Smoothing of Spatial Filter by Graph Fourier Transform for EEG Signals

Hiroshi Higashi*, Toshihisa Tanaka†, and Yuichi Tanaka†
* Toyohashi University of Technology, Aichi, Japan.
E-mail: higashi@tut.jp Tel: +81-532-44-6779
† Tokyo University of Agriculture and Technology, Tokyo, Japan.
E-mail: tanakat@cc.tuat.ac.jp, ytnk@cc.tuat.ac.jp Tel: +81-42-388-7123, +81-42-388-7150

***Abstract*—Spatial filtering is useful for extracting features from multichannel EEG signals. In order to enhance robustness of the spatial filter against low SNR and small samples, we propose a smoothing method for the spatial filter using spectral graph theory. This method is based on an assumption that the electrodes installed in nearby locations observe the electrical activities of the same source. Therefore the spatial filter's coefficients corresponding to the nearby electrodes are supposed to be taken similar values, that is, the coefficients should be spatially smooth. To introduce the smoothness, we define a graph whose edge weights represent the physical distances between the electrodes. The spatial filter spatially smoothed is found out in the subspace that is spanned by the smooth basis of the graph Fourier transform. We evaluate the method with artificial signals and a dataset of motor imagery brain computer interface. The smoothness of the spatial filter given by the method provides robustness of the spatial filter in the condition that the small amount of the samples is available.**

I. Introduction

Electroencephalogram (EEG) is a record electrical activities of a brain. Because analyzing spatial features of the activities is important in several situations, a multichannel measurement of EEG where electrodes are installed in various locations on a head surface has been widely used. One of the effective methods to analyze the multichannel EEG signals is spatial filtering [1], which is used for improving the signal to noise ratio (SNR) of observed signals or separating some activities from the observations [2]. The spatial filtering is formulated by a weighted-sum of the signals observed at electrodes [3], [4]. The spatially filtered signal is given as;

$$y(t) = \sum_{i=1}^{M} w_i x_i(t), \tag{1}$$

where $x_i(t)$ denotes a signal observed at the ith electrode at time t, M denotes the number of the electrodes, w_i denotes a spatial weight for the ith electrode, and $y(t)$ denotes an extracted signal. The problem to extract certain components from observations of the sensor array is to find the weight vector denoted by $\boldsymbol{w} = [w_1, \ldots, w_M]^\top$ under a certain criterion, where $\cdot^\top$ is the transpose of a vector or a matrix. For this purpose, learning approaches using observed signals are widely adopted [5], [6]. A number of signal processing techniques such as Wiener filtering, principal component analysis (PCA), independent component analysis (ICA) [4], and so forth are involved in this problem.

However, the learning procedure can be ill-posed because of a limited number of sensors and samples. Hence regularization is widely used to prevent to solve an ill-posed problem or overfitting in signal processing and machine learning [7], [8]. The regularization for an optimization problem is to add a penalty term, which represents additional information such as smoothness or bounds of the vector norm of parameters to be optimized, to an original cost function. In this way, the regularization can help to design more robust spatial weights against ill-posed problems [9].

Recently, the regularizations making the optimized spatial filter spatially smooth have been proposed for EEG signal processing [9]–[11]. The regularizations are motivated by the following 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. The EEG signal reflects the summation of the synchronous activity of thousands or millions of neurons [12]–[14]. Hence, the nearby sensors may observe the activities which are induced from the same source. This idea also motivates the popular signal preprocessing methods for EEG signals such as the Laplacian filter that averages the signals observed in the nearby sensors and improves SNR in EEG signal processing [2], [15]. The researches of the regularizations based on the assumption have suggested that the smoothing with the spatial filter improves its robustness [9]–[11]. However, it is difficult to choose the regularization parameter because the parameter should be chosen out of all real values.

In this paper, we propose a novel method giving the spatial smoothness for the spatial filter optimization by a spectral graph based method. The proposed method constrains the variable space for the spatial filter. The constrained variable space is designed by the graph Fourier transform (GFT) [16], [17]. The GFT is a transform for a graph signal that consists of the signal values and the signal structure represented by

H. Higashi and T. Tanaka are also affiliated with RIKEN Brain Science Institute, Saitama, Japan.

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

Y. Tanaka is supported in part by MEXT Tenure Track Promotion Program.

the graph. The GFT can make the smooth basis for the target graph [18], [19]. By utilizing the feature of GFT, we make the graph representing the electrode arrangement and design the spatially smoothing basis functions, which gradually change across the electrode positions, for the variable space. By designing the spatial filter as a linear combination of the basis functions, the proposed method can design the spatially smoothed spatial filter. We show the solution of the spatial filter for an optimization problem formulated by a generalized Rayleigh quotient with the constraint of the search space. In this formulation, the tuning of the regularization parameter becomes simple, because we do not choose the parameter out of all real values but choose the number of the basis function out of the limited number of candidates. The maximum number of the basis function (the number of the parameter candidates) is the same as the number of the vertexes of the graph, that is, the number of the electrodes. We show some efficiencies of the proposed method with PCA and common spatial pattern (CSP) [6] in experiments with artificial signals and BCI datasets, respectively.

II. Spatially Smoothing by Graph Fourier Transform

A. EEG Signals on Graph

The electrodes for EEG recording have a geometric structure that is the locations or arrangement of the electrodes on the head surface. By describing the geometric structure by a graph, EEG signals and its spatial filter can be treated as the graph signal.

The EEG signals connected by the graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \boldsymbol{W}\}$, where $\mathcal{V}$ is a finite set of vertices (electrodes), $\mathcal{E}$ is a set of edges, and $\boldsymbol{W}$ is a weighted adjacency matrix [17]. The entry of the ith column and the jth row of $\boldsymbol{W}$ represents the weight of the edge connecting the ith and jth vertices.

In this paper, the weight matrix $\boldsymbol{W}$ is determined based on the distance between the electrodes by the following way. First, we define a distance between electrodes on the arrangements used for the EEG measurement. International 10-20, 10-10, and 10-5 methods [12], [20], [21] have stood as the de facto standard of the electrode arrangements. 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 locations on head surface can be described by coordinates denoted by $\xi = \{x, y, z\}$. We define the orthogonal coordinates described in Fig. 1, which illustrates the electrode positions of the international 10-20 method.

Given the positions of two electrodes as $\xi_i = \{x_i, y_i, z_i\}$ and $\xi_j = \{x_j, y_j, z_j\}$. The question arising here is: how to define the distance between two points on the head. The Euclidean distance defined by $d_{ij} = \|\xi_i - \xi_j\|$ is a straightforward solution. The distance is the same as the length of the line segment connecting two electrodes. The example of the line segment between $F3$ and Cz electrodes is illustrated in Fig. 1.

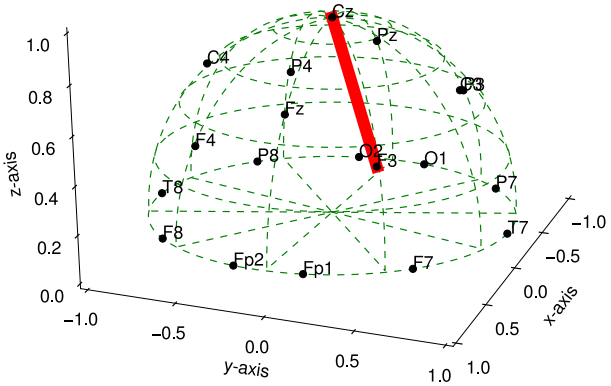


Fig. 1. The electrode arrangement of the international 10-20 method on the orthogonal coordinates.

Consider a multichannel system consisting of M electrodes. We obtain the distances between each electrode as $\{d_{ij}\}_{i,j=1}^{M}$. To connect the vertices corresponding to the nearby electrodes, the Gaussian metric between two points;

$$[\boldsymbol{W}]_{ij} = \begin{cases} \exp\left(-\frac{d_{ij}^2}{2q^2}\right), & d_{ij} \leq \kappa \\ 0, & \text{otherwise} \end{cases}, \quad \forall i, j = 1, \ldots, M \tag{2}$$

is employed, where q denotes a parameter to tune the closeness of the two electrodes and κ is a threshold.

The connections between electrodes are described by (2), then we can define the graph Laplacian [17] defined as

$$\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}, \tag{3}$$

where $\boldsymbol{D}$ is a diagonal matrix whose diagonal elements are defined as $[\boldsymbol{D}]_{ii} = \sum_{i=1}^{M} [\boldsymbol{W}]_{ij}$. A basis of GFT on $\mathcal{G}$ is obtained as a set of the eigenvectors of $\boldsymbol{L}$ [17]. The orthonormal basis $\{\boldsymbol{u}_i\}_{i=0}^{M-1}$ satisfies

$$\boldsymbol{L}\boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i, \quad \forall i = 0, \ldots, M-1. \tag{4}$$

The indexes of the eigenvalues and the eigenvectors are decided in such a way that $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{M-1}$. The orthonormal basis of GFT have similar features to those of the discrete Fourier transform. The graph Laplacian eigenvector $\boldsymbol{u}_0$ is constant, and ones corresponding to the lower eigenvalues vary slowly across the graph.

B. Design of Spatial Filter on a Subspace

In order to give the smoothness to the optimizing spatial filter, we introduce a constraint term that restricts the space in which the spatial filter is found. An optimization problem with the constraint is defined as

$$\begin{aligned} \max_{\boldsymbol{w}} \quad & J(\boldsymbol{w}), \\ \text{subject to} \quad & \boldsymbol{w} \in \mathcal{U}_{\mathcal{P}} = \text{Span}\{\boldsymbol{u}_0, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_{P-1}\}, \end{aligned} \tag{5}$$

where $J(\boldsymbol{w})$ is a cost function of the spatial filter $\boldsymbol{w}$ and $\text{Span}\{\cdots\}$ is a subspace spanned by vectors. The constraint means that $\boldsymbol{w}$ can be represented by a linear combination of

the basis function in $\mathcal{U}_P$. We next show the solution of the cost function formulated as a generalized Rayleigh quotient.

The criteria based on a generalized Rayleigh quotient are well known in problems finding projections [22]. For example, PCA, linear discriminant analysis (LDA) [8], and CSP [6] are formulated with this criterion. The optimization problem with the constraint of the subspace induced from the GFT is represented by

$$\begin{aligned} \max_{\boldsymbol{w}} \quad & \frac{\boldsymbol{w}^\top \boldsymbol{A}\boldsymbol{w}}{\boldsymbol{w}^\top \boldsymbol{B}\boldsymbol{w}}, \\ \text{subject to} \quad & \boldsymbol{w} \in \mathcal{U}_P, \end{aligned} \tag{6}$$

where $\boldsymbol{A}$ and $\boldsymbol{B}$ are symmetric matrices. The set $\{\boldsymbol{u}_0, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_{P-1}\}$ can be the orthogonal basis of $\mathcal{U}_P$ because they can be chosen to be orthogonal each other. Therefore, $\boldsymbol{w}$ can be represented as

$$\boldsymbol{w} = v_0\boldsymbol{u}_0 + v_1\boldsymbol{u}_1 + \cdots + v_{P-1}\boldsymbol{u}_{P-1} = \boldsymbol{U}_p\boldsymbol{v}, \tag{7}$$

where $\{v_i\}_{i=0}^{P-1}$ are arbitrary scalar coefficients, $\boldsymbol{U}_P$ is a matrix defined as $\boldsymbol{U}_P = [\boldsymbol{u}_0, \ldots, \boldsymbol{u}_{P-1}]$, and $\boldsymbol{v}$ is a vector defined as $\boldsymbol{v} = [v_0, \ldots, v_{P-1}]^\top$. Then, the problem (6) can be translated to a problem finding $\boldsymbol{v}$ by change of variables. The problem for finding $\boldsymbol{v}$ is formulated by

$$\max_{\boldsymbol{v}} \quad \frac{\boldsymbol{v}^\top \hat{\boldsymbol{A}}\boldsymbol{v}}{\boldsymbol{v}^\top \hat{\boldsymbol{B}}\boldsymbol{v}}, \tag{8}$$

where $\hat{\boldsymbol{A}}$ and $\hat{\boldsymbol{B}}$ are the matrices defined as

$$\hat{\boldsymbol{A}} = \boldsymbol{U}_P^\top \boldsymbol{A}\boldsymbol{U}_P, \quad \hat{\boldsymbol{B}} = \boldsymbol{U}_P^\top \boldsymbol{B}\boldsymbol{U}_P. \tag{9}$$

The optimal solution $\hat{\boldsymbol{v}}$ of (8) is given as the generalized eigenvector corresponding to the largest generalized eigenvalue of the generalized eigenvalue problem:

$$\hat{\boldsymbol{A}}\boldsymbol{v} = \lambda\hat{\boldsymbol{B}}\boldsymbol{v}. \tag{10}$$

Then, the optimal $\boldsymbol{w}$ is obtained by (7).

III. Experiment with Artificial Signals

An analysis of the proposed method by a toy experiment with artificial signals is given. We assume the feature being distributed spatially and smoothly. Observations are generated as the mixture of the synthetic feature component and noise. We show the result of the principal component estimated with the proposed method of the observations.

A. Generating Artificial Signal

The artificial signal is generated as follows. The feature components;

$$[\boldsymbol{s}]_j = \exp\left(-\frac{\|\boldsymbol{\epsilon}_j - \boldsymbol{o}_s\|^2}{2\sigma^2}\right), \quad j = 1, \ldots, M, \tag{11}$$

are introduced for each channel, where $\boldsymbol{\epsilon}_j$ is the coordinates of the jth electrode, σ^2 is a parameter controlling the spatial extent of the components, $\boldsymbol{o}_s$ is the coordinates representing the location in which the feature component is supposed to be generated, and M is the number of the channels. The model of observations are defined as

$$[\boldsymbol{x}_i]_j = [\boldsymbol{s}]_j + \eta, \quad i = 1, \ldots, N, \tag{12}$$

where η is a stochastic noise and N is the number of the observed samples.

B. Spatially Smoothing PCA

We generated some samples by (12) and extracted the feature component as the principal component. The principal component $\hat{\boldsymbol{p}}$ is estimated by solving the problem;

$$\begin{aligned} \hat{\boldsymbol{p}} &= \arg\min_{\boldsymbol{p}} \sum_{i=1}^{N} \|\boldsymbol{p}^\top(\boldsymbol{x}_i - \bar{\boldsymbol{x}})\|^2 \\ &= \arg\max_{\boldsymbol{x}} \ \boldsymbol{p}^\top \boldsymbol{\Sigma}\boldsymbol{p}, \end{aligned} \tag{13}$$

where $\bar{\boldsymbol{x}}$ is the sample mean defined as $\bar{\boldsymbol{x}} = (1/N)\sum_{i=1}^{N} \boldsymbol{x}_i$ and $\boldsymbol{\Sigma}$ is a sample covariance matrix defined as

$$\boldsymbol{\Sigma} = \frac{1}{N-1}\sum_{i=1}^{N}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^\top. \tag{14}$$

This problem (13) is equivalent to the problem (6) if $\boldsymbol{B} = \boldsymbol{I}$. The principal component is regarded as the spatial filter in this experimental condition. Therefore, the proposed constraint can be applied and the optimal solution can be found.

C. Results

The conditions of the experiment are as follows. The electrode arrangement is shown in Fig. 2. Although the electrode location is given by 3 dimensional coordinates, in order to represent the arrangement on 2 dimensional plot in Fig. 2, we transform the coordinates into the 2 dimensional coordinates $\hat{\boldsymbol{\epsilon}}_j$ as

$$\hat{\boldsymbol{\epsilon}}_j = \boldsymbol{\mu}_j \frac{\|\boldsymbol{\epsilon}_j - \boldsymbol{o}\|}{\|\boldsymbol{\mu}_j\|}, \tag{15}$$

where $\boldsymbol{\mu}_j$ is the coordinates being removed the value of z-axis defined as $\boldsymbol{\mu}_j = [[\boldsymbol{\epsilon}_j]_1, [\boldsymbol{\epsilon}_j]_2]^\top$ and the origin $\boldsymbol{o}$ is defined as $\boldsymbol{o} = [0, 0, 1]^\top$. From now on, we use the same metric for 2 dimensional topographical plot. The source location $\boldsymbol{o}_s$ is $[0.0, -0.7, 0.7]^\top$ and the variance σ^2 is 0.4. The feature components illustrated in Figs. 3a and 3b. The noise η is generated from a Gaussian distribution with zero-mean and unit variance. As an example, we show the observations in Fig. 4.

The results of the standard PCA are shown in Fig. 5. We can observe that the principal component estimated with 100 samples ($N_L = 100$) can represent the desired feature component, because the coefficients corresponding to the electrodes that observe the large feature components are large. On the other hand, the principal component estimated with 5 samples ($N_L = 5$) is influenced strongly by noise and the large coefficients scatter spatially at the various electrodes.

With the same procedure for PCA, we applied PCA with the proposed constraint to the artificial signals. The basis of GFT of the graph describing the electrode arrangement shown

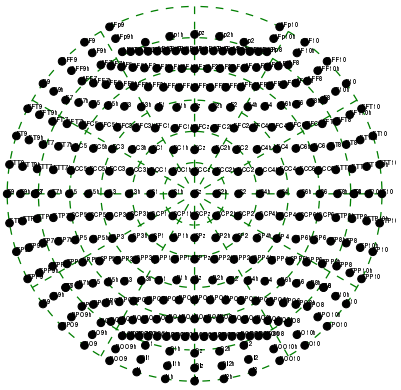
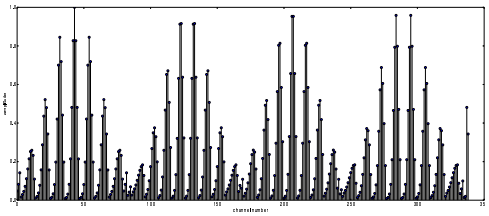
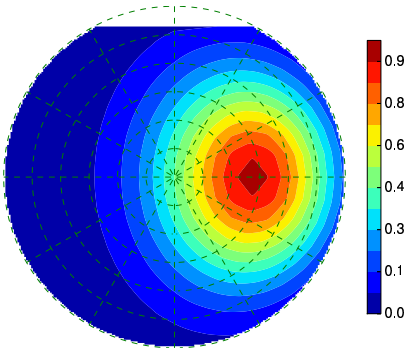


Fig. 2. The electrode arrangement for the artificial signal.



(a) Value for each channel



(b) Topographical plot

Fig. 3. The feature component.

in Fig. 2 is calculated by the way described in Sec. II-A with $\kappa = 0.5$ and $q = 0.1$. The examples of the graph are shown in Fig. 6 and the examples of the basis functions of GFT are shown in Figs. 7 and 8. The principal components of the proposed method are shown in Figs. 9 and 10. In those figures, the principal components $\boldsymbol{w}_p$ is the principal components found in the space of $\{\boldsymbol{x} \in \mathbb{R}^M | \boldsymbol{x} = v_0\boldsymbol{u}_0 + \cdots + v_{p-1}\boldsymbol{u}_{p-1}, v_0 \in \mathbb{R}, \ldots, v_{p-1} \in \mathbb{R}\}$. It should be noted that the principal component found in the 338 dimensional subspace is the same as that by the standard PCA. It is observed in Fig. 10 that the proposed method can find the principal components which are similar to the true feature component even with the small amount of the samples.

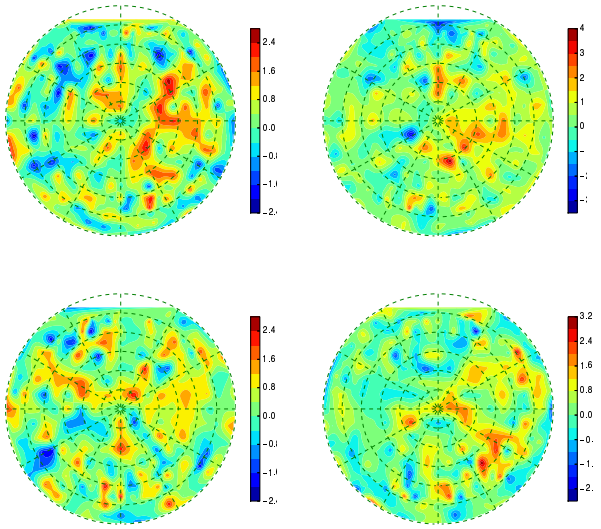
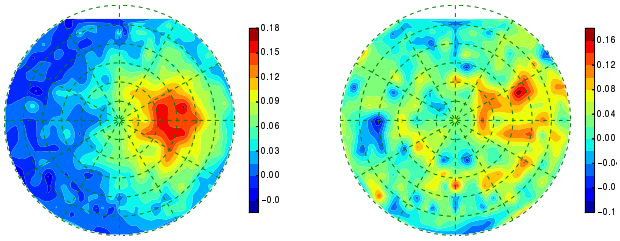


Fig. 4. Four examples of the observations.



(a) $N_L = 100$ (b) $N_L = 5$

Fig. 5. Principal components by the standard PCA.

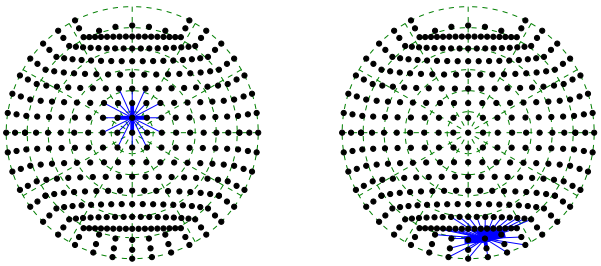
IV. Experiment with BCI Dataset

We evaluated the proposed method in a two-class classification problem in which EEG signals are classified into motor imagery tasks. We applied the proposed constraint in the CSP method that designs the spatial filters. We compared the classification performance of the CSP method with the proposed constraint to the standard CSP and the spatially regularized CSP (SRCSP) [9].

A. Data Description

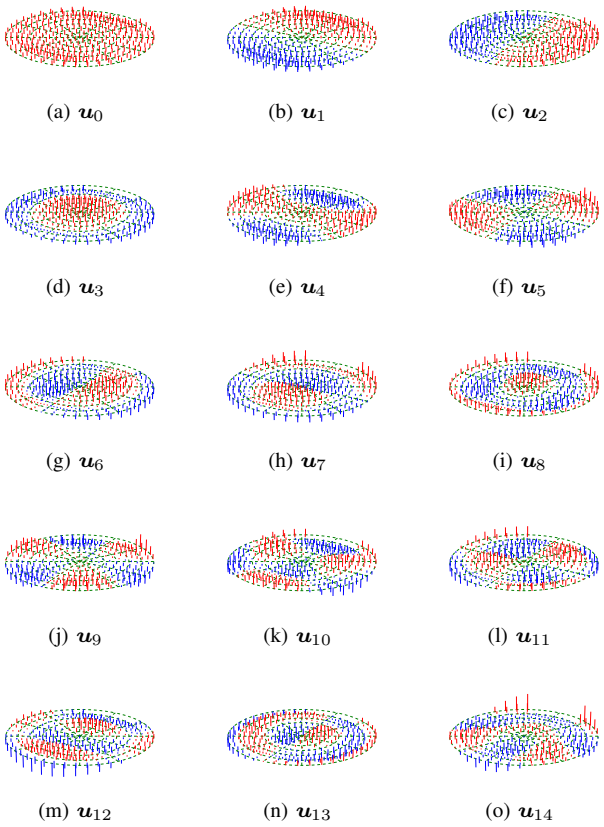
We used dataset IVa from BCI Competition III (for details of the dataset, see http://www.bbci.de/competition/iii/). This dataset consists of EEG signals during right hand and right foot motor-imageries. The EEG signals were recorded from five subjects labeled *aa*, *al*, *av*, *aw*, and *ay*. The electrodes were installed with the electrode arrangement called the extended International 10-20 method [23] (see Fig. 11). The measured signal was bandpass filtered with the passband of 0.05–200 Hz, and then digitized at 1000 Hz.

Moreover, the lowpass filter whose cutoff frequency is 50 Hz was applied to recorded signals and the filtered signals was donwsampled to 100 Hz. Furthermore, the signals were bandpass filtered with the passband of 7–30 Hz that is a band



(a) *FCCz* (b) *O2h*

Fig. 6. Edges at each electrode for the graph. The width of the edges represents the weights for each edge.



(a) $\boldsymbol{u}_0$ (b) $\boldsymbol{u}_1$ (c) $\boldsymbol{u}_2$

(d) $\boldsymbol{u}_3$ (e) $\boldsymbol{u}_4$ (f) $\boldsymbol{u}_5$

(g) $\boldsymbol{u}_6$ (h) $\boldsymbol{u}_7$ (i) $\boldsymbol{u}_8$

(j) $\boldsymbol{u}_9$ (k) $\boldsymbol{u}_{10}$ (l) $\boldsymbol{u}_{11}$

(m) $\boldsymbol{u}_{12}$ (n) $\boldsymbol{u}_{13}$ (o) $\boldsymbol{u}_{14}$

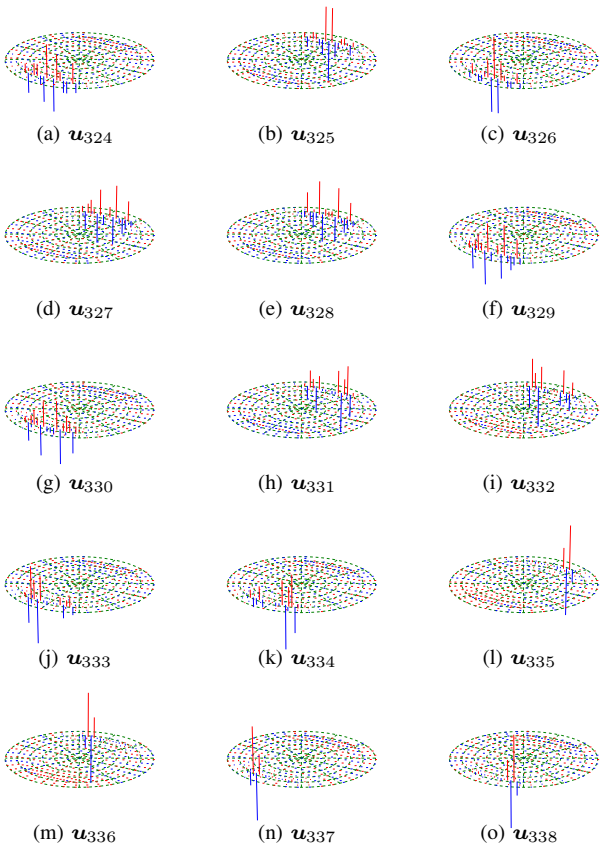
Fig. 7. The basis function of GFT $(\boldsymbol{u}_0, \ldots, \boldsymbol{u}_{14})$.

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.

B. Feature Extraction and Classification

We classified the observed signals into two classes corresponding to the motor-imagery tasks by feature extraction with the spatial filters designed by the CSP method [6], [24] and classification with an LDA classifier [8].

In the classification, the CSP method finds the spatial filters as follows. Let $\boldsymbol{X} \in \mathbb{R}^{M \times N}$ be an observed signal, where M is the number of channels and N is the number of samples and let $\boldsymbol{x}_n$ be the nth column of $\boldsymbol{X}$. The CSP method extracts



(a) $\boldsymbol{u}_{324}$ (b) $\boldsymbol{u}_{325}$ (c) $\boldsymbol{u}_{326}$

(d) $\boldsymbol{u}_{327}$ (e) $\boldsymbol{u}_{328}$ (f) $\boldsymbol{u}_{329}$

(g) $\boldsymbol{u}_{330}$ (h) $\boldsymbol{u}_{331}$ (i) $\boldsymbol{u}_{332}$

(j) $\boldsymbol{u}_{333}$ (k) $\boldsymbol{u}_{334}$ (l) $\boldsymbol{u}_{335}$

(m) $\boldsymbol{u}_{336}$ (n) $\boldsymbol{u}_{337}$ (o) $\boldsymbol{u}_{338}$

Fig. 8. The basis function of GFT $(\boldsymbol{u}_{322}, \ldots, \boldsymbol{u}_{338})$.

a feature values with the spatial filter $\boldsymbol{w}$ in such a way that

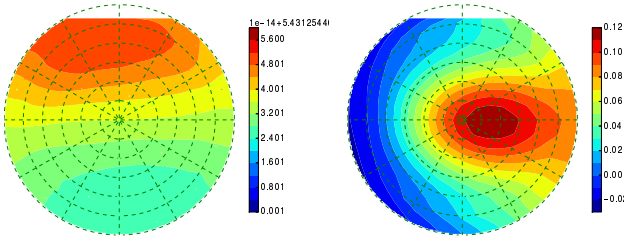
$$\sigma^2(\boldsymbol{X}, \boldsymbol{w}) = \frac{1}{N} \sum_{n=1}^{N} |\boldsymbol{w}^\top (\boldsymbol{x}_n - \boldsymbol{\mu})|^2, \tag{16}$$

where $\boldsymbol{\mu}$ is the time average of $\boldsymbol{X}$ given by $\boldsymbol{\mu} = (1/N) \sum_{n=1}^{N} \boldsymbol{x}_n$. We assume that sets of the observed signals, $\mathcal{C}_1$ and $\mathcal{C}_2$, where $\mathcal{C}_d$ contains the signals belonging to class d, $d \in \{1, 2\}$ is a class label, $\mathcal{C}_1 \cap \mathcal{C}_2 = \emptyset$, and $\emptyset$ is a set having no elements. CSP is defined the weight vector that maximizes the intra-class variance in $\mathcal{C}_c$ under the normalization of samples, where c is a class label. More specifically, for c fixed, the weight vector is found by solving the following optimization problem [5], [6];

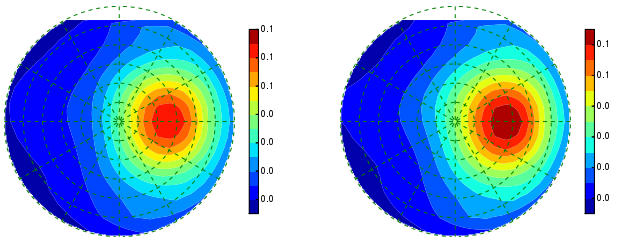
$$\begin{aligned} &\max_{\boldsymbol{w}} \quad E_{\boldsymbol{X} \in \mathcal{C}_c} \left[\frac{1}{N} \sum_{n=1}^{N} |\boldsymbol{w}^\top (\boldsymbol{x}_n - \boldsymbol{\mu})|^2 \right], \\ &\text{subject to} \sum_{d=1,2} E_{\boldsymbol{X} \in \mathcal{C}_d} \left[\frac{1}{N} \sum_{n=1}^{N} |\boldsymbol{w}^\top (\boldsymbol{x}_n - \boldsymbol{\mu})|^2 \right] = 1, \end{aligned} \tag{17}$$

where $E_{\boldsymbol{X} \in \mathcal{C}_d}[\cdot]$ denotes the expectation over $\mathcal{C}_d$ and $|\cdot|$ is the absolute value of a scalar. The solution of (17) is given by the generalized eigenvector corresponding to the largest generalized eigenvalue of the generalized eigenvalue problem described as

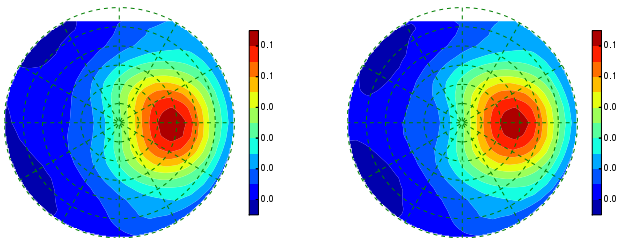
$$\boldsymbol{\Sigma}_c \boldsymbol{w} = \lambda (\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2) \boldsymbol{w}, \tag{18}$$



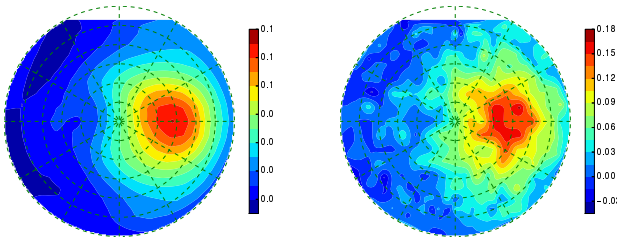
(a) $\boldsymbol{w}_0$ (b) $\boldsymbol{w}_5$



(c) $\boldsymbol{w}_{10}$ (d) $\boldsymbol{w}_{15}$

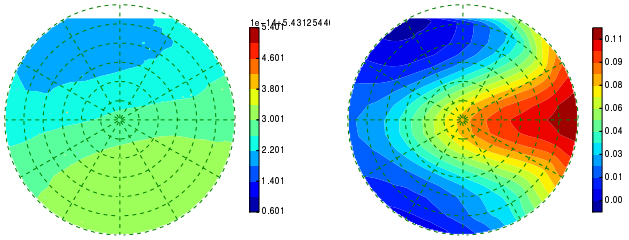


(e) $\boldsymbol{w}_{20}$ (f) $\boldsymbol{w}_{25}$

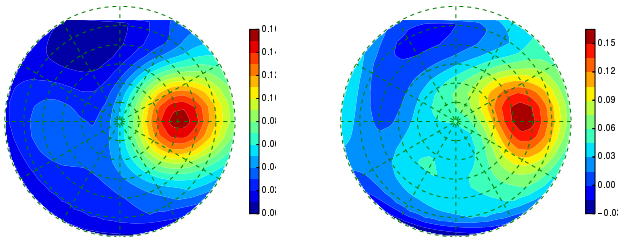


(g) $\boldsymbol{w}_{30}$ (h) $\boldsymbol{w}_{338}$

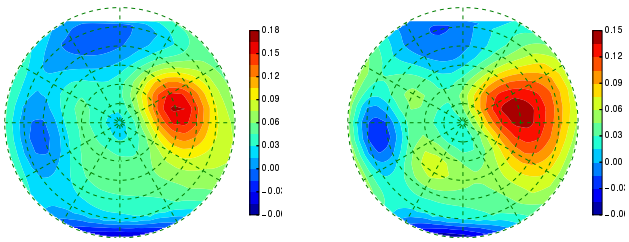
Fig. 9. Principal components by PCA with the proposed constraint ($N_L = 100$).



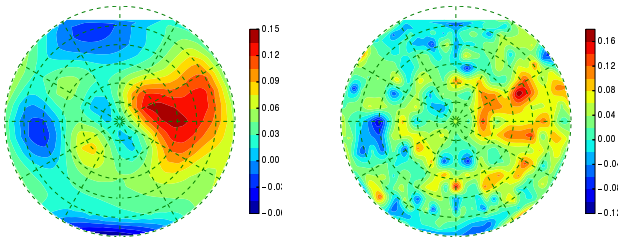
(a) $\boldsymbol{w}_0$ (b) $\boldsymbol{w}_5$



(c) $\boldsymbol{w}_{10}$ (d) $\boldsymbol{w}_{15}$



(e) $\boldsymbol{w}_{20}$ (f) $\boldsymbol{w}_{25}$



(g) $\boldsymbol{w}_{30}$ (h) $\boldsymbol{w}_{338}$

Fig. 10. Principal components by PCA with the proposed constraint ($N_L = 5$).

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

$$\boldsymbol{\Sigma}_d = E_{\boldsymbol{X} \in \mathcal{C}_d} \left[\frac{1}{N} \sum_{n=1}^{N} (\boldsymbol{x}_n - \boldsymbol{\mu})(\boldsymbol{x}_n - \boldsymbol{\mu})^\top \right]. \tag{19}$$

The solution of the problem (17) is equal to the solution of the problem:

$$\max_{\boldsymbol{w}} \quad \frac{\boldsymbol{w}^\top \boldsymbol{\Sigma}_c \boldsymbol{w}}{\boldsymbol{w}^\top (\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2) \boldsymbol{w}}. \tag{20}$$

This formulation with the constraint that $\boldsymbol{w}$ is optimized in the subspace is the same as the problem (8) where $\boldsymbol{A} = \boldsymbol{\Sigma}_c$ and $\boldsymbol{B} = \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2$. Therefore, we can incorporate the graph based smoothing method into the CSP method.

The feature vector was extracted by the following way. In each case of $c = 1$ and $c = 2$, we solve (17) or (20). We got the eigenvectors corresponding to the largest eigenvalues in each eigenvalue problem defined as $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$, respectively. By using the weight vectors, the feature vector was defined as $\boldsymbol{y} = [\sigma^2(\boldsymbol{X}, \boldsymbol{w}_1), \sigma^2(\boldsymbol{X}, \boldsymbol{w}_2)]^\top$.

The feature vector was projected onto the 1-dimensional space by the LDA projector designed with learning data. The projected feature is classified by the threshold which is defined

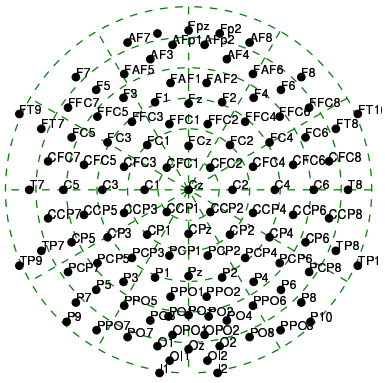


Fig. 11. The electrode arrangement of BCI Competition III dataset IVa.

TABLE I
ACCURACY RATES [%] GIVEN BY 100 TRAINING SAMPLES PER CLASS.

	Subject					
	aa	*al*	*av*	*aw*	*ay*	Ave.
CSP	80.96	97.89	51.93	92.56	82.90	82.90
SRCSP	81.38	97.96	57.35	92.35	90.35	83.95
Proposed	82.13	98.21	71.98	95.20	91.74	87.85

as the middle point of two class averages over the projected learning data.

C. Results

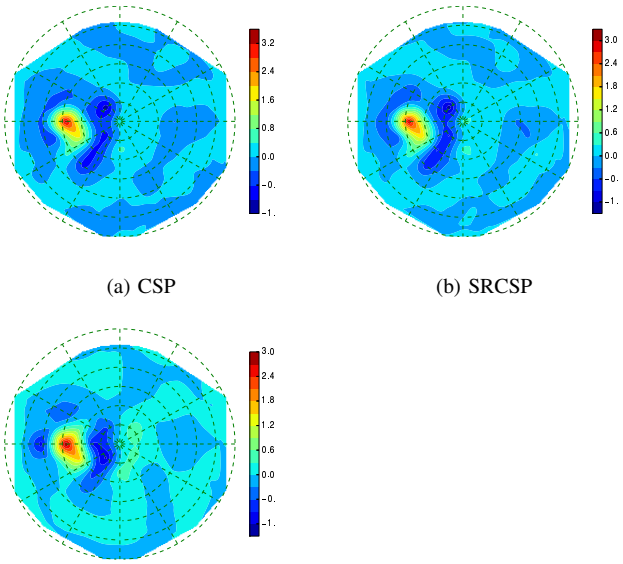
The classification accuracy rates were given by training the spatial filters and the classifier with randomly chosen 100 samples, and testing with the remaining samples. The accuracy rates averaged over 100 times of this procedure are shown in Table I. The parameters in (2) were set to $\kappa = 0.5$ and $q = 0.1$ for SRCSP [9] and the proposed method. Moreover, the regularization parameter for SRCSP was chosen by the nested 5×5 cross validation (CV) out of $\{0, 10^{-7}, 10^{-6.8}, \ldots, 10^{-0.2}, 10^{0}\}$. The dimension P of the subspace for the proposed method was chosen by the nested 5×5 CV out of $\{1, 2, \ldots, 117, 118\}$.

Table II also shows the classification accuracy rates in the case where the number of the training samples is considerably reduced to only five samples. We can observe a significant improvement of the accuracy rates for subjects *al* and *ay* by the proposed constraint. The results suggest that the proposed method can improve the accuracy even if the number of training samples available is small.

Examples of the topographically plotted spatial filters for subject *al* are shown in Figs. 12 and 13. In Fig. 12 ($N_L = 100$), the regularization parameters for SRCSP and P for the proposed method are set to 10^{-3} and 67, respectively. In Fig. 13 ($N_L = 5$) the regularization parameters for SRCSP and P for the proposed method are set to 10^{-3} and 59, respectively. Those parameters are chosen as the parameters performing the best classification accuracy rate in the 5×5 CV. In the case that the enough number of the learning samples is available (Fig. 12), the spatial filters among the CSP, SRCSP,

TABLE II
ACCURACY RATES [%] GIVEN BY 5 TRAINING SAMPLES PER A CLASS.

	Subject					
	aa	*al*	*av*	*aw*	*ay*	Ave.
CSP	52.49	77.62	52.32	66.62	55.66	60.94
SRCSP	50.94	83.19	52.39	67.04	53.34	61.38
Proposed	56.43	91.16	55.39	75.11	69.60	69.60



(a) CSP (b) SRCSP

(c) Proposed

Fig. 12. Spatial filters ($N_L = 100$).

and the proposed method have a common feature of the large coefficients concentrating around the left motor cortex. However, if the number of the available learning samples is small, the spatial filter designed only by the proposed method has such a feature.

V. CONCLUSIONS

We have proposed the method to provide the smoothness for the problem of designing the spatial filter for multichannel EEG signals. Our experiments suggest that compared with the existing methods that give the smoothness by the regularizations, the proposed method provides more robust spatial filter because of the less number of the candidates of the parameter to be tuned.

REFERENCES

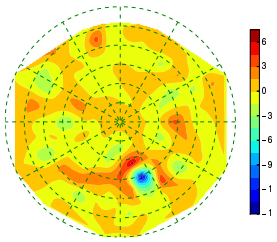
[1] R. Srinivasan, P. L. Nunez, and R. B. Silberstein, "Spatial filtering and neocortical dynamics: Estimates of EEG coherence," *IEEE Transactions on Biomedical Engineering*, vol. 45, no. 7, pp. 814–826, 1998.

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

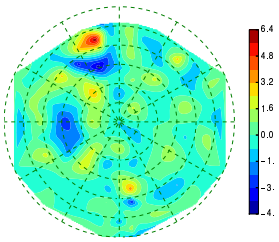
[3] H. Krim and M. Viberg, "Two decades of array signal processing research," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67–94, 1996.

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

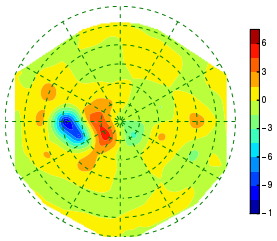
[5] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," *Clinical Neurophysiology*, vol. 110, no. 5, pp. 787–798, 1999.



(a) CSP



(b) SRCSP



(c) Proposed

Fig. 13. Spatial filters ($N_L = 5$).

[6] 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.

[7] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*. V. H. Winston & Sons, 1977.

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

[9] 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.

[10] 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.

[11] H. Higashi and T. Tanaka, "Regularization using similarities of signals observed in nearby sensors for feature extraction of brain signals," in *Proceedings of 2013 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013, pp. 7420–7423.

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

[13] 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.

[14] H. Yuan and B. He, "Brain-computer interfaces using sensorimotor rhythms: Current state and future perspectives," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 5, pp. 1425–1435, 2014.

[15] B. He, S. Gao, H. Yuan, and J. Wolpaw, "Brain—Computer Interfaces," in *Neural Engineering*, B. He, Ed. Springer US, 2013, pp. 87–151.

[16] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 129–150, 2011.

[17] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 83–98, 2013.

[18] G. Taubin, "A signal processing approach to fair surface design," in *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '95)*, 1995, pp. 351–358.

[19] G. Taubin, T. Zhang, and G. Golub, "Optimal surface smoothing as filter design," in *Computer Vision — ECCV '96*, ser. Lecture Notes in Computer Science, B. Buxton and R. Cipolla, Eds. Springer Berlin Heidelberg, 1996, vol. 1064, pp. 283–292.

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

[21] 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.

[22] B. Parlett, *The Symmetric Eigenvalue Problem*. Society for Industrial and Applied Mathematics, Jan. 1998.

[23] 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.

[24] 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.