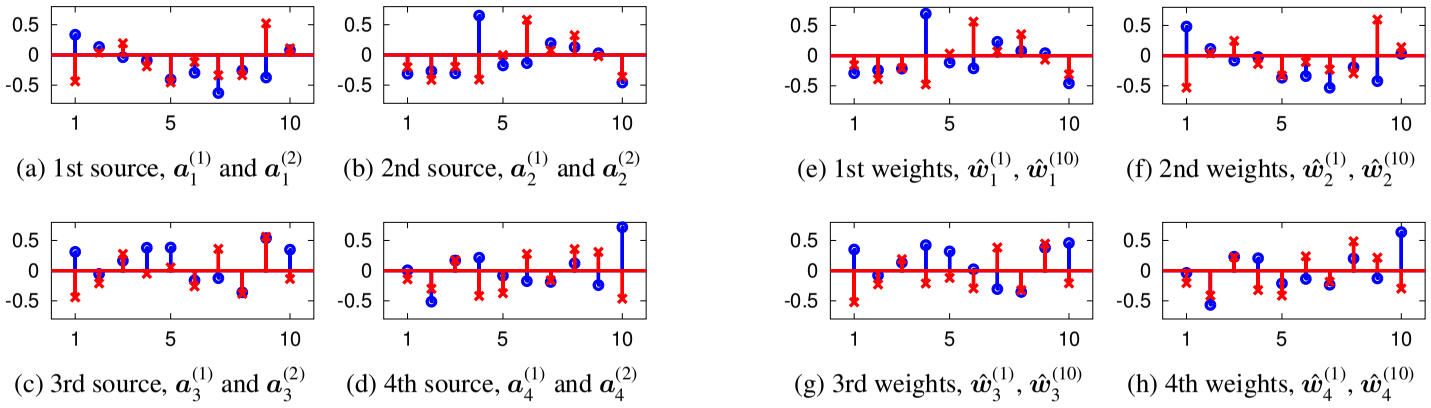
Figure 4.2: Frequency patterns as the spectra of the source signals (Figs. 4.2a–4.2d) and the amplitude characteristics of the designed frequency filters (Figs. 4.2e–4.2h). The horizontal axis represents frequency and its unit is Hz.



(a) 1st source, s_1 (b) 2nd source, s_2 (e) 1st window, $\boldsymbol{b}_1$ (f) 2nd window, $\boldsymbol{b}_2$

(c) 3rd source, s_3 (d) 4th source, s_4 (g) 3rd window, $\boldsymbol{b}_3$ (h) 4th window, $\boldsymbol{b}_4$

Figure 4.3: Time patterns as the time windows of the source signals (Figs. 4.3a–4.3d) and the designed time windows (Figs. 4.3e–4.3h). The horizontal axis represents time and its unit is second.



(a) 1st source, $\boldsymbol{a}_1^{(1)}$ and $\boldsymbol{a}_1^{(2)}$ (b) 2nd source, $\boldsymbol{a}_2^{(1)}$ and $\boldsymbol{a}_2^{(2)}$ (e) 1st weights, $\hat{\boldsymbol{w}}_1^{(1)}$, $\hat{\boldsymbol{w}}_1^{(10)}$ (f) 2nd weights, $\hat{\boldsymbol{w}}_2^{(1)}$, $\hat{\boldsymbol{w}}_2^{(10)}$

(c) 3rd source, $\boldsymbol{a}_3^{(1)}$ and $\boldsymbol{a}_3^{(2)}$ (d) 4th source, $\boldsymbol{a}_4^{(1)}$ and $\boldsymbol{a}_4^{(2)}$ (g) 3rd weights, $\hat{\boldsymbol{w}}_3^{(1)}$, $\hat{\boldsymbol{w}}_3^{(10)}$ (h) 4th weights, $\hat{\boldsymbol{w}}_4^{(1)}$, $\hat{\boldsymbol{w}}_4^{(10)}$

Figure 4.4: Spatial patterns as the spatial amplitudes of the source signals (Figs. 4.4a–4.4d) and the designed spatial weights (Figs. 4.4e–4.4h). The amplitudes plotted by circles are $\boldsymbol{a}_i^{(1)}$ and the amplitudes plotted by cross are $\boldsymbol{a}_i^{(2)}$ for $i = 1, \ldots, 4$. The weights plotted by circles are $\boldsymbol{w}_i^{(1)}$ and the weights plotted by cross are $\boldsymbol{w}_i^{(10)}$ for $i = 1, \ldots, 4$. The horizontal axis represents the channel number.

lar to each other. For instance, the correlation coefficient between $\boldsymbol{a}_1^{(1)}$ shown in Fig. 4.4a with circles and $\boldsymbol{w}_2^{(1)}$ shown in Fig. 4.4f 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 first 25 samples according to Fig. 4.3a, the time window for extracting $s_1[n]$ is expected to remove first 25 samples. Because $|S_1[k]|$ and the amplitude characteristics shown in Fig. 4.2f have the similar center frequency, we can decide that the time window for extracting $s_1[n]$ is the time window shown in Fig. 4.3f. Although the designed time windows does not coincide with $t_i[n]$ because $\boldsymbol{b}$ is applied to an FIR-filtered signal that is shorter than than 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 Fig. 4.3c 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 Fig. 4.3c), the time window shown in Fig. 4.3c removes samples in the last 20 samples.

4.7 Classification Experiment of BMI Datasets

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 [43], common sparse spectral spatial patterns (CSSSP) [45], filter bank CSP (FBCSP) [50], and discriminative filter bank CSP (DFBCSP) [48]).

Table 4.2: Description of the datasets.

	dataset IVa	dataset 1	original dataset
Classes	right hand, right foot	2 tasks from foot, left hand, and right hand	left hand, right hand
Subject labels	*aa*, *al*, *av*, *aw*, *ay*	*a*, *b*, *f*, *g*	*sa*, *sb*, *sc*
Number of channels	118	59	29
Signal length	3.5 secs	4 secs	4 secs
Sampling rate	100 Hz	100 Hz	128 Hz
Number of the trials per class	140	100	100

4.7.1 Data Description

We used dataset IVa from BCI competition III [29], 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) [40] 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) [113]. The condition for each dataset is shown in Table 4.2. 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 lowpass filter whose cutoff frequency is 50 Hz and filter order is 4, and downsampled to 100 Hz.

We additionally conducted the classification experiment with a dataset we recorded by ourselves. We call our dataset the original dataset in the thesis. The recording for the original dataset was approved by the research ethics committee of Tokyo University of Agriculture and Technology. The condition for the dataset is also shown in the third column of Table 4.2. During the recording, the subjects who participated in the dataset performed the MI tasks that were represented by the visual cue. The cue was as an arrow on an LCD screen. The right arrow represented the direction that the subject performs the motor imagery task of the right hand. The left arrow represented the direction that the subject performs the motor imagery task of the left hand. The subject performed the tasks repeatedly with an interval of around 3 seconds. The EEG signals were recorded with Ag/AgCl active electrodes (g.LADYbird, g.LADYbirdGND, and g.GAMMAearclip produced

by Guger Technologies) and a power supply (g.GAMMAbox produced by Guger Technologies). The electrodes were placed at F3, Fz, F4, FC5, FC3, FC6, FCz, FC2, FC4, FC6, T7, C5, C3, C1, Cz, C2, C4, C6, T8, CP5, CP3, CP1, CPz, CP2, CP4, CP6, P3, Pz, and P4 (the positions are represented by the notation of International 10-10 system [78]). The signals observed by the electrodes were amplified by an amplifier (MEG-6116 produced by Nihon Kohden). The amplifier moreover analog-filtered the signals with the passband of 0.5–100 Hz. The signals through the amplifier were sampled by an A/D converter (AIO-163202F-PE produced by Contec) with the sampling rate of 256 Hz. The converted signals were recorded with Data Acquisition Toolbox that is one of the toolboxs of MATLAB (MathWorks). We furthermore applied to this dataset a Butterworth lowpass filter whose cutoff frequency is 50 Hz and filter order is 4, and downsampled to 128 Hz.

4.7.2 Result

For the experiments, as a sample for each trial, we used a signal observed in the period of 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, CSTFP), the parameters for the feature extraction were obtained as follows.

- **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 (3.5). The eigenvectors corresponding to the r largest and r smallest eigenvalues of the eigenvalue problem (3.8) were given as the spatial weights.

- **CSP-Exh:** The parameters determined in this method are spatial weights and a passband of the Butterworth filter. The passband of the Butterworth filters were 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 the CSP method.

- **CSSSP:** The parameters determined in this method are spatial weights and a bandpass filter. The bandpass filter between 7–30 Hz was applied as preprocessing [45]. CSSSP was applied with a regularization parameters, C, and the parameter for the number of the spatial weights, r. The order of the filter was fixed to 16 [45].

- **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 [50]. 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 [50].

- **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 [48]. In optimization, we stopped iteration when error of the cost function between successive iterations becomes under 10^{-5}.

- **CSTFP:** The parameters determined in this method are F FIR filters, the corresponding F time windows, and r spatial weights associated to each FIR filter We fixed $T_1 = 0$ to observe the behavior of the resulting time window. CSTFP applied with the FIR filter order, 41 [48], and the following candidate set for the time windows. The candidate set for the time windows, $\mathcal{B}$, is consisted of the vectors defined as

$$\hat{\boldsymbol{b}}_l = [\underbrace{0, \ldots, 0}_{D}, \underbrace{1, \ldots, 1}_{O}, 0, \ldots, 0]^T, \tag{4.35}$$

  where we choose D and O out of a set $\{0, 5, 10, \ldots, K\}$ such that $O > 50$ and $D + O \leq K$. The regularization parameter, ϵ, was set to 0.1. In alternating optimization, we initialize $\boldsymbol{h}_i$ as a random vector which is orthonormalized from $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{D_i}$ and $\boldsymbol{b}_i$ as a vector all 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 [114]. We conducted the nested CV [114] 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 4.3.

Table 4.3: The parameters decided by the nested CV in the learning data in the classification experiments.

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

Table 4.4: Classification accuracy rates [%] given by 5×5 CV in dataset IVa from BCI competition III. The figure with ± represents the standard deviation (S.D.) over CV.

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

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 of the feature vectors. Then, the LDA for the learning samples was used for obtaining 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 4.4, 4.5, and 4.6 in which we show the classification accuracies of each subject and method, 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 4.3 were fixed, to show effects of changes of the parameters in the classification accuracy rates. Figures 4.5 and 4.6 show the accuracy rates of CSP and the rates of each method averaged over all subjects with each T_1.

Table 4.5: Classification accuracy rates [%] given by 5×5 CV in dataset 1 from BCI competition IV. The figure with ± represents the standard deviation (S.D.) over CV.

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

For each subject, the parameters shown in Table 4.3 except for T_1 were fixed to the combinations of the parameters that performed the highest accuracy rates with each T_1. In Fig. 4.5, T_1 that perform the highest accuracy rates are different among the subjects. Moreover, Fig. 4.6 shows that the accuracy rates highly depend on T_1 in the conventional methods. Figures 4.7 and 4.8 show 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 Figs. 4.9 and 4.10. As we can observe in Figs. 4.7, 4.8, 4.9a, and 4.10a, the short time windows caused by small ϵ result in poor classification accuracy rates. All of the time windows shown in Figs. 4.9 and 4.10 remove the samples ob-

Table 4.6: Classification accuracy rates [%] given by 5×5 CV in the original dataset. The figure with ± represents the standard deviation (S.D.) over CV.

Method	Subject			
	sa	*sb*	*sc*	Ave.
CSP	87.3±3.9	54.7±8.8	92.0±4.8	78.0
CSP-Ref	**92.8**±3.3	55.3±7.6	92.0±3.9	80.0
CSSSP	91.0±4.1	56.1±9.7	93.1±3.3	80.1
FBCSP	75.2±17.7	50.1±8.8	74.5±16.6	66.6
DFBCSP	90.8±4.3	52.5±6.7	93.9±4.0	79.1
CSTFP	91.7±4.5	**57.4**±7.4	**94.1**±4.0	**81.1**

served within a few hundreds milliseconds after the cue. The results suggest that the brain activities related to the task can not be observed just after the cue. In Figs. 4.9c–4.9e and 4.10c–4.10e, 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 Figs. 4.7 and 4.8 in which the classification accuracy rates do not strongly depend on the regularization parameter, ϵ, more than 0.12. Moreover, the FIR filters do not highly differ even with various ϵ. However, the spatial weights with the small windows (Figs. 4.9a and 4.10a) 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*.

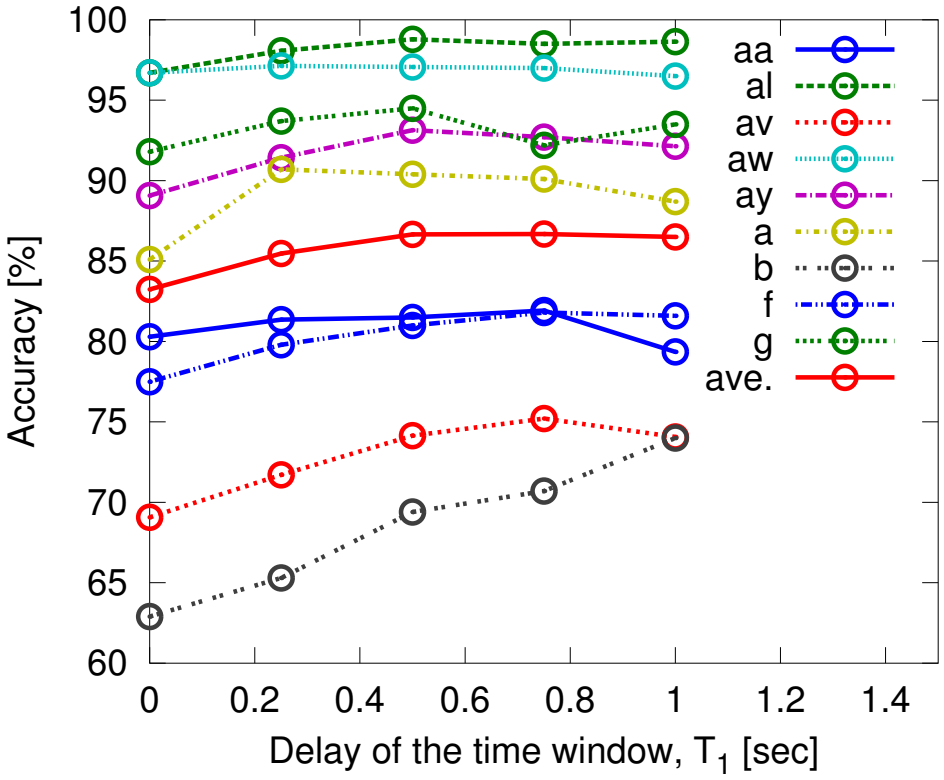


Figure 4.5: The variation of the accuracy rates by various T_1 in CSP.

4.8 Conclusion

We have proposed a novel method called CSTFP for 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. The method 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

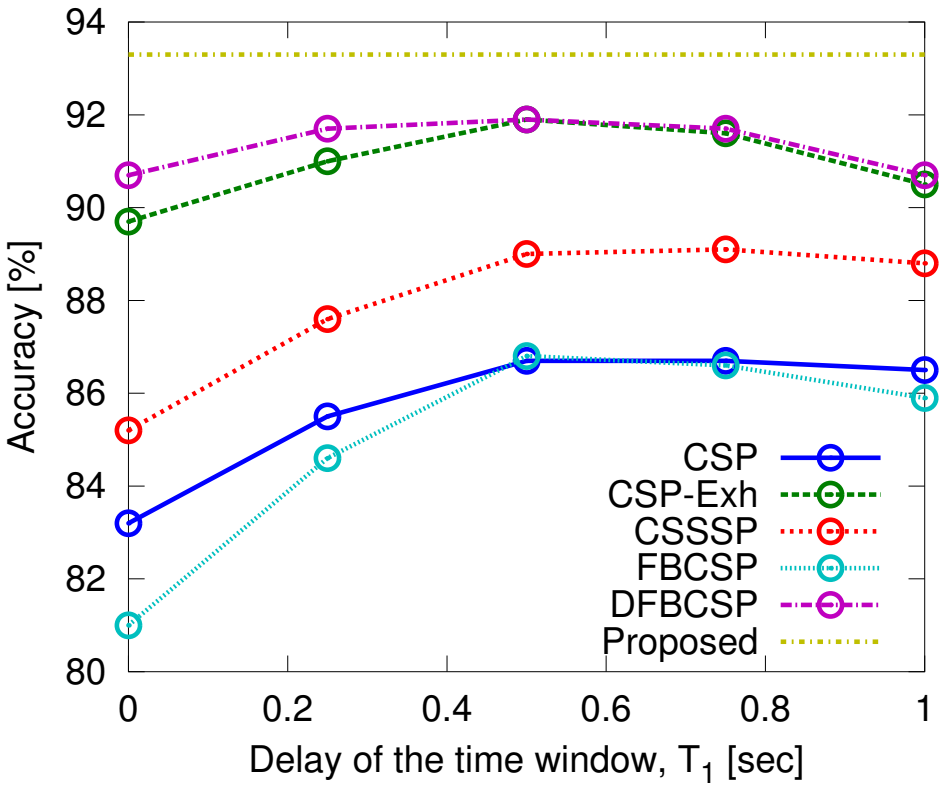


Figure 4.6: The variation of the accuracy rates by various T_1 in each method. The accuracy rates are averaged over all subjects.

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 to a mental task can be observed.

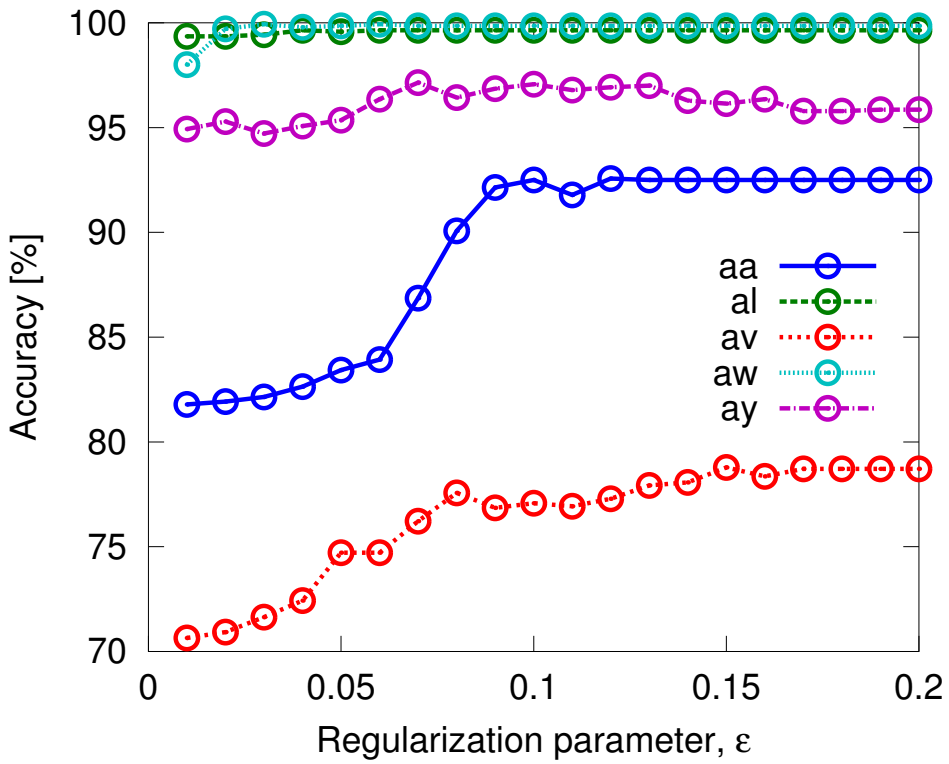


Figure 4.7: The variation of the accuracy rates by various regularization parameters, ϵ, in BCI competition III dataset IVa.

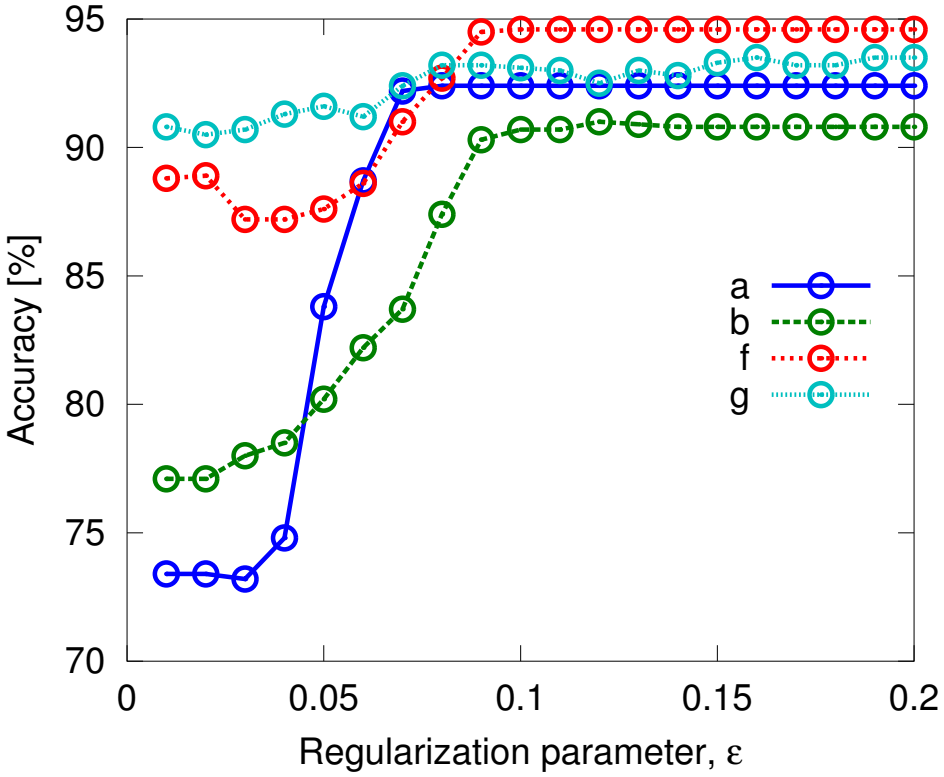
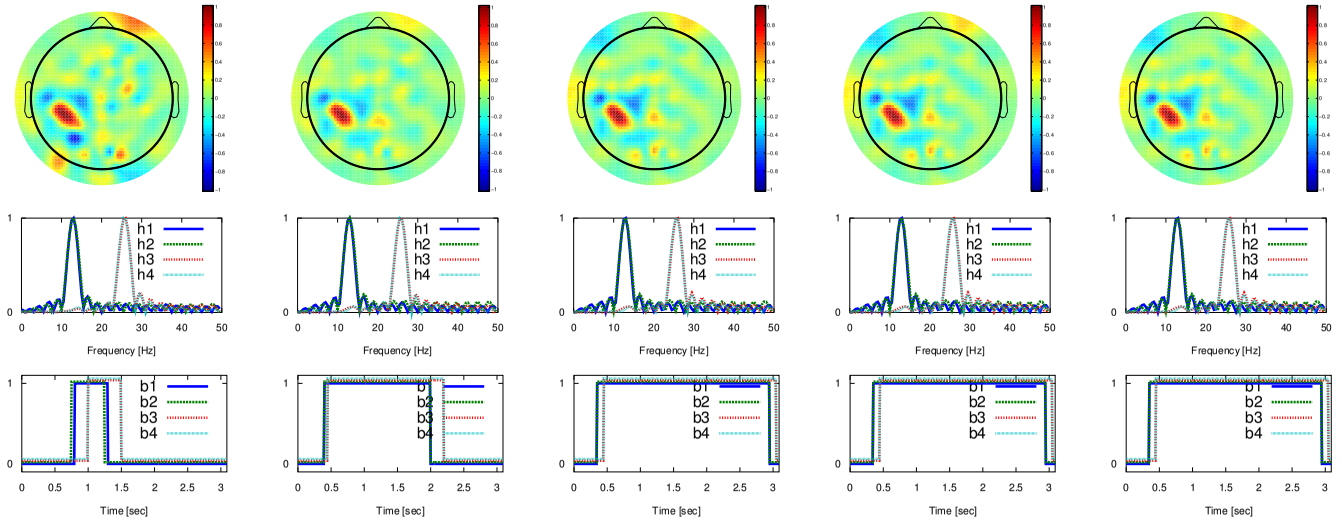
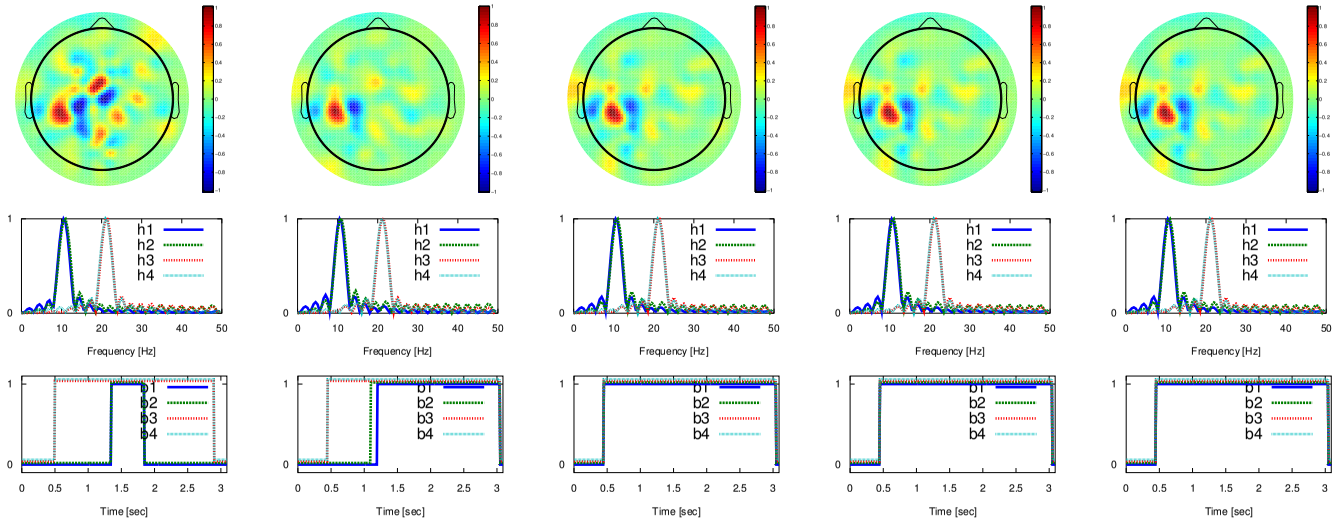


Figure 4.8: The variation of the accuracy rates by various regularization parameters, ϵ, in BCI competition IV dataset 1.



(a) $\epsilon = 0.04$ (b) $\epsilon = 0.08$ (c) $\epsilon = 0.12$ (d) $\epsilon = 0.16$ (e) $\epsilon = 0.4$

Figure 4.9: Examples of spatio-time-frequency patterns as the spatial weights, $\hat{\boldsymbol{w}}_1^{(1)}$ (top), the amplitude response of the filter, $\boldsymbol{h}_1, \ldots, \boldsymbol{h}_4$ (center), and the time windows, $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_4$ (bottom), designed by the CSTFP method with various ϵ in the result of subject *aa*. The vertical axis represents normalized amplitude.



(a) $\epsilon = 0.04$ (b) $\epsilon = 0.08$ (c) $\epsilon = 0.12$ (d) $\epsilon = 0.16$ (e) $\epsilon = 0.4$

Figure 4.10: Examples of spatio-time-frequency patterns as the spatial weights, $\hat{\boldsymbol{w}}_1^{(1)}$ (top), the amplitude response of the filter, $\boldsymbol{h}_1, \ldots, \boldsymbol{h}_4$ (center), and the time windows, $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_4$ (bottom), designed by the CSTFP method with various ϵ in the result of subject *av*. The vertical axis represents normalized amplitude.

Chapter 5

Regularization with Sensor Positions

This chapter proposes the regularization of using similarity of weighted signals in nearby sensors. First, a distance between electrodes of EEG measurement system is defined in Sec. 5.1. Then we introduce the regularization derived with the defined distance in Sec. 5.2. The regularization is adopted to the CSP method in Sec. 5.3. We illustrate the ability of the proposed regularized CSP method for extracting a local feature. The proposed method is demonstrated with artificial signals in Sec. 5.4. Furthermore, the result of classification of the BMI datasets by spatially weighting of the proposed method is shown in Sec. 5.5.

5.1 Distance Between Electrodes

We define a distance between electrodes on the arrangements used for EEG measurement. International 10-20, 10-10, and 10-5 methods [62, 77, 79] have stood as the de-facto standard of electrode arrangement. In these systems, locations on a head surface are described by relative distances between cranial landmarks over the head surface. Under an assumption that the shape of head is a sphere, the

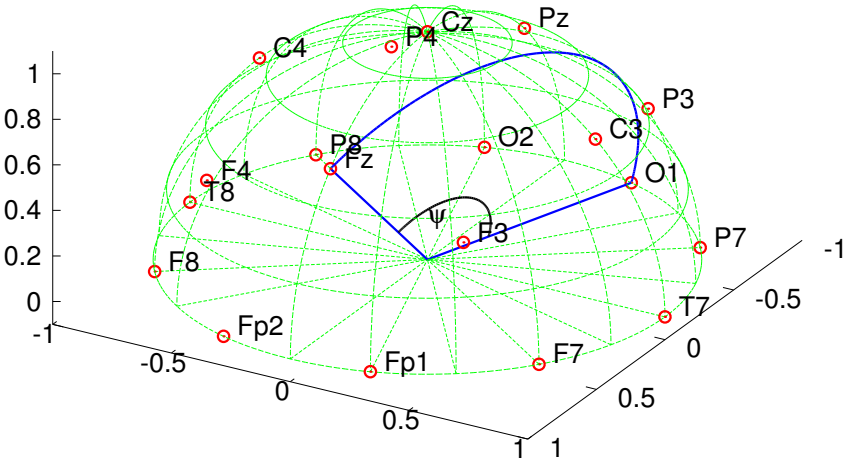


Figure 5.1: The electrode arrangement of the International 10-20 method on the orthogonal coordinates. The red circles represent the electrodes.

locations on head surface can be describe coordinates represented by $\xi = \{x, y, z\}$. We define the coordinates such as the axes of Fig. 5.1 that illustrates the electrode positions of the international 10-20 method.

Given the positions of two electrodes as

$$\xi_i = \{x_i, y_i, z_i\}, \tag{5.1}$$

and

$$\xi_j = \{x_j, y_j, z_j\}. \tag{5.2}$$

The question arising here is: how to define the distance between two points, on the head. The Euclidean distance defined by

$$d_{i,j} = \|\xi_i - \xi_j\| \tag{5.3}$$

is a straightforward solution. In this study, we define the perimeter of a sector the two sides of which are line segments between the origin and two electrode

position on the coordinates as the distance between two electrodes. Let ψ_{ij} be the angle between the line segments between the origin and ξ_i, and the origin and ξ_j. The distance by the perimeter is defined as $d_{ij} = \nu\psi_{ij}$, where $\nu = \|\epsilon_i\| = \|\epsilon_j\|$. Moreover, because

$$\cos\psi = \frac{<\xi_i, \xi_j>}{\nu^2} \tag{5.4}$$

and $\nu = 1$, d_{ij} can be represented by

$$d_{ij} = \arccos(x_i x_j + y_i y_j + z_i z_j). \tag{5.5}$$

The metric is illustrated in Fig. 5.1. In the figure, we show the distance between F_z and O_1 as an example. The length of the curve connecting F_z and O_1 is the defined distance by the metric.

5.2 Regularization

Consider a sensor array consisting of M sensors. A signal sample observed in the ith channel at a time instance is denoted by x_i. A set, $\{x_i\}_{i=1}^M$, forms a vector $\boldsymbol{x}$ defined as $\boldsymbol{x} = [x_1, \ldots, x_M]^T$. We obtain d_{ij} for $i, j = 1, \ldots, M$ as the distances between sensors by the metric defined in Sec. 5.1. To mainly evaluate the regularization costs between each sensor and its nearby sensors, the Gaussian metric between two points;

$$g_{ij} = \exp\left(-\frac{d_{ij}^2}{2p^2}\right), \tag{5.6}$$

is employed, where p denotes a parameter to tune the closeness of the two sensors. Then we define the cost;

$$P(\boldsymbol{w}) = E_{\boldsymbol{x}}\left[\sum_{i=1}^{M-1}\sum_{j=i+1}^{M} g_{ij}|w_i x_i - w_j x_j|^2\right], \tag{5.7}$$

which evaluates the mean of squared error between weighted signals observed in sensors that are located near each other, Note that the cost (5.7) becomes small as the weighted signals become similar in the nearby sensors.

Equation (5.7) can be transformed to matrix vector form as

$$P(\boldsymbol{w}) = E_{\boldsymbol{x}}[\boldsymbol{w}^T \boldsymbol{D}_x(\boldsymbol{C} - \boldsymbol{G})\boldsymbol{D}_x\boldsymbol{w}] = \boldsymbol{w}^T \boldsymbol{B}\boldsymbol{w}, \tag{5.8}$$

where $\boldsymbol{C}$ and $\boldsymbol{D}_x$ are diagonal matrices defined as

$$[\boldsymbol{C}]_{ii} = \sum_{k=1}^{M} g_{ik}, \quad [\boldsymbol{D}_x]_{ii} = x_i, \quad i = 1, \ldots, M, \tag{5.9}$$

each element of $\boldsymbol{G} \in \mathbb{R}^{M\times M}$ is defined as

$$[\boldsymbol{G}]_{ij} = g_{ij}, \quad i, j = 1, \ldots, M, \tag{5.10}$$

and

$$\boldsymbol{B} = E_{\boldsymbol{x}}[\boldsymbol{D}_x(\boldsymbol{C} - \boldsymbol{G})\boldsymbol{D}_x]. \tag{5.11}$$

To take expectation over $\boldsymbol{x}$ for obtaining $\boldsymbol{B}$, we can use the sample average of observed signals.

5.3 CSP with Regularization with Sensor Positions

The CSP method [42, 43] is effective in feature extraction and classification for two-class MI-BMI. In this section, we first review the standard CSP method. Then, we exhibit how to apply the regularization in described Sec. 5 for finding CSP.

By adding the regularization term given as (5.7), the modified regularized optimization problem is defined as

$$\begin{aligned} &\min_{\boldsymbol{w}} \boldsymbol{w}^T(\boldsymbol{\Sigma}_c + \gamma\boldsymbol{B})\boldsymbol{w}, \\ &\text{subject to } \boldsymbol{w}^T(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)\boldsymbol{w} = 1, \end{aligned} \tag{5.12}$$

where γ is a combination coefficient. If the matrices of $\boldsymbol{\Sigma}_c + \gamma \boldsymbol{B}$ and $\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2$ are nonsingular, (5.12) is equivalent to the generalized eigenvalue problem:

$$(\boldsymbol{\Sigma}_c + \gamma \boldsymbol{B})\boldsymbol{w} = \lambda(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)\boldsymbol{w}. \tag{5.13}$$

5.4 Simulation by Artificial Signals

An analysis of the proposed method by a toy experiment with artificial signals is given. We used the mixture of synthetic source and noise signals. We assume to know the spatial distributions of the source signals. The spatial weights derived by the CSP method and the regularized CSP method were compared with the true distribution.

We assumed a 2-class BMI where observed EEG signals are modeled by a mixture of narrow-band signals. In this model, two signals, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, belonging to class 1 and class 2, respectively, are given by

$$\boldsymbol{x}_1[n] = \boldsymbol{a}_1[n]s[n] + \boldsymbol{\eta}, \quad \boldsymbol{x}_2[n] = \boldsymbol{a}_2[n]s[n] + \boldsymbol{\eta}, \tag{5.14}$$

for $n = 1, \ldots, N$, where $\boldsymbol{x}_1[n], \boldsymbol{x}_2[n] \in \mathbb{R}^M$ denote vectors representing a signal observed at discrete time instance n, N denotes the number of time instances, M denotes the number of channels, $s[n] \in \mathbb{R}$ denotes a source signal of feature component, $\boldsymbol{a}_1, \boldsymbol{a}_2 \in \mathbb{R}^M$ denote vectors defined by $\boldsymbol{a}_i = [a_{i1}, \ldots, a_{iM}]^T$, $a_{im} \in \mathbb{R}$ denote an amplitude of the source at the mth channel for class i, and $\boldsymbol{\eta} \in \mathbb{R}^M$ denote a stochastic noise.

The simulation settings for generating artificial signals were shown in Table 5.1. The observed signal in 2 channels are shown in Fig. 5.5.

The topographically plotted spatial weights given by by the standard CSP and the proposed methods are shown in Fig. 5.6. The parameters for the proposed

Table 5.1: The settings for generating artificial signals

Parameter	Value and distribution
Number of channels	118
Electrodes arrangement	Int'l extended 10-20
Number of samples	512
Sampling frequency	512
Spectrum of the source signals	Fig. 5.3
Distributions	Fig. 5.4
Noise $[\eta]_m$	$\mathcal{N}(0, 0.1)$

method are 0.05 for p and 10^8 for γ. Compared to the standard CSP, the weight designed by the proposed method resulted in the large weight coefficients concentrated at the certain spots. Moreover we can observe in Fig. 5.6b that the topographical maps of the spatial weights given by the proposed method are similar to the true distribution maps shown in Fig 5.4.

5.5 Classification Experiment of BMI Datasets

We compared performance in a two-class classification of EEG signals during motor imagery using the proposed method to those using the standard CSP and the spatially regularized CSP (SRCSP) [115], respectively.

5.5.1 Data Description

We used dataset IVa from BCI competition III. The details of the dataset are shown in 4.7.1. The lowpass filter whose cutoff frequency is 50 Hz was applied to recorded signals and the filtered signals was donwsampled to 100 Hz. Further-

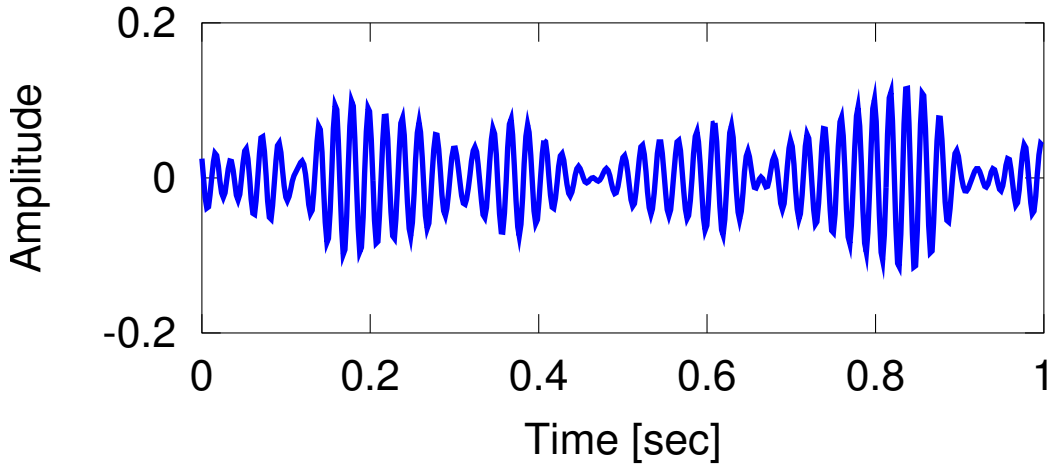


Figure 5.2: Source signals in the artificial signals

more, the signals were bandpass filtered with the passband of 7–30 Hz that is a band including mu and beta rhythms. The dataset for each subject consisted of signals of 140 trials per a class. The signal length for each trial is 3.5 seconds.

5.5.2 Features for Classification

The following feature vector was used for classification. In each case of $c = 1$ and $c = 2$, we solve (3.8) or (5.13), and then we got the eigenvectors corresponding to the largest eigenvalues in each eigenvalue problem defined by $\hat{\boldsymbol{w}}_1$ and $\hat{\boldsymbol{w}}_2$, respectively. By using the weight vectors, the feature vector was defined as

$$\boldsymbol{y} = [\sigma^2(\boldsymbol{X}, \hat{\boldsymbol{w}}_1), \sigma^2(\boldsymbol{X}, \hat{\boldsymbol{w}}_2)]^T. \tag{5.15}$$

5.5.3 Results

Linear discriminant analysis [61] were used for classifying the extracted feature vectors. For the proposed methods, we used the signals that were observed in the

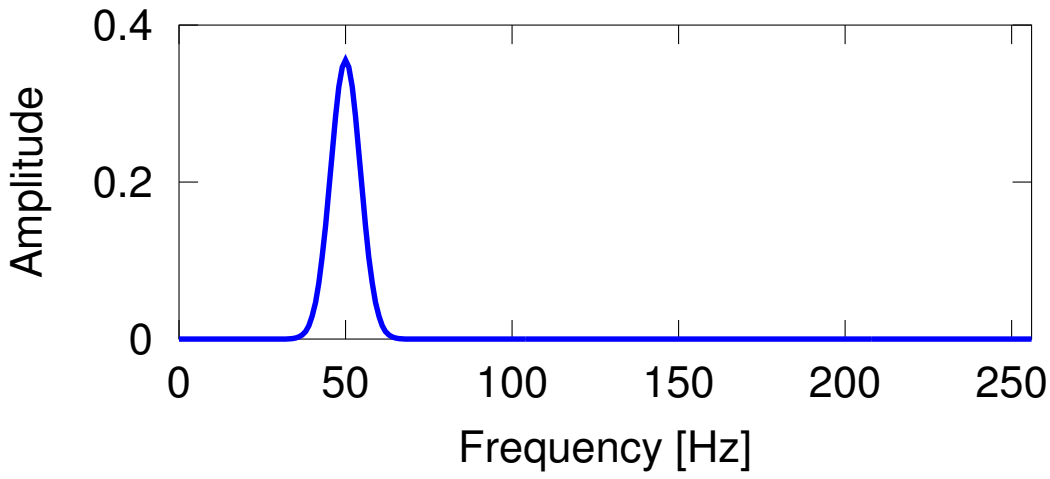


Figure 5.3: Spectrum of the source signals in the artificial signals

intervals between the tasks to form the regularization term (5.7) for each subject.

The classification accuracy was given by training for the spatial weights and the classifier with randomly chosen 100 samples, and testing with the remaining samples. An average accuracy over 100 times of this procedure is shown in in Table 5.2. The parameter in (5.6) were set to $p = 0.05$ for SRCSP and the proposed method, The parameter was chosen out of $\gamma \in \{10^0, 10^{0.1}, \ldots, 10^{30}\}$. The best accuracy among the parameters for each subject is shown in Table 5.2. In the result of Table 5.2, for *aa*, *al*, *av*, *aw*, and *ay*, γ were set to $10^{9.4}$, 10^{12}, $10^{11.1}$, 10^{10}, and $10^{12.5}$, respectively, in SRCSP. In the proposed method, γ were set to $10^{9.9}$, $10^{14.3}$, $10^{12.7}$, $10^{11.1}$, and $10^{14.4}$, respectively. The both of the regularized CSP slightly outperform the standard CSP method in the classification accuracy for all subjects.

Table 5.3 also shows classification accuracy, however when the number of the training samples is considerably reduced to only five samples. As the same as in Table 5.2, the parameters performing the best classification accuracy were chosen

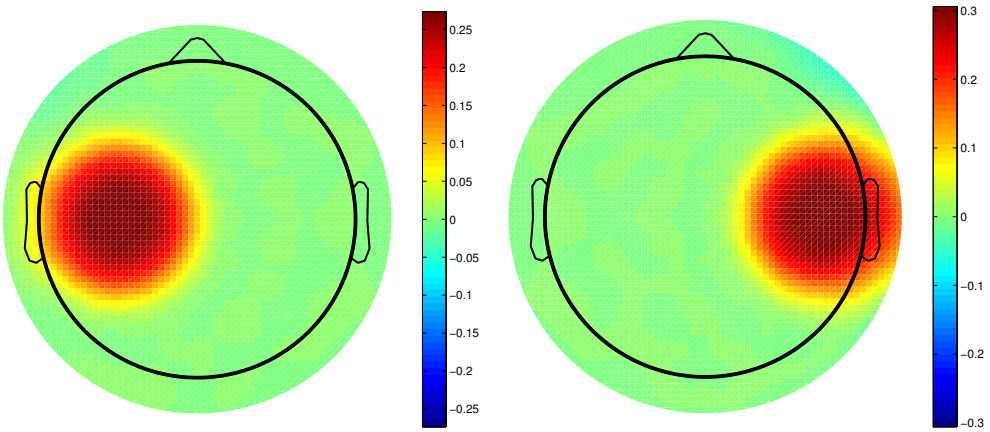


Figure 5.4: Spatial distributions of the source signal in the artificial signals.

Table 5.2: Classification accuracy rates [%] given by 100 training samples per a class.

	Subject					
	aa	*al*	*av*	*aw*	*ay*	Ave.
CSP	81.4	94.8	53.1	92.9	89.6	82.3
SRCSP	82.2	95.2	64.2	94.3	92.7	85.7
Proposed	82.0	95.4	66.2	94.6	93.0	86.3

out of the candidates. For *aa*, *al*, *av*, *aw*, and *ay*, γ were set to $10^{10.4}$, $10^{10.5}$, $10^{12.5}$, $10^{10.8}$, and $10^{11.4}$, respectively, in SRCSP. In the proposed method, γ were set to $10^{13.6}$, $10^{13.5}$, $10^{16.9}$, $10^{13.9}$, and $10^{15.0}$, respectively. We can observe significant improvement of the accuracy rates for subjects *al* and *ay* by the regularizations. The results suggest that the proposed regularization can improve the accuracy even if the number of training samples available is small.

The topographically plotted spatial weights for subject *ay* is shown in Fig. 5.7. All samples in the dataset were used to find the spatial weights. The parameters

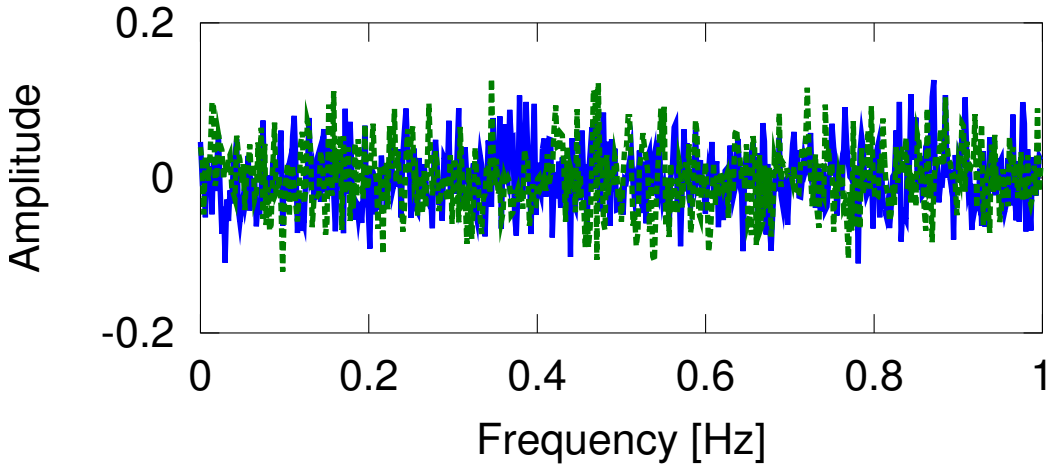


Figure 5.5: Examples of the observed signals in the artificial signals

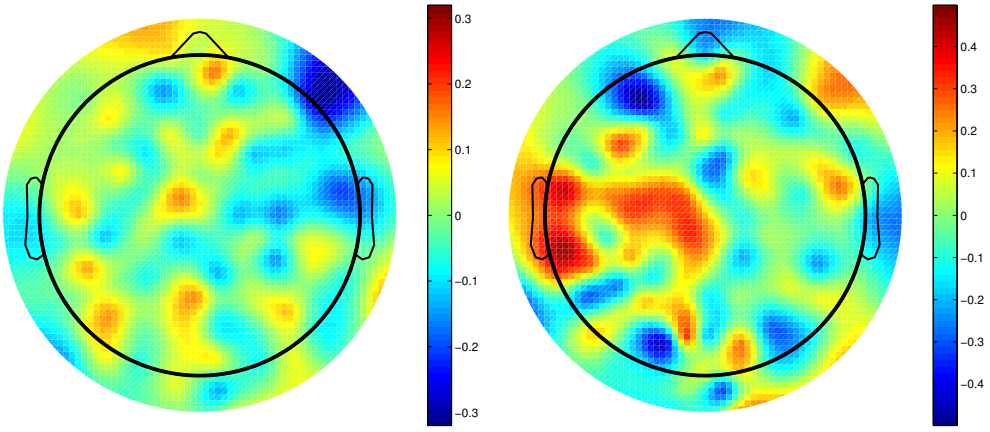
Table 5.3: Classification accuracy rates [%] given by 5 training samples per a class.

	Subject					
	aa	*al*	*av*	*aw*	*ay*	Ave.
CSP	52.6	67.0	50.3	61.2	51.0	56.4
SRCSP	58.4	77.8	54.2	71.6	78.6	68.1
Proposed	59.2	81.3	54.4	71.3	79.3	69.1

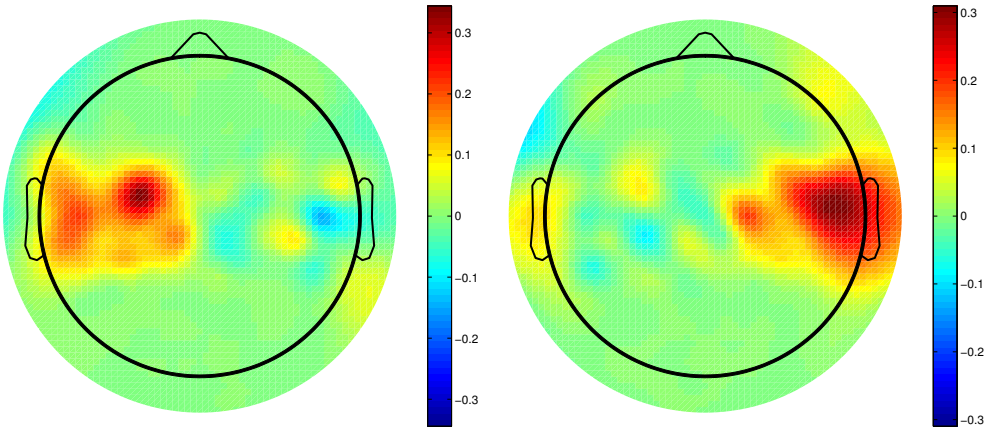
of the proposed method, p and γ, were set to 0.05 and 10^{15}, respectively. Comparing to the standard CSP, the electrodes which have large coefficients do not be scattered spatially in the proposed method.

5.6 Conclusion

We have proposed the regularization based on the similarity of observed signals in the nearby sensors for feature extraction problem in an EEG sensor array. More-

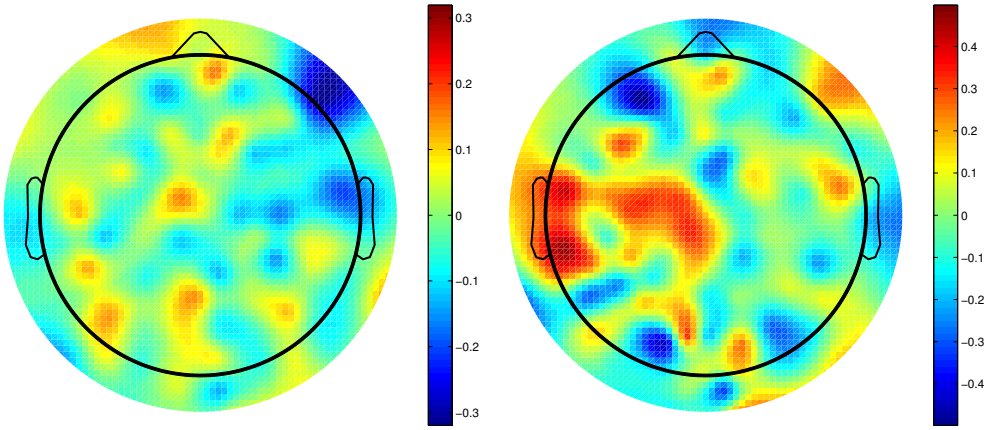


(a) The standard CSP

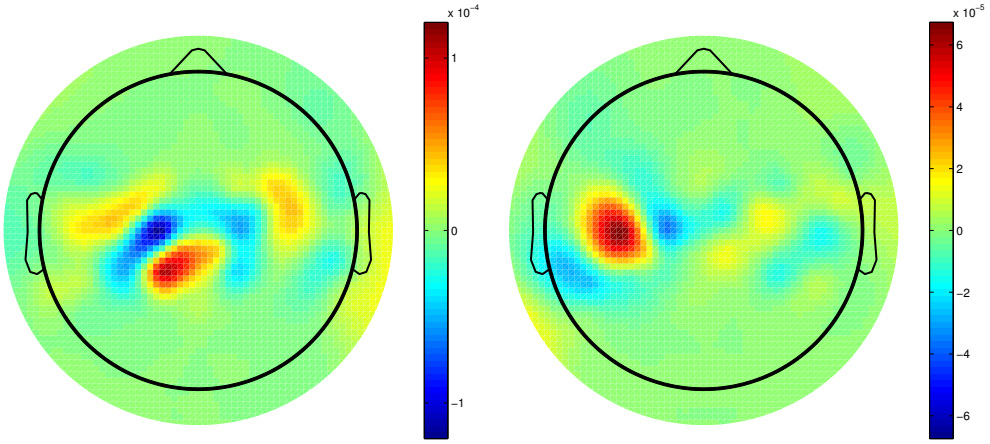


(b) CSP with the proposed method

Figure 5.6: Topographical maps of the spatial weights that are as the eigenvectors corresponding to the largest eigenvalues of (3.8) and (5.13) ($c = 1$ (left) and $c = 2$ (right)) in the experiment using the artificial signals.



(a) The standard CSP



(b) CSP with the regularization

Figure 5.7: Topographical maps of the spatial weights, $\hat{\boldsymbol{w}}_1^{(1)}$ and $\hat{\boldsymbol{w}}_1^{(2)}$, for subject *ay* in the experiment using the real-world EEG signals.

over, we have illustrated how to apply the proposed regularization to the CSP method. The experimental results demonstrated that the proposed regularization improves classification accuracy in a setting of the small number of samples.

Chapter 6

Conclusions and Open Problems

6.1 Conclusions

We conclude the study in this section. We have discussed the data-driven feature extraction methods for classification of EEG signals in this study. The methods are proposed to address the problems;

1) simultaneous design of spatial weights, frequency filters, and time windows is impossible

2) extraction of multiple feature components modeled with different spatial patterns, frequency patterns, and time patterns for each other is impossible,

3) overfitting of the designed parameters for the spatial weights often happens.

The CSTFP method has been proposed to address the problems 1) and 2). We define a signal extraction model composed of the sets of the parameters of the spatial filters, the frequency filters, and the time windows. The extraction model includes the multiple filter sets having different spatial patterns, different

frequency patterns, and different time patterns and each set extracts different feature components. The parameters of the sets are optimized with the criterion based on the CSP method. Therefore, the filter sets are optimized for a specific dataset. The conducted experiments suggest that the CSTFP method can extract the feature components that have different patterns. The CSTFP achieves higher classification accuracy rates than those of the existing methods. The experimental results with the CSTFP method reveal a possibility that the feature components associated with the MI are in the several bands.

The regularization using information of sensor arrangement has been proposed to address the problem 3). Generally, the regularizations aim to add additional information such as bounds of the vector norm and smoothness to an original cost function. The proposed regularization adds an assumption that spatially-weighted signals observed in nearby sensors are similar. We have proposed the CSP algorithm with the regularization. In the experiment of classification of the BMI datasets, the regularized CSP performs high classification accuracy even in the situation where only small learning datasets are available.

6.2 Open Problems

Finally, we show some open problems in the data-driven feature extraction for MI-BMI. Although we have proposed the solutions for the problems, the following challenging problems have been remaining.

6.2.1 Adaptive System for Unstationary Feature Components

It is impossible to pursue feature components that are unstationary over time. The feature components can be unstationary, because the feature components are affected by the mental state of a subject. Moreover, feedbacking to a user in the BMI systems has influence to the change of the feature components [3, 9]. To pursue the changes of the feature components, the parameters must be updated while the subject uses BMI. For the adaptive updating, methods which take advantage of common components [116] among the different MI tasks and fast algorithms for designing the parameters are required.

6.2.2 Application to Self-Paced BMIs

The interesting other problem is application to self-paced (asynchronous) BMIs [25, 117]. In this study, our target is target cue-based BMIs in which a subject starts to perform a desire task when the cue appeared. We set a period for classification based on the cue information. On the other hand, an objective of the asynchronous BMIs is that a subject performs a task without cue when the subject wants to input it. The asynchronous BMIs require classification of idle state. And the BMI works as an interface only if a subject performs specific tasks. In this way, the asynchronous BMIs provide more user-friendly interface than the cue-based BMIs. However, the data-driven methods for MI-BMI do not achieve enough accuracy on classifying the idle state [113]. The low accuracy is caused by the idle state including any brain states except for the BMI tasks. To solve the problem, we need to define brain states to be possible to occur as the idle state. And a multiclass classification technique for classifying them is needed.

6.2.3 Application to Rehabilitation

In the application for rehabilitation for patient with brain damage, we need the adaptive methods shown in the above to pursue the change of brain functions. Additionally, a learning dataset may be not available because of lack of correct labels in case of the rehabilitation. For dealing with this problem, a method with the combination of heuristic and adaptive methods is required. For example, at the beginning of the rehabilitation the parameters are designed with with datasets given by the other patients and healthy subjects. And then, as the patient does the rehabilitation, the parameters are designed for the patient by the adaptive method.

Bibliography

[1] S. Y. Bookheimer, M. H. Strojwas, M. S. Cohen, A. M. Saunders, M. A. Pericak-Vance, J. C. Mazziotta, and G. W. Small, "Patterns of brain activation in people at risk for Alzheimer's disease," *New England Journal of Medicine*, vol. 343, no. 7, pp. 450–456, Aug. 2000.

[2] C. Iadecola, "Neurovascular regulation in the normal brain and in Alzheimer's disease," *Nature Reviews Neuroscience*, vol. 5, no. 5, pp. 347–360, May 2004.

[3] G. Dornhege, J. d. R. Millan, T. Hinterberger, D. McFarland, and K.-R. Muller, Eds., *Toward Brain-Computer Interfacing*. The MIT Press, 2007.

[4] P. A. Bandettini, E. C. Wong, R. S. Hinks, R. S. Tikofsky, and J. S. Hyde, "Time course EPI of human brain function during task activation," *Magnetic Resonance in Medicine*, vol. 25, no. 2, pp. 390–397, Jun. 1992.

[5] M. A. L. Nicolelis, "Brain-machine interfaces to restore motor function and probe neural circuits," *Nature Reviews Neuroscience*, vol. 4, no. 5, pp. 417–422, May 2003.

[6] A. Villringer and B. Chance, "Non-invasive optical spectroscopy and imaging of human brain function," *Trends in Neurosciences*, vol. 20, no. 10, pp. 435–442, Oct. 1997.

[7] K. K. Kwong, J. W. Belliveau, D. A. Chesler, I. E. Goldberg, R. M. Weisskoff, B. P. Poncelet, D. N. Kennedy, B. E. Hoppel, M. S. Cohen, and R. Turner, "Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation." *Proceedings of the National Academy of Sciences*, vol. 89, no. 12, pp. 5675–5679, Jun. 1992.

[8] S. Sanei and J. Chambers, *EEG Signal Processing*. Wiley-Interscience, 2007.

[9] 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.

[10] M. A. Lebedev and M. A. L. Nicolelis, "Brain machine interfaces: past, present and future," *Trends in Neurosciences*, vol. 29, no. 9, pp. 536–546, 2006.

[11] J. K. Chapin, K. A. Moxon, R. S. Markowitz, and M. A. Nicolelis, "Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex." *Nature Neuroscience*, vol. 2, no. 7, pp. 664–670, 1999.

[12] M. Velliste, S. Perel, M. C. Spalding, A. S. Whitford, and A. B. Schwartz, "Cortical control of a prosthetic arm for self-feeding," *Nature*, vol. 453, no. 7198, pp. 1098–1101, 2008.

[13] L. R. Hochberg, M. D. Serruya, G. M. Friehs, J. A. Mukand, M. Saleh, A. H. Caplan, A. Branner, D. Chen, R. D. Penn, and J. P. Donoghue, "Neuronal ensemble control of prosthetic devices by a human with tetraplegia." *Nature*, vol. 442, no. 7099, pp. 164–71, 2006.

[14] 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.

[15] E. Niedermeyer and F. H. L. Da Silva, *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 2005.

[16] J. Mellinger, G. Schalk, C. Braun, H. Preissl, W. Rosenstiel, N. Birbaumer, and A. Kbler, "An MEG-based brain computer interface (BCI)," *NeuroImage*, vol. 36, no. 3, pp. 581–593, 2007.

[17] R. Sitaram, A. Caria, R. Veit, T. Gaber, G. Rota, A. Kuebler, and N. Birbaumer, "fMRI brain-computer interface: a tool for neuroscientific research and treatment," *Intell. Neuroscience*, vol. 2007, pp. 1:1–1:10, 2007.

[18] S. M. Coyle, T. E. Ward, and C. M. Markham, "Brain—computer interface using a simplified functional near-infrared spectroscopy system," *Journal of Neural Engineering*, vol. 4, no. 3, p. 219, 2007.

[19] 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.

[20] C. Zhang, Y. Kimura, H. Higashi, and T. Tanaka, "A simple platform of brain-controlled mobile robot and its implementation by SSVEP," in *Pro-*

ceedings of The 2012 International Joint Conference on Neural Networks (IJCNN), 2012, pp. 1–7.

[21] K.-R. Müller, M. Tangermann, G. Dornhege, M. Krauledat, G. Curio, and B. Blankertz, "Machine learning for real-time single-trial EEG-analysis: From brain—computer interfacing to mental state monitoring," *Journal of Neuroscience Methods*, vol. 167, no. 1, pp. 82–90, Jan. 2008.

[22] R. Scherer, G. R. Muller, C. Neuper, B. Graimann, and G. Pfurtscheller, "An asynchronously controlled EEG-based virtual keyboard: improvement of the spelling rate," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 979–984, 2004.

[23] G. Townsend, B. K. LaPallo, C. B. Boulay, D. J. Krusienski, G. E. Frye, C. K. Hauser, N. E. Schwartz, T. M. Vaughan, J. R. Wolpaw, and E. W. Sellers, "A novel P300-based brain—computer interface stimulus presentation paradigm: Moving beyond rows and columns," *Clinical Neurophysiology*, vol. 121, no. 7, pp. 1109–1120, Jul. 2010.

[24] F. Galán, M. Nuttin, E. Lew, P. W. Ferrez, G. Vanacker, J. Philips, and J. d. R. Millán, "A brain-actuated wheelchair: Asynchronous and non-invasive Brain—computer interfaces for continuous control of robots," *Clinical Neurophysiology*, vol. 119, no. 9, pp. 2159–2169, Sep. 2008.

[25] R. Leeb, D. Friedman, G. R. Müller-Putz, R. Scherer, M. Slater, and G. Pfurtscheller, "Self-paced (asynchronous) BCI control of a wheelchair in virtual environments: a case study with a tetraplegic," *Computational Intelligence and Neuroscience*, vol. 2007, pp. 7:1—-7:12, 2007.

[26] K. Choi and A. Cichocki, "Control of a Wheelchair by Motor Imagery in Real Time," in *Intelligent Data Engineering and Automated Learning — IDEAL 2008*, ser. Lecture Notes in Computer Science, C. Fyfe, D. Kim, S.-Y. Lee, and H. Yin, Eds. Springer Berlin / Heidelberg, 2008, vol. 5326, pp. 330–337.

[27] C. J. B. Rao, P. Shenoy, R. Chalodhorn, and R. P. N, "Control of a humanoid robot by a noninvasive brain—computer interface in humans," *Journal of Neural Engineering*, vol. 5, no. 2, p. 214, 2008.

[28] B. Blankertz, K.-R. Müller, G. Curio, T. M. Vaughan, G. Schalk, J. R. Wolpaw, A. Schlogl, C. Neuper, G. Pfurtscheller, T. Hinterberger, M. Schroder, and N. Birbaumer, "The BCI competition 2003: Progress and perspectives in detection and discrimination of EEG single trials," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1044–1051, 2004.

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

[30] G. Pfurtscheller, G. R. Muller-Putz, R. Scherer, and C. Neuper, "Rehabilitation with brain-computer interface systems," *Computer*, vol. 41, no. 10, pp. 58–65, 2008.

[31] B. Várkuti, C. Guan, Y. Pan, K. S. Phua, K. K. Ang, C. W. K. Kuah, K. Chua, B. T. Ang, N. Birbaumer, and R. Sitaram, "Resting state changes in functional connectivity correlate with movement recovery for BCI and

robot-assisted upper-extremity training after stroke," *Neurorehabilitation and Neural Repair*, vol. 27, no. 1, pp. 53–62, Jan. 2013.

[32] A. Pascual-Leone, C. Freitas, L. Oberman, J. Horvath, M. Halko, M. Eldaief, S. Bashir, M. Vernet, M. Shafi, B. Westover, A. Vahabzadeh-Hagh, and A. Rotenberg, "Characterizing brain cortical plasticity and network dynamics across the age-span in health and disease with TMS-EEG and TMS-fMRI," *Brain Topography*, vol. 24, no. 3-4, pp. 302–315, 2011.

[33] S. J. Page, P. Levine, S. A. Sisto, and M. V. Johnston, "Mental practice combined with physical practice for upper-limb motor deficit in subacute stroke," *Physical Therapy*, vol. 81, no. 8, pp. 1455–1462, Aug. 2001.

[34] J. J. Daly and J. R. Wolpaw, "Brain—computer interfaces in neurological rehabilitation," *The Lancet Neurology*, vol. 7, no. 11, pp. 1032–1043, Nov. 2008.

[35] K. K. Ang, C. Guan, K. Sui Geok Chua, B.-T. Ang, C. Kuah, C. Wang, K. S. Phua, Z. Y. Chin, and H. Zhang, "Clinical study of neurorehabilitation in stroke using EEG-based motor imagery brain-computer interface with robotic feedback," in *Proceedings of 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2010, pp. 5549–5552.

[36] V. V. Nikulin, F. U. Hohlefeld, A. M. Jacobs, and G. Curio, "Quasi-movements: A novel motor-cognitive phenomenon," *Neuropsychologia*, vol. 46, no. 2, pp. 727–742, 2008.

[37] J. R. Wolpaw and D. J. McFarland, "Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans," *The National Academy of Sciences*, vol. 101, no. 51, pp. 17 849–17 854, 2004.

[38] S. Palva and J. M. Palva, "New vistas for alpha-frequency band oscillations," *Trends in neurosciences*, vol. 30, no. 4, pp. 150–158, Apr. 2007.

[39] G. Pfurtscheller, C. Neuper, D. Flotzinger, and M. Pregenzer, "EEG-based discrimination between imagination of right and left hand movement," *Electroencephalography and Clinical Neurophysiology*, vol. 103, no. 6, pp. 642–651, Dec. 1997.

[40] 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.

[41] M. Grosse-Wentrup and M. Buss, "Multiclass common spatial patterns and information theoretic feature extraction," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 8, pp. 1991–2000, 2008.

[42] 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.

[43] H. Ramoser, J. Muller-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.

[44] 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.

[45] 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.

[46] R. Tomioka, G. Dornhege, G. Nolte, B. Blankertz, K. Aihara, and K. R. Müller, "Spectrally weighted common spatial pattern algorithm for single trial eeg classification," *Dept. Math. Eng., Univ. Tokyo, Tokyo, Japan, Tech. Rep*, vol. 40, 2006.

[47] 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 (ICASSP)*, 2011, pp. 585–588.

[48] ——, "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.

[49] N. Tomida, H. Higashi, and T. Tanaka, "A joint tensor diagonalization approach to active data selection for EEG classification," in *Proceedings of 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 983–987.

[50] 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 2008 International Joint Conference on Neural Networks (IJCNN)*, 2008, pp. 2390–2397.

[51] ——, "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.

[52] 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.

[53] 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.

[54] 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, p. 229, 2006.

[55] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 41–56, 2008.

[56] W. Wu, X. Gao, B. Hong, and S. Gao, "Classifying single-trial EEG during Motor imagery by iterative spatio-spectral patterns learning (ISSPL),"

IEEE Transactions on Biomedical Engineering, vol. 55, no. 6, pp. 1733–1743, 2008.

[57] Q. Novi, C. Guan, T. H. Dat, and P. Xue, "Sub-band common spatial pattern (SBCSP) for brain-computer interface," in *Proceedings of 3rd International IEEE/EMBS Conference on Neural Engineering*, 2007, pp. 204–207.

[58] K. P. Thomas, C. Guan, C. T. Lau, A. P. Vinod, and K. K. Ang, "A new discriminative common spatial pattern method for motor imagery brain computer interfaces," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 11, pp. 2730–2733, 2009.

[59] G. Sun, J. Hu, and G. Wu, "A novel frequency band selection method for common spatial pattern in motor imagery based brain computer interface," in *Proceedings of The 2010 International Joint Conference on Neural Networks (IJCNN)*, 2010, pp. 1–6.

[60] A. N. Tikhonov and V. Y. Arsenin, *Solutions of ill-posed problems*. V. H. Winston & Sons, 1977.

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

[62] J. Malmivuo and R. Plonsey, *Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields*. Oxford University Press, 1995.

[63] S. S. Haykin, *Adaptive Filter Theory*. Pearson Education India, 2005.

[64] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. Wiley, 2002.

[65] A. Cichocki and S. Sanei, "EEG/MEG signal processing," *Computational intelligence and neuroscience*, p. 97026, 2007.

[66] M. Welling, "Fisher linear discriminant analysis," *Science*, vol. 1, no. 2, pp. 1–3, 2009.

[67] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.

[68] R. G. Brereton and G. R. Lloyd, "Support vector machines for classification and regression," *The Analyst*, vol. 135, no. 2, pp. 230–267, 2010.

[69] A. Lecuyer, F. Lotte, R. B. Reilly, R. Leeb, M. Hirose, and M. Slater, "Brain-computer interfaces, virtual reality, and videogames," *Computer*, vol. 41, no. 10, pp. 66–72, 2008.

[70] N. Sharma, V. M. Pomeroy, and J.-C. Baron, "Motor imagery: A backdoor to the motor system after stroke?" *Stroke*, vol. 37, no. 7, pp. 1941–1952, Jul. 2006.

[71] R. Llinas, "Neuron," *Scholarpedia*, vol. 3, p. 1490, 2008.

[72] S. A. Huettel, A. W. Song, and G. McCarthy, *Functional Magnetic Resonance Imaging*. Sinauer Associates Sunderland, 2004.

[73] C. S. Roy and C. S. Sherrington, "On the regulation of the blood-supply of the brain," *The Journal of physiology*, vol. 11, no. 1-2, p. 85, 1890.

[74] A. Villringer, J. Planck, C. Hock, L. Schleinkofer, and U. Dirnagl, "Near infrared spectroscopy (NIRS): A new tool to study hemodynamic changes

during activation of brain function in human adults," *Neuroscience Letters*, vol. 154, no. 1—2, pp. 101–104, May 1993.

[75] S. Baillet, J. C. Mosher, and R. M. Leahy, "Electromagnetic brain mapping," *IEEE Signal Processing Magazine*, vol. 18, no. 6, pp. 14–30, 2001.

[76] P. L. Nunez and R. Srinivasan, "Electroencephalogram," *Scholarpedia*, vol. 2, p. 1348, 2007.

[77] M. R. Nuwer, "Recording electrode site nomenclature," *Journal of Clinical Neurophysiology*, vol. 4, no. 2, p. 121, 1987.

[78] E. Suarez, M. D. Viegas, M. Adjouadi, and A. Barreto, "Relating induced changes in EEG signals to orientation of visual stimuli using the ESI-256 machine." *Biomedical sciences instrumentation*, vol. 36, p. 33, 2000.

[79] R. Oostenveld and P. Praamstra, "The five percent electrode system for high-resolution EEG and ERP measurements," *Clinical Neurophysiology*, vol. 112, no. 4, pp. 713–719, 2001.

[80] V. Jurcak, D. Tsuzuki, and I. Dan, "10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems." *NeuroImage*, vol. 34, no. 4, pp. 1600–1611, 2007.

[81] D. B. Ryan, G. E. Frye, G. Townsend, D. R. Berry, S. Mesa-G, N. A. Gates, and E. W. Sellers, "Predictive spelling with a P300-based brain—computer interface: Increasing the rate of communication," *International Journal of Human-Computer Interaction*, vol. 27, no. 1, pp. 69–84, Dec. 2010.

[82] A. Lenhardt, M. Kaper, and H. J. Ritter, "An adaptive P300-based online brain computer interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 16, no. 2, pp. 121–130, 2008.

[83] S. J. Luck, *An introduction to the event-related potential technique*, ser. Cognitive neuroscience. MIT Press, 2005.

[84] P. Sajda, A. Gerson, K.-R. Müller, B. Blankertz, and L. Parra, "A data analysis competition to evaluate machine learning algorithms for use in brain-computer interfaces," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2, pp. 184–185, 2003.

[85] 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. 581–617, 2006.

[86] W. S. Pritchard, "Psychophysiology of P300." *Psychological Bulletin*, vol. 89, no. 3, pp. 506–540, 1981.

[87] S. A. Hillyard and M. Kutas, "Electrophysiology of Cognitive Processing," *Annual Review of Psychology*, vol. 34, no. 1, pp. 33–61, Jan. 1983.

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

[89] E. W. Sellers and E. Donchin, "A P300-based brain—computer interface: Initial tests by ALS patients," *Clinical Neurophysiology*, vol. 117, no. 3, pp. 538–548, Mar. 2006.

[90] E. Donchin, K. M. Spencer, and R. Wijesinghe, "The mental prosthesis: assessing the speed of a P300-based brain-computer interface," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 2, pp. 174–179, 2000.

[91] D. J. Krusienski, E. W. Sellers, F. Cabestaing, S. Bayoudh, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "A comparison of classification techniques for the P300 Speller," *Journal of Neural Engineering*, vol. 3, no. 4, p. 299, 2006.

[92] D. J. Krusienski, E. W. Sellers, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "Toward enhanced P300 speller performance," *Journal of Neuroscience Methods*, vol. 167, no. 1, pp. 15–21, 2008.

[93] N. R. Draper and H. Smith, *Applied Regression Analysis*, J. Wiley, Ed. Springer-Verlag, 1998.

[94] D. Zhu, J. Bieger, G. G. Molina, and R. M. Aarts, "A survey of stimulation methods used in SSVEP-based BCIs," *Intell. Neuroscience*, vol. 2010, pp. 1:1–1:12, 2010.

[95] F.-B. Vialatte, M. Maurice, J. Dauwels, and A. Cichocki, "Steady-state visually evoked potentials: Focus on essential paradigms and future perspectives," *Progress in Neurobiology*, vol. 90, no. 4, pp. 418–438, Apr. 2010.

[96] Y. Wang, X. Gao, B. Hong, C. Jia, and S. Gao, "Brain-computer interfaces based on visual evoked potentials," *IEEE Engineering in Medicine and Biology Magazine*, vol. 27, no. 5, pp. 64–71, 2008.

[97] D. Regan, "Some characteristics of average steady-state and transient responses evoked by modulated light," *Electroencephalography and Clinical Neurophysiology*, vol. 20, no. 3, pp. 238–248, 1966.

[98] ——, "Steady-state evoked potentials," *Journal of the Optical Society of America*, vol. 67, no. 11, pp. 1475–1489, 1977.

[99] X. Gao, D. Xu, M. Cheng, and S. Gao, "A BCI-based environmental controller for the motion-disabled," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2, pp. 137–140, 2003.

[100] 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, 2000.

[101] G. R. Muller-Putz, R. Scherer, C. Neuper, and G. Pfurtscheller, "Steady-state somatosensory evoked potentials: suitable brain signals for brain-computer interfaces?" *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 1, pp. 30–37, 2006.

[102] D. S. Goodin, K. C. Squires, and A. Starr, "Long latency event-related components of the auditory evoked potential in dementia." *Brain : a Journal of Neurology*, vol. 101, no. 4, pp. 635–648, Dec. 1978.

[103] H. Higashi, T. M. Rutkowski, Y. Washizawa, A. Cichocki, and T. Tanaka, "EEG auditory steady state responses classification for the novel BCI," in *Proceedings of 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2011, pp. 4576–4579.

[104] H. Bakardjian, T. Tanaka, and A. Cichocki, "Optimization of SSVEP brain responses with application to eight-command Brain—Computer Interface," *Neuroscience Letters*, vol. 469, no. 1, pp. 34–38, Jan. 2010.

[105] Y. Kimura, T. Tanaka, H. Higashi, and N. Morikawa, "SSVEP-based brain–computer interfaces using FSK-modulated visual stimuli," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 10, pp. 2831–2838, 2013.

[106] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3/4, pp. 321–377, 1936.

[107] T. W. Anderson, *An introduction to multivariate statistical analysis*. Wiley New York, 1958, vol. 2.

[108] Z. Lin, C. Zhang, W. Wu, and X. Gao, "Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 12, pp. 2610–2614, 2006.

[109] A. Kübler, F. Nijboer, J. Mellinger, T. M. Vaughan, H. Pawelzik, G. Schalk, D. J. McFarland, N. Birbaumer, and J. R. Wolpaw, "Patients with ALS can use sensorimotor rhythms to operate a brain-computer interface," *Neurology*, vol. 64, no. 10, pp. 1775–1777, May 2005.

[110] H. Higashi, T. Tanaka, and A. Funase, "Classification of single trial EEG during imagined hand movement by rhythmic component extraction," in

Proceedings of 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2009, pp. 2482–2485.

[111] A. Pascual-Leone, A. Amedi, F. Fregni, and L. B. Merabet, "The plastic human brain cortex," *Annual Review of Neuroscience*, vol. 28, no. 1, pp. 377–401, Jul. 2005.

[112] W. Penfield, "The interpretive cortex: The stream of consciousness in the human brain can be electrically reactivated," *Science*, vol. 129, no. 3365, pp. 1719–1725, Jun. 1959.

[113] 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.

[114] S. Varma and R. Simon, "Bias in error estimation when using cross-validation for model selection," *BMC Bioinformatics*, vol. 7, no. 1, p. 91, 2006.

[115] H. Higashi, A. Cichocki, and T. Tanaka, "Regularization using geometric information between sensors capturing features from brain signals," in *Proceedings of 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 721–724.

[116] W. Samek, V. Carmen, K.-R. Müller, and M. Kawanabe, "Stationary common spatial patterns for brain—computer interfacing," *Journal of Neural Engineering*, vol. 9, no. 2, p. 26013, 2012.

[117] G. Townsend, B. Graimann, and G. Pfurtscheller, "Continuous EEG classification during motor imagery-simulation of an asynchronous BCI," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 12, no. 2, pp. 258–265, 2004.

Appendix A

Proofs

A.1 Proof of Theorem 1

For convenience, we define $\boldsymbol{Q} = \boldsymbol{Q}_1 + \boldsymbol{Q}_2$ and $\boldsymbol{v}_{ij} = \boldsymbol{v}_j$, $j = 1, \ldots, D_i$. Any vector $\boldsymbol{h}_i$ can be normalized such that $\boldsymbol{h}_i^T \boldsymbol{Q} \boldsymbol{h}_i = 1$ and the cost function keeps unchanged. Therefore, the maximization of the cost function is equivalent to the maximization of $\boldsymbol{h}_i^T \boldsymbol{Q}_c \boldsymbol{h}_i$ with the constraint that $\boldsymbol{h}_i^T \boldsymbol{Q} \boldsymbol{h}_i = 1$. Then the Lagrangian of the cost function is

$$L = \boldsymbol{h}_i^T \boldsymbol{Q}_c \boldsymbol{h}_i - \zeta(\boldsymbol{h}_i \boldsymbol{Q} \boldsymbol{h}_i - 1) - \sum_{j=1}^{D_i} \nu_j \boldsymbol{h}_i^T \boldsymbol{v}_j, \tag{A.1}$$

where ζ and $\nu_1, \ldots, \nu_{D_i}$ are Lagrange multipliers. The partial derivative of L with respect to $\boldsymbol{h}_i$ is

$$\frac{\partial L}{\partial \boldsymbol{h}_i} = 2\boldsymbol{Q}_c \boldsymbol{h}_i - 2\zeta \boldsymbol{Q} \boldsymbol{h}_i - \sum_{j=1}^{D_i} \nu_j \boldsymbol{v}_j. \tag{A.2}$$

Then, $\partial L / \partial \boldsymbol{h}_i$ is zero when

$$\zeta = \frac{\boldsymbol{h}_i^T \boldsymbol{Q}_c \boldsymbol{h}_i}{\boldsymbol{h}_i^T \boldsymbol{Q} \boldsymbol{h}_i}, \tag{A.3}$$

because by left-multiplying $\boldsymbol{h}_i^T$ in $\partial L/\partial \boldsymbol{h}_i = 0$, we obtain

$$2\boldsymbol{h}_i^T \boldsymbol{Q}\boldsymbol{h}_i - 2\zeta \boldsymbol{h}_i^T \boldsymbol{Q}_c \boldsymbol{h}_i - \sum_{j=1}^{D_i} \nu_j \boldsymbol{h}_i^T \boldsymbol{v}_j = 0, \tag{A.4}$$

where the third term in the left-hand is zero because $\boldsymbol{h}_i^T \boldsymbol{v}_j = 0$ for $i \neq j$. We define $\boldsymbol{V}$ by (4.15) and $\boldsymbol{\nu}$ as

$$\boldsymbol{\nu} = [\nu_1, \nu_2, \ldots, \nu_{D_i}]^T \tag{A.5}$$

and rewrite A.4 to

$$2\boldsymbol{h}_i^T \boldsymbol{Q}\boldsymbol{h}_i - 2\zeta \boldsymbol{h}_i^T \boldsymbol{Q}_c \boldsymbol{h}_i - \boldsymbol{V}\boldsymbol{\nu} = 0. \tag{A.6}$$

Then, multiplying $\partial L/\partial \boldsymbol{h}_i = 0$ by $\boldsymbol{V}^T \boldsymbol{Q}^{-1}$ is

$$2\boldsymbol{V}^T \boldsymbol{Q}^{-1} \boldsymbol{Q}_c \boldsymbol{h}_i - \boldsymbol{V}^T \boldsymbol{Q}^{-1} \boldsymbol{V}\boldsymbol{\nu} = 0, \tag{A.7}$$

because

$$2\zeta \boldsymbol{V}^T \boldsymbol{Q}^{-1} \boldsymbol{Q}\boldsymbol{h}_i = 0. \tag{A.8}$$

Thus,

$$\boldsymbol{\nu} = 2(\boldsymbol{V}^T \boldsymbol{Q}^{-1} \boldsymbol{V})^{-1} \boldsymbol{V}^T \boldsymbol{Q}^{-1} \boldsymbol{Q}_c \boldsymbol{h}_i. \tag{A.9}$$

By substituting (A.9) into $\partial L/\partial \boldsymbol{h}_i = 0$, we obtain

$$\begin{aligned} 2\boldsymbol{Q}_c \boldsymbol{h}_i - 2\zeta \boldsymbol{Q}_c \boldsymbol{h}_i - 2\boldsymbol{V}(\boldsymbol{V}^T \boldsymbol{Q}^{-1} \boldsymbol{V})^{-1} \boldsymbol{V}^T \boldsymbol{Q}^{-1} \boldsymbol{Q}_c \boldsymbol{h}_i = 0 \\ \boldsymbol{G}\boldsymbol{Q}_c \boldsymbol{h}_i = \zeta \boldsymbol{Q}\boldsymbol{h}_i, \end{aligned} \tag{A.10}$$

where $\boldsymbol{G}$ is defined in (4.14). Since ζ is the criterion to be maximized, the maximum solution of the optimization problem is achieved by the unit-length generalized eigenvector corresponding to the largest generalized eigenvalue of (4.12).

A.2 Proof of Proposition 1

We first consider an upper bound of $\hat{J}(\boldsymbol{w}, \boldsymbol{h})$. For convenience, we define

$$P_d = E_{\boldsymbol{X} \in \mathcal{C}_d}[\alpha_{\boldsymbol{X}}(\boldsymbol{w}, \boldsymbol{h}, \boldsymbol{b})], \tag{A.11}$$

for $d = 1, 2$. Therefore, we can represent $\hat{J}(\mathcal{P})$ as $\hat{J}(\mathcal{P}) = P_c/(P_1 + P_2)$, where $\mathcal{P} = \{\boldsymbol{w}, \boldsymbol{h}, \boldsymbol{b}\}$. Then,

$$\hat{J}(\mathcal{P}) = \frac{P_c}{P_1 + P_2} \leq 1, \tag{A.12}$$

because,

$$P_1, P_2 < P_1 + P_2. \tag{A.13}$$

It is the upper bound of the cost function, $\hat{J}(\mathcal{P})$.

Note that

$$\hat{J}(\boldsymbol{w}^{(k)}, \boldsymbol{h}^{(k)}, \boldsymbol{b}^{(k)}) = \hat{J}(\mathcal{P}^{(k)}) > 0. \tag{A.14}$$

Given $\boldsymbol{w}^{(k+1)}$ by solving (4.10) with fixed $\boldsymbol{h}^{(k)}$ and $\boldsymbol{b}^{(k)}$. Then the relationship;

$$0 \leq \hat{J}(\boldsymbol{w}^{(k)} | \boldsymbol{h}^{(k)}, \boldsymbol{b}^{(k)}) \leq \hat{J}(\boldsymbol{w}^{(k+1)} | \boldsymbol{h}^{(k)}, \boldsymbol{b}^{(k)}) \leq 1, \tag{A.15}$$

is given, because $\hat{J}(\boldsymbol{w} | \boldsymbol{h}^{(k)}, \boldsymbol{b}^{(k)})$ with the constraint and fixing $\boldsymbol{h}^{(k)}$, $\boldsymbol{b}^{(k)}$ achieves its maximum by $\boldsymbol{w}^{(k+1)}$. Moreover the relationship yields

$$0 \leq \hat{J}(\boldsymbol{w}^{(k)}, \boldsymbol{h}^{(k)}, \boldsymbol{b}^{(k)}) \leq \hat{J}(\boldsymbol{w}^{(k+1)}, \boldsymbol{h}^{(k)}, \boldsymbol{b}^{(k)}) \leq 1. \tag{A.16}$$

In a similar way, it holds that;

$$0 \leq \hat{J}(\boldsymbol{w}^{(k)}, \boldsymbol{h}^{(k)}, \boldsymbol{b}^{(k)}) \leq \hat{J}(\boldsymbol{w}^{(k)}, \boldsymbol{h}^{(k+1)}, \boldsymbol{b}^{(k)}) \leq 1, \tag{A.17}$$

and

$$0 \leq \hat{J}(\boldsymbol{w}^{(k)}, \boldsymbol{h}^{(k)}, \boldsymbol{b}^{(k)}) \leq \hat{J}(\boldsymbol{w}^{(k)}, \boldsymbol{h}^{(k)}, \boldsymbol{b}^{(k+1)}) \leq 1, \tag{A.18}$$

This completes the proof.

Appendix B

List of Publications

International Journal Papers

1. H. Higashi and T. Tanaka, "Common spatio-time-frequency patterns for motor-imagery based brain machine interfaces," *Computational Intelligence and Neuroscience*, vol. 2013, Article ID 537218, 12 pages, 2013.
   DOI: 10.1155/2013/537218

2. Y. Kimura, T. Tanaka, H. Higashi, and N. Morikawa, "SSVEP-Based brain–Computer interfaces using FSK-modulated visual stimuli," *IEEE Transactions on Biomedical Engineering*, 2013. (accepted for publication)
   DOI: 10.1109/TBME.2013.2265260

3. 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.
   DOI: 10.1109/TBME.2012.2215960

International Conference Papers

1. H. Higashi and T. Tanaka, "Band selection by criterion of common spatial patterns for motor imagery based brain machine interfaces," in *Proceedings of The 2013 Asia-Pacific Signal and Information Prpcessing Association Annual Summit and Conference (APSIPA ASC 2011)*, pp. 1–7, 2013.

2. H. Higashi and T. Tanaka, "Regularization using similarities of signals observed in nearby sensors for feature extraction of brain signals," in *Proceedings of 35th annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2013)*, pp. 7420–7423, 2013.

3. N. Tomida, H. Higashi and T. Tanaka, "A joint tensor diagonalization approach to active data selection for EEG classification," in *Proceedings of 2013 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2013)*, pp. 983–987, 2013.

4. S. Yoshimoto, Y. Washizawa, T. Tanaka, H. Higashi and J. Tamura, "Toward multi-command auditory brain computer interfacing using speech stimuli," in *Proceedings of APSIPA Annual Summit and Conference 2012 (APSIPA 2012)*, pp. 1–4, 2012.

5. H. Higashi and T. Tanaka, "Time sparsification of EEG signals in motor-imagery based brain computer interfaces," in *Proceedings of 34th annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2012)*, pp. 4271–4274, 2012.
   DOI: 10.1109/EMBC.2012.6346910

6. C. Zhang, Y. Kimura, H. Higashi, and T. Tanaka, "A simple plat form of

brain-contralled mobile robot and its implementation by SSVEP," in *Proceedings of 2012 International Joint Conference on Neural Networks (IJCNN 2012)*, pp. 1–7, 2012.
DOI: 10.1109/IJCNN.2012.6252579

7. H. Higashi, A. Cichocki, and T. Tanaka, "Regularization using geometric information between sensors capturing features from brain signals," in *Proceedings of 2012 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2012)*, pp. 721–724, 2012.
DOI: 10.1109/ICASSP.2012.6287985

8. L. Zhang, C. Zhang, H. Higashi, J. Cao, and T. Tanaka, "Common spatial pattern using multivariate EMD for EEG classification," in *Proceedings of The 2011 Asia-Pacific Signal and Information of Association Annual Summit and Conference (APSIPA 2011),* Wed-PM. SS4 (5 pages), 2011.

9. H. Higashi and T. Tanaka, "Optimal design of a bank of spatio-temporal filters for EEG signal classification," in *Proceedings of 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2011)*, pp. 6100–6103, 2011.
DOI: 10.1109/IEMBS.2012.6091507

10. H. Higashi, T. Rutkowski, Y. Yashizawa, A. Cichocki, and T. Tanaka, "EEG auditory steady state responses classification for the novel BCI," in *Proceedings of 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2011)*, pp. 4576–4579, 2011.
DOI: 10.1109/IEMBS.2011.6091133

11. 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 (ICASSP 2011)*, pp. 585–588, 2011.
DOI: 10.1109/ICASSP.2011.5946471

12. 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 (Proceedings of 17th International Conference on Neural Information Processing (ICONIP 2010)*, vol. 6444, pp. 26–33, 2010.
DOI: 10.1007/978-3-642-17534-3_4

13. 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 (IJCNN 2010)*, pp. 3508–3513, 2010.
DOI: 10.1109/IJCNN.2010.5596476

14. T. Tanaka, H. Higashi, and Y. Saito, "Rhythmic component extraction for EEG signals with reduced computational complexity," in *Proceedings of 2009 IEEE International Symposium on Biomedical Engineering (ISBME 2009)*, no. 1054, 2009.

15. 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 (Proceedings of The 2nd International Conference on Cognitive Neurodynamics (ICCN 2009)*, pp. 686–688, 2010.

DOI: 10.1007/978-3-642-17534-3_4

16. 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 Engineering in Medicine and Biology Society (EMBC 2009)*, pp. 2482–2485, 2009.
DOI: 10.1109/IEMBS.2009.5334806

17. Y. Saito, T. Tanaka, and H. Higashi, “Adaptive rhythmic component extraction with regularization for EEG data analysis,” in *Proceedings of 2009 International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2009)*, pp. 353–356, 2009.
DOI: 10.1109/ICASSP.2009.4959593

Domestic Conference Papers

1. 冨田尚規, 東広志, 田中聡久, “テンソル同時対角化によるサンプル重み付けを用いた脳波識別,” 信学技報, vol. 112, no. 423, SIP2012-107, pp. 153–158, 2013.

2. 東広志, 田中聡久, “脳波識別のためのフィルタバンク・空間重み・時間窓の最適設計,” 第 *27* 回電子情報通信学会信号処理シンポジウム論文集 , pp. 235–240, 2012.

3. 東広志, 田中聡久, “運動想起中の脳波識別における時間窓のスパース化,” 信学技報, vol. 112, no. 115, SIP2012-34, pp. 7–12, 2012.

4. 田村潤, 鷲沢嘉一, 東広志, 森川直樹, 田中聡久, “音声刺激による聴覚ブレイン・コンピュータ・インターフェイスの可能性,” 信学技報, vol. 111, no. 466, SIP2011-177, pp. 281–286, 2012.

5. 木村陽介, 東広志, 田中聡久, “FSK 変調した視覚刺激による BCI,” 信学技報, vol. 111, no. 315, NC2011-80, pp. 47–52, 2011 年.

6. 張誠, 木村陽介, 東広志, 田中聡久, “An SSVEP-Based BCI to Control a Mobile Robot via the Internet,” 信学技報, vol. 111, no. 315, NC2011-79, pp. 41–46, 2011 年.

7. 張誠, 木村陽介, 東広志, 田中聡久, “Control of a Mobile Robot via Wi-Fi with an SSVEP-Based BCI,” 第 26 回電子情報通信学会信号処理シンポジウム, pp. 602–607, 2011 年.

8. 東広志, Andrzej Cichocki, 田中聡久, “脳波電極間の距離情報を用いた正則化,” 第 26 回電子情報通信学会信号処理シンポジウム, pp. 497–502, 2011 年.

9. 東広志, 田中聡久, “脳波識別のための時空間フィルタバンクの最適設計,” 信学技報, vol. 111, no. 102, SIP2011-44, pp. 85–90, 2011 年.

10. 木村陽介, 東広志, 田中聡久, “定常的視覚誘発電位を用いたディジタル通信,” 2011 年 電子情報通信学会総合大会 基礎・境界講演論文集, A-4-27, pp. 105, 2011 年.

11. 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,” 2011 年 電子情報通信学会総合大会 基礎･境界講演論文集, A-4-34, pp. 112, 2011 年.

12. 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.

13. 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,” 第 25 回電子情報通信学会信号処理シンポジウム, pp. 121–126, 2010 年.

14. 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,” 第 24 回電子情報通信学会信号処理シンポジウム, pp. 402–407, 2009 年.

15. 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, vol. 109, No. 280, NC2009-54, pp. 23–28, 2009 年.

16. 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, vol. 109, No. 112, SIP2009-27, pp. 49–53, 2009 年.

17. 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,” 第 23 回電子情報通信学会信号処理シンポジウム, pp. 58–63, 2008 年.