Multilinear Discriminant Analysis With Subspace Constraints for Single-Trial Classification of Event-Related Potentials

Hiroshi Higashi, *Member, IEEE*, Tomasz M. Rutkowski, *Senior Member, IEEE*, Toshihisa Tanaka, *Senior Member, IEEE*, and Yuichi Tanaka, *Member, IEEE*

***Abstract*—The classification accuracy of a brain–computer interface (BCI) frequently suffers from ill-posed and overfitting problems. To avoid and alleviate these problems, we propose a method of a multilinear discriminant analysis with constraints to augment parameter reduction, regularization, and additional prior information for event-related potential (ERP)-based BCIs. The method reduces the number of parameters by multilinearization, regularizes the ill-posedness via subspaces that constrain the parameter spaces, and incorporates a brain functional connectivity through the constraints. The experimental results show that the proposed method improved the classification accuracy rates in a single-trial ERP processing.**

***Index Terms*—Brain–computer/machine interface (BCI/BMI), event-related potentials, electroencephalogram (EEG), single-trial classification, linear discriminant analysis, multilinear algebra.**

I. Introduction

THE decoding of brain activities from an electroencephalogram (EEG) is an important and challenging technology [1]. One of the applications of EEG decoding is a brain–computer/machine interface (BCI/BMI). The BCI uses neurophysiological signals to directly connects a human brain with an external device. By detecting the brain activities evoked by certain tasks, the BCI generates an output. The brain activities include the imagination of muscular movements or certain other procedures, such as paying attention to external stimuli [2], [3]. The BCI provides a nonmuscular communication tool for conveying mental messages or commands to the external world [3]–[5].

Manuscript received November 1, 2015; revised April 7, 2016 and July 19, 2016; accepted July 30, 2016. Date of publication August 10, 2016; date of current version September 23, 2016. This work was supported by JSPS KAKENHI under Grants 15K21079, 26240043, and 15H04002, and the Tateishi Science and Technology Foundation under Grant 2151030. The work of Y. Tanaka was supported in part by the MEXT Tenure Track Promotion Program. The guest editor coordinating the review of this manuscript and approving it for publication was Dr. Stephen Strother.

H. Higashi is with the Graduate School of Engineering, Toyohashi University of Technology, Aichi 441-8580, Japan, and also with RIKEN Brain Science Institute, Saitama 351-0106, Japan (e-mail: higashi@tut.jp).

T. M. Rutkowski is with the Graduate School of Education, University of Tokyo, Tokyo 113-8654, Japan, and also with RIKEN Brain Science Institute, Saitama 351-0106, Japan (e-mail: tomek@bci-lab.info).

T. Tanaka is with the Institute of Engineering, Tokyo University of Agriculture and Technology, Tokyo 183-0057, Japan, and also with RIKEN Brain Science Institute, Saitama 351-0106, Japan (e-mail: tanakat@cc.tuat.ac.jp).

Y. Tanaka is with the Institute of Engineering, Tokyo University of Agriculture and Technology, Tokyo 183-0057, Japan (e-mail: ytnk@cc.tuat.ac.jp).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTSP.2016.2599297

One of the promising BCI tasks is based on the oddball paradigm [6], where multiple stimuli (symbols or tones) appear randomly [2]. The user responds by counting the number of occurrences of target stimuli or, in psychophysical experiments with able-bodied users, by clicking buttons. During the response, slow fluctuations, called event-related potentials (ERPs) [6], are observed in the EEG. An ERP consists of several electrophysiological components, such as P300 and N100. Detection of these ERP components identifies the stimulus the user is heeding and indicates the command the user wants to enter. A well-known BCI, called a P300-speller, uses the oddball paradigm for the input of the alphabet [7], [8]. The main problem for ERP-based BCIs is the classification of the observed EEG signals as either target (the user pays attention to the stimulus) and nontarget (the user does not pay attention to the stimulus). To solve this, linear discriminant analysis (LDA), as a supervised dimensionality reduction technique, is used widely [8], [9].

Spatial dense, multichannel, systems for recording ERPs would improve the accuracy of classifying BCIs. As the number of electrodes and channels for capturing the brain activities increases, the size of the EEG signal matrices also increases [10]. Because of rearrangement of the EEG signals, the input vectors in the LDA projection may be large. Because brain patterns depend on the individual subject and measurement environment, the projections and classifiers need calibration. However, it can be difficult to obtain a sufficient number of samples (trials) for calibration, because the recording of larger EEG datasets is time-consuming and tires the user. When the signal size is large and the number of the samples is small, the calculations can be ill-posed or overfitted [11].

To alleviate such problems, unsupervised dimensionality reduction techniques, such as principal component analysis (PCA) [11], regularizations [12]–[16], calibrations with data from different users or sessions [10], [17], [18], and the generation of artificial signals [13] are often used. For ERP-based BCIs, regularization using the vector norms [19] or shrunk covariance [9] has been proposed. Additionally, multilinear/multiway signal processing [20] that manipulates a multidimensional array called a tensor, which is typicaly has more than three "modes" can be applied (a vector is a single-mode tensor and a matrix is a dual-mode tensor). In the multilinear approach, the structures of the recorded signals and converted data can be retained. Since the structure provides additional information, multilinear approaches can be considered to be regularizations [21], [22]. Moreover, if an alternating optimization for each mode is

applied [20], [23]–[25], the number of parameters to be optimized can be small. In this way, the multilinear approaches are less ill-posed. However, multilineary formed LDA (multilinear discriminant analysis (MLDA)) can be ill-posed and overfitted, because the dimension of the projection for each mode can be large, even after multilinearization.

In this paper, we propose a novel MLDA projection with subspace constraints that produce regularization, parameter reduction, and additional prior information transferred from different datasets. The proposed method is based on uncorrelated MLDA (UMLDA) [26], which has the advantage of fewer parameters compared to other MLDA algorithms [27]. To address the known ill-posed and overfitted problems of MLDA, we regularize the parameter spaces by subspace constraints. Unlike regularization techniques that use the vector norm [9], [19], which restricts the statistics of the parameter vector, our approach restricts the parameter space to the subspaces where the solution can exist.

For the constraints, the subspaces should ideally include discriminative projections. However, such subspaces are unknown. Therefore, we propose the use of the functional connectivity of the brain [28], [29] which represents the synchronization of two brain regions or time segments. The idea behind the use of the functional connectivity for the subspace design is that the projection coefficients for strongly-connective channels or time segments take similar values.

The functional connectivity is also unknown, and we need to estimate it. Usually, the number of training samples is too small to accurately estimate the functional connectivity in the classification problem. We propose the estimation of the functional connectivity by using data collected from different subjects/sessions that follow the same experimental procedures. The subspaces are derived from the functional connectivity via graph Fourier bases [16]. In this way, the proposed constraints let the functionally connected coefficients of the projection take similar values. Although regularization techniques using the vector norm of the projection [9], [19], [30] are not able to incorporate the structure of the variables, our approach can provide this function. To evaluate the proposed method, we conducted tests with conventional approaches in single-trial ERP processing setting.

The rest of this paper is organized as follows. Section II briefly summarizes the mathematical notation and the basics of multilinear algebra. Related works are reviewed in Section III. We propose the addition of the subspace constraints to the UMLDA algorithm in Section IV. Ideas for the subspace design for the constraints based on functional brain connectivity are introduced in Section V. Section VI presents the experimental results of single-trial classification of ERPs in the oddball paradigm. Section VII concludes this paper.

II. Basic Multilinear Algebra

In this paper, lower-case boldface characters represent vectors (e.g., $\boldsymbol{x}$ and $\boldsymbol{y}$). Upper-case boldface notation denotes matrices (e.g., $\boldsymbol{X}$ and $\boldsymbol{Y}$), and upper-case calligraphic capitals denote tensors (e.g., $\mathcal{X}$ and $\mathcal{Y}$). A tensor is a multidimensional array represented as $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$. This is a tensor with N modes and the dimension of I_N for its nth mode. The ith element of a vector $\boldsymbol{x}$ is given by $[\boldsymbol{x}]_i$. An entry in the ith row and jth column of a matrix $\boldsymbol{X}$ is given by $[\boldsymbol{X}]_{i,j}$. The entry of the i_nth index in the n-mode for $n = 1, \ldots, N$ of a tensor $\mathcal{X}$ is given by $[\mathcal{X}]_{i_1, i_2, \ldots, i_N}$.

The n-mode vectors of $\mathcal{X}$ are defined as the I_n-dimensional vectors obtained from $\mathcal{X}$ by varying the index i_n while keeping the other indices fixed. The unfolding [20] of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ along the nth mode is represented by an operator denoted by $(\cdot)_{(n)}$. The matrix transformed by $(\mathcal{X})_{(n)}$ is denoted as $\boldsymbol{X}_{(n)} \in \mathbb{R}^{I_n \times \bar{I}_n}$ where $\bar{I}_n = \prod_{i=1, i \neq n}^{N} I_i$. The column vectors of $\boldsymbol{X}_{(n)}$ are the n-mode vectors of $\mathcal{X}$. The n-mode product of a tensor $\mathcal{X}$ with a matrix $\boldsymbol{A} \in \mathbb{R}^{J \times I_n}$ is denoted by $\mathcal{X} \times_n \boldsymbol{A} \in \mathbb{R}^{I_1 \times \cdots \times I_{n-1} \times J \times I_{n+1} \times \cdots \times I_N}$. The elements of the n-mode product of $\mathcal{X}$ and $\boldsymbol{A}$ are defined as

$$[\mathcal{X} \times_n \boldsymbol{A}]_{i_1, \ldots, i_{n-1}, j_n, i_{n+1}, \ldots, i_N} = \sum_{i_n = 1}^{I_n} [\mathcal{X}]_{i_1, \ldots, i_{n-1}, i_n, i_{n+1}, \ldots, i_N} [\boldsymbol{A}]_{j_n, i_n}. \quad (1)$$

The n-mode unfolding of the n-mode product can be obtained by

$$(\mathcal{X} \times_n \boldsymbol{A})_{(n)} = \boldsymbol{A} \boldsymbol{X}_{(n)}. \quad (2)$$

For convenience, we denote

$$\mathcal{X} \prod_{n=1}^{N} \times_n \boldsymbol{A}_n = \mathcal{X} \times_1 \boldsymbol{A}_1 \times_2 \boldsymbol{A}_2 \times_3 \cdots \times_N \boldsymbol{A}_N. \quad (3)$$

III. Related Works

A. Linear Discriminant Analysis

LDA is a classical supervised linear projection [11] defined as

$$\boldsymbol{\gamma} = \boldsymbol{U}^\top \boldsymbol{x}, \quad (4)$$

where $\boldsymbol{x} \in \mathbb{R}^{N_d}$ is a feature vector to be classified, $\boldsymbol{\gamma} \in \mathbb{R}^{N_r}$ is the projected vector, and $\boldsymbol{U} \in \mathbb{R}^{N_d \times N_r}$ is a matrix working as a projection, given as follows.

Consider the problem in which an observed sample $\boldsymbol{x}$ is classified into a class ψ out of N_c classes ($\psi \in \{\omega_1, \ldots, \omega_{N_c}\}$). Let $\{\boldsymbol{x}^{(m)}, \psi^{(m)}\}_{m=1}^{N_m}$ be N_m pairs of the observed sample and its class label in a given training set, where $\boldsymbol{x}^{(m)} \in \mathbb{R}^{N_d}$, and $\psi^{(m)} \in \{\omega_1, \ldots, \omega_{N_c}\}$. Let $\boldsymbol{\gamma}^{(m)}$ be the projected vector given as $\boldsymbol{\gamma}^{(m)} = \boldsymbol{U}^\top \boldsymbol{x}^{(m)}$. The projection $\boldsymbol{U}$ is found by solving

$$\max_{\boldsymbol{U}} \quad \frac{\sum_{c=1}^{N_c} |\Omega_c| \|\bar{\boldsymbol{\gamma}}_{\omega_c} - \bar{\boldsymbol{\gamma}}\|^2}{\sum_{m=1}^{N_m} \|\boldsymbol{\gamma}^{(m)} - \bar{\boldsymbol{\gamma}}_{\psi^{(m)}}\|^2}, \quad (5)$$

where $\bar{\boldsymbol{\gamma}}$ is the mean of the projected vectors given by $\bar{\boldsymbol{\gamma}} = \frac{1}{N_m} \sum_{m=1}^{N_m} \boldsymbol{\gamma}^{(m)}$, $\bar{\boldsymbol{\gamma}}_{\omega_c}$ is the mean of the projected vectors belonging to class ω_c given by $\bar{\boldsymbol{\gamma}}_{\omega_c} = \frac{1}{|\Omega_c|} \sum_{m \in \Omega_c} \boldsymbol{\gamma}^{(m)}$, Ω_c is a set of sample indices belonging to class ω_c given by $\Omega_c = \{m' \mid \psi^{(m')} = \omega_c, m' = 1, \ldots, N_m\}$, and the operator $|\cdot|$ for a set gives the number of the elements in the set. We define the

between-class scatter matrix $\boldsymbol{B}$ as

$$\boldsymbol{B} = \sum_{c=1}^{N_c} |\Omega_c| (\bar{\boldsymbol{x}}_{\omega_c} - \bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_{\omega_c} - \bar{\boldsymbol{x}})^\top, \tag{6}$$

the total scatter matrix $\boldsymbol{T}$ as

$$\boldsymbol{T} = \sum_{m=1}^{N_m} (\boldsymbol{x}^{(m)} - \bar{\boldsymbol{x}})(\boldsymbol{x}^{(m)} - \bar{\boldsymbol{x}})^\top, \tag{7}$$

and the within-class scatter matrix $\boldsymbol{W}$ as

$$\boldsymbol{W} = \boldsymbol{T} - \boldsymbol{B}, \tag{8}$$

where $\bar{\boldsymbol{x}}$ is the mean vector of $\boldsymbol{x}^{(m)}$ defined as $\bar{\boldsymbol{x}} = \frac{1}{N_m}\sum_{m=1}^{N_m} \boldsymbol{x}^{(m)}$ and $\bar{\boldsymbol{x}}_{\omega_c}$ is the mean vector of $\boldsymbol{x}$ belonging to class ω_c defined as $\bar{\boldsymbol{x}}_{\omega_c} = \frac{1}{|\Omega_c|}\sum_{m\in\Omega_c} \boldsymbol{x}^{(m)}$. By the scatter matrices, the optimization problem in (5) is transformed into

$$\max_{\boldsymbol{U}} \quad \frac{tr\left(\boldsymbol{U}^\top \boldsymbol{B}\boldsymbol{U}\right)}{tr\left(\boldsymbol{U}^\top \boldsymbol{W}\boldsymbol{U}\right)}. \tag{9}$$

B. Linear Discriminant Analysis Using the Shrunk Covariance

To solve the optimization problem in (9), the generalized eigenvalue decomposition for $\boldsymbol{B}$ and $\boldsymbol{W}$ can be used. However, for a high-dimensional feature space with a few samples, the decomposition is ill-posed. To alleviate this situation, we can use a regularization technique called shrinkage LDA (sLDA) [31]. For sLDA, the shrinkage total scatter matrix

$$\tilde{\boldsymbol{T}} = (1-\alpha)\boldsymbol{T} + \alpha\nu \boldsymbol{I}_{N_d}, \tag{10}$$

is used instead of the total scatter matrix in (7) to obtain the within-class scatter matrix in (8), where α is a regularization parameter to be tuned, and ν is defined as $\nu = tr(\boldsymbol{T})/N_d$. The parameter α is tuned analytically or empirically. For analytical selection, we can use the Ledoit-Wolf estimator [32]–[34]. For empirical selection, we can use cross-validation. This regularization has been applied to the problem of single-trial classification of ERPs [9].

C. Uncorrelated Multilinear Discriminant Analysis

UMLDA creates a feature tensor by a tensor-to-vector projection (TVP) [27] with the elementary multilinear projection (EMP) [26], [27]. The EMPs for an N-mode tensor $\mathcal{X} \in \mathbb{R}^{I_1\times I_2\times\cdots\times I_N}$ projected onto a scalar by N vectors are defined as

$$y(\mathcal{X}, U) = \mathcal{X}\prod_{n=1}^{N} \times_n \boldsymbol{u}_n^\top, \tag{11}$$

where U is the set of the EMP vectors defined as $U = \{\boldsymbol{u}_n \in \mathbb{R}^{I_n}\}_{n=1}^{N}$. A set of EMPs results in a TVP. A TVP that projects a tensor $\mathcal{X}$ onto an R-dimensional vector $\boldsymbol{y}$ by NR vectors is defined as

$$\boldsymbol{y}(\mathcal{X}, \Theta) = \left[y(\mathcal{X}, U^{(1)}), \ldots, y(\mathcal{X}, U^{(R)})\right]^\top, \tag{12}$$

where $U^{(r)} = \{\boldsymbol{u}_n^{(r)} \in \mathbb{R}^{I_n}\}_{n=1}^{N}$, and $\Theta = \{U^{(r)}\}_{r=1}^{R}$ for later convenience.

The set of EMPs for UMLDA is obtained as follows. Consider the problem in which an observed sample $\mathcal{X}$ is classified into a class ψ out of N_c classes ($\psi \in \{\omega_1, \ldots, \omega_{N_c}\}$). Let $\{\mathcal{X}^{(m)}, \psi^{(m)}\}_{m=1}^{N_m}$ be the set of N_m pairs of the observed sample and its class label in a given training set, where $\mathcal{X}^{(m)} \in \mathbb{R}^{I_1\times\cdots\times I_N}$, and $\psi^{(m)} \in \{\omega_1, \ldots, \omega_{N_c}\}$. Let $\boldsymbol{g}^{(r)}$ be the vector of the training samples projected by the rth EMP defined as

$$\boldsymbol{g}^{(r)} = \left[y(\mathcal{X}^{(1)}, U^{(r)}), \ldots, y(\mathcal{X}^{(N_m)}, U^{(r)})\right]^\top. \tag{13}$$

UMLDA finds the EMPs in such a way that the vectors $\{\boldsymbol{g}^{(r)}\}_{r=1}^{R}$ are uncorrelated with each other [35]. The optimization problem with the constraints of the uncorrelation is given as [26]

$$\begin{aligned} \max_{\Theta} \quad & J(\Theta) = \frac{\sum_{c=1}^{N_c} |\Omega_c| \|\bar{\boldsymbol{y}}_{\omega_c} - \bar{\boldsymbol{y}}\|^2}{\sum_{m=1}^{N_m} \|\boldsymbol{y}(\mathcal{X}^{(m)}, \Theta) - \bar{\boldsymbol{y}}_{\psi^{(m)}}\|^2}, \\ \text{subject to} \quad & \boldsymbol{g}^{(r)\top}\boldsymbol{g}^{(r')} = 0, \quad r, r' = 1, \ldots, R,\ r \neq r', \end{aligned} \tag{14}$$

where $\bar{\boldsymbol{y}}$ is the mean of the projected vectors defined as $\bar{\boldsymbol{y}} = \frac{1}{N_m}\sum_{m=1}^{N_m} \boldsymbol{y}(\mathcal{X}^{(m)}, \Theta)$, and $\bar{\boldsymbol{y}}_{\omega_c}$ is the mean of the projected vectors belonging to class ω_c defined as $\bar{\boldsymbol{y}}_{\omega_c} = \frac{1}{|\Omega_c|}\sum_{m\in\Omega_c} \boldsymbol{y}(\mathcal{X}^{(m)}, \Theta)$.

It is difficult to find all of the parameters in (14) simultaneously. Therefore, we adopt sequential optimization, which finds the parameters one by one with respect to the EMP index n [26]. In sequential optimization, after $\{U^{(r)}\}_{r=1}^{p-1}$ are found, $U^{(p)}$ is obtained under a constraint of $\{U^{(r)}\}_{r=1}^{p-1}$. By this sequential solution, the problem in (14), for only the first EMP, is given as

$$U^{(1)} = \arg\max_{U}\ J'(U) \tag{15}$$

with

$$J'(U) = \frac{\sum_{c=1}^{N_c} |\Omega_c| \|\bar{y}_{\omega_c} - \bar{y}\|^2}{\sum_{m=1}^{N_m} \|y(\mathcal{X}^{(m)}, U) - \bar{y}_{\psi^{(m)}}\|^2}, \tag{16}$$

where $\bar{y}$ is the mean of the projected scalar defined as $\bar{y} = \frac{1}{N_m}\sum_{m=1}^{N_m} y(\mathcal{X}^{(m)}, U)$, and $\bar{y}_{\omega_c}$ is the mean of the projected scalar belonging to class ω_c defined as $\bar{y} = \frac{1}{|\Omega_c|}\sum_{m\in\Omega_c} y(\mathcal{X}^{(m)}, U)$. Next, for the pth EMP ($p > 1$), we solve (15) with the constraints

$$\begin{aligned} U^{(p)} = \arg\max_{U} \quad & J'(U), \\ \text{subject to} \quad & \boldsymbol{g}^{(p)\top}\boldsymbol{g}^{(r)} = 0, \quad r = 1, \ldots p-1. \end{aligned} \tag{17}$$

For the above problem, the EMPs $(U^{(1)}, \ldots, U^{(p-1)})$ should be given in advance for $\boldsymbol{g}^{(1)}, \ldots, \boldsymbol{g}^{(p-1)}$.

The problem in (17) does not have a closed-form solution. Therefore, we apply an alternating solution over the tensor mode [26].

IV. Uncorrelated Multilinear Discriminant Analysis With Subspace Constraints

The proposed subspace-constrained UMLDA (SMLDA) adds subspace constraints for the parameter spaces to the cost function of UMLDA in (14). The optimization problem for SMLDA

is

$$\begin{aligned} \max_{\Theta} \quad & J(\Theta), \\ \text{subject to} \quad & {\boldsymbol{g}^{(r)}}^{\top}\boldsymbol{g}^{(r')} = 0, \quad r, r' = 1, \ldots, R,\ r \neq r' \\ & \boldsymbol{u}_n^{(r)} \in S_n, \quad \forall r, \forall n, \end{aligned} \tag{18}$$

where S_n is a subspace in the parameter space for $\{\boldsymbol{u}_n^{(r)}\}_{r=1}^{R}$. As for UMLDA, sequential optimization can be adopted to solve (18). For sequential optimization, the subproblem for the pth EMP is

$$\begin{aligned} U^{(p)} = & \arg\max_{U} \ J'(U), \\ \text{subject to} \quad & {\boldsymbol{g}^{(p)}}^{\top}\boldsymbol{g}^{(r)} = 0, \quad r = 1, \ldots p-1, \\ & \boldsymbol{u}_n \in S_n, \quad \forall n. \end{aligned} \tag{19}$$

The above problem does not have a closed-form solution similar to that of UMLDA. Therefore, we apply an alternating optimization. In an alternating optimization, the problem is divided into N subproblems with respect to the modes. First, we set the initial values for the parameters. We solve the subproblems in an alternating manner and stop optimizing if the cost function converges. The subproblem for $\boldsymbol{u}_n^{(p)}$ in (19) is

$$\begin{aligned} \boldsymbol{u}_n^{(p)} = & \arg\max_{\boldsymbol{u}} \ \frac{\boldsymbol{u}^{\top}\boldsymbol{B}_n^{(p)}\boldsymbol{u}}{\boldsymbol{u}^{\top}\boldsymbol{W}_n^{(p)}\boldsymbol{u}}, \\ \text{subject to} \quad & {\boldsymbol{g}^{(p)}}^{\top}\boldsymbol{g}^{(r)} = 0, \quad r = 1, \ldots p-1, \\ & \boldsymbol{u} \in S_n, \end{aligned} \tag{20}$$

where $\boldsymbol{B}_n^{(p)}$ is a matrix corresponding to a between-class scatter matrix for LDA defined as

$$\boldsymbol{B}_n^{(p)} = \sum_{c=1}^{N_c} |\Omega_c| (\bar{\boldsymbol{z}}_{\omega_c,n}^{(p)} - \bar{\boldsymbol{z}}_n^{(p)})(\bar{\boldsymbol{z}}_{\omega_c,n}^{(p)} - \bar{\boldsymbol{z}}_n^{(p)})^{\top}, \tag{21}$$

$\boldsymbol{W}_n^{(p)}$ is a matrix corresponding to a within-class scatter matrix for LDA defined as

$$\boldsymbol{W}_n^{(p)} = \sum_{m=1}^{N_m} (\boldsymbol{z}_n^{(p,m)} - \bar{\boldsymbol{z}}_{\psi^{(m)},n}^{(p)})(\boldsymbol{z}_n^{(p,m)} - \bar{\boldsymbol{z}}_{\psi^{(m)},n}^{(p)})^{\top}, \tag{22}$$

$\boldsymbol{z}_n^{(p,m)}$ is the mth sample projected by $\{\boldsymbol{u}_{n'}^{(p)}\}_{n'=1,n'\neq n}^{N}$ defined as

$$\boldsymbol{z}_n^{(p,m)} = \left(\mathcal{X}^{(m)} \prod_{n'=1,n'\neq n}^{N} \times_{n'} \boldsymbol{u}_{n'}^{(p)} \right)_{(n)}, \tag{23}$$

the mean vector of $\boldsymbol{z}_n^{(p,m)}$ belonging to ω_c is defined as $\bar{\boldsymbol{z}}_{\omega_c,n}^{(p)} = \frac{1}{|\Omega_c|}\sum_{m\in\Omega_i} \boldsymbol{z}_n^{(p,m)}$, and the mean vector of $\boldsymbol{z}_n^{(p,m)}$ is defined as $\bar{\boldsymbol{z}}_n^{(p)} = \frac{1}{N_m}\sum_{m=1}^{N_m} \boldsymbol{z}_n^{(p,m)}$. The optimization problem in (20) has $p-1+1$ constraints. The first $p-1$ constraints (${\boldsymbol{g}^{(p)}}^{\top}\boldsymbol{g}^{(r)} = 0,\ r = 1, \ldots, p-1$) are equivalent to the equations

$$\boldsymbol{u}^{\top}\left[\boldsymbol{z}_n^{(p,1)}, \ldots, \boldsymbol{z}_n^{(p,N_m)}\right]\boldsymbol{g}^{(r)} = 0, \quad r = 1, \ldots, p-1. \tag{24}$$

As we define $\boldsymbol{H}_n^{(p)} \in \mathbb{R}^{I_n \times (p-1)}$ as

$$\boldsymbol{H}_n^{(p)} = \left[\boldsymbol{z}_n^{(p,1)}, \ldots, \boldsymbol{z}_n^{(p,N_m)}\right]\left[\boldsymbol{g}^{(1)}, \ldots, \boldsymbol{g}^{(p-1)}\right], \tag{25}$$

these $p-1$ constraints can be represented in a matrix-vector form as

$$\boldsymbol{u}^{\top}\boldsymbol{H}_n^{(p)} = \boldsymbol{0}, \tag{26}$$

where $\boldsymbol{0} = [0, \ldots, 0]^{\top} \in \mathbb{R}^{p-1}$. Let $H_n^{(p)}$ be a subspace spanned by the column vectors of $\boldsymbol{H}_n^{(p)}$. The orthogonal complement of $H_n^{(p)}$ is denoted by ${H_n^{(p)}}^{\perp}$. The $p-1+1$ constraints of (20) are equivalent to the constraint

$$\boldsymbol{u} \in V_n^{(p)} = \left\{ S_n \cap {H_n^{(p)}}^{\perp} \right\}, \tag{27}$$

where $V_n^{(p)}$ is the intersection of S_n and ${H_n^{(p)}}^{\perp}$. If $V_n^{(p)} = \{\boldsymbol{o}\}$, there is no space where $\boldsymbol{u}$ can exist, and $\boldsymbol{u}$ has no solution. Therefore, for SMLDA to have a solution, the problem must satisfy $R \leq \min\{I_1, \ldots, I_N\}$ and $\mathrm{rank}(S_n) > 0$ for $n = 1, \ldots, N$. In addition, if $\mathrm{rank}(S_n) = I_n$ for $n = 1, \ldots, N$, SMLDA is equivalent to UMLDA. Let $\{\boldsymbol{v}_{n,1}^{(p)}, \ldots, \boldsymbol{v}_{n,N_v}^{(p)}\}$ be an orthogonal basis of $V_n^{(p)}$, where $N_v = \mathrm{rank}(V_n^{(p)})$. The projection vector $\boldsymbol{u}$ can be formulated by the linear combination

$$\boldsymbol{u} = a_1\boldsymbol{v}_{n,1}^{(p)} + \cdots + a_{N_v}\boldsymbol{v}_{n,N_v}^{(p)} = \boldsymbol{V}_n^{(p)}\boldsymbol{a}, \tag{28}$$

where $\boldsymbol{V}_n^{(p)} \in \mathbb{R}^{I_n \times N_v}$ is a matrix defined as $\boldsymbol{V}_n^{(p)} = [\boldsymbol{v}_{n,1}^{(p)}, \ldots, \boldsymbol{v}_{n,N_v}^{(p)}]$, $\{a_s\}_{s=1}^{N_v}$ are coefficients for the linear combination, and $\boldsymbol{a} \in \mathbb{R}^{N_v}$ is a vector defined as $\boldsymbol{a} = [a_1, \ldots, a_{N_v}]^{\top}$. The optimization problem in (20) can be translated in finding $\boldsymbol{a}$ by a change of variables. The approach for finding $\boldsymbol{a}$ is

$$\boldsymbol{a} = \arg\max_{\boldsymbol{a}} \ \frac{\boldsymbol{a}^{\top}\tilde{\boldsymbol{B}}_n^{(p)}\boldsymbol{a}}{\boldsymbol{a}^{\top}\tilde{\boldsymbol{W}}_n^{(p)}\boldsymbol{a}}, \tag{29}$$

where $\tilde{\boldsymbol{B}}_n^{(p)}$ and $\tilde{\boldsymbol{W}}_n^{(p)}$ are $N_v \times N_v$ matrices defined as

$$\tilde{\boldsymbol{B}}_n^{(p)} = {\boldsymbol{V}_n^{(p)}}^{\top}\boldsymbol{B}_n^{(p)}\boldsymbol{V}_n^{(p)} \tag{30}$$

and

$$\tilde{\boldsymbol{W}}_n^{(p)} = {\boldsymbol{V}_n^{(p)}}^{\top}\boldsymbol{W}_n^{(p)}\boldsymbol{V}_n^{(p)}. \tag{31}$$

The solution of (29) is given by generalized eigenvalue decomposition as follows:

$$\tilde{\boldsymbol{B}}_n^{(p)}\boldsymbol{a} = \lambda\tilde{\boldsymbol{W}}_n^{(p)}\boldsymbol{a}, \tag{32}$$

where λ is an eigenvalue. The solution is given as the eigenvector corresponding to the largest eigenvalue of (32).

The procedure of UMLDA with subspace constraints is summarized in Algorithm 1 as a pseudocode. The pseudocode includes the procedure for the design of the subspaces presented in Section V.

V. Design for Subspace Constraints

In this section, the subspace defined in (18) for SMLDA are proposed for EEG classification in ERP-based BCIs. The idea

Algorithm 1: SMLDA with the functional connectivity.

Input: $\{\mathcal{X}^{(m)}, \psi^{(m)}\}_{m=1}^{N_m}$: the training samples, $\{\mathcal{T}^{(m)}\}_{m=1}^{N_a}$: the complementary data.
Parameters: R: the reduced dimension, $\{D_n\}_{n=1}^{N}$: the dimensions of the subspaces.
Output: $\Theta = \{\boldsymbol{u}_n^{(r)}\}_{n=1,r=1}^{N,R}$: the EMPs.
{Design of the subspaces}
for $n = 1, \ldots, N$ **do**
  Calculate $\boldsymbol{A}_n$ using $\{\mathcal{T}^{(m)}\}_{m=1}^{N_a}$ by (33)
  Obtain S_n from the graph Fourier basis of $\bar{\boldsymbol{L}}_n$ by (45)
end for
{SMLDA}
for $p = 1, \ldots, R$ **do**
  Initialize $\{\boldsymbol{u}_n^{(p)}\}_{n=1}^{N}$.
  Calculate $\{\boldsymbol{g}^{(r)}\}_{r=1}^{p-1}$ by (13).
  Set the index of the iteration as $k = 0$.
  repeat
    for $n = 1, \ldots, N$ **do**
      $k \leftarrow k + 1$.
      Calculate $\boldsymbol{H}_n^{(p)}$ by (25).
      Calculate an orthogonal basis $\{\boldsymbol{v}_{n,1}^{(p)}, \ldots, \boldsymbol{v}_{n,N_v}^{(p)}\}$ of $S_n \cap H_n^{(p)\perp}$.
      Obtain $\boldsymbol{a}$ by solving (29) as the largest generalized eigenvector of (32).
      Obtain $\boldsymbol{u}_n^{(p)}$ by calculating (28).
      Calculate the cost as $C_k \leftarrow J'(U^{(p)})$ defined in (15).
    end for
  until $C_k - C_{k-1}$ is sufficiently small.
end for

behind the subspaces is to incorporate brain functional connectivity into SMLDA. In Section V-A, functional connectivity is defined as an adjacency matrix estimated from the data observed in different experimental subjects/sessions. In Section V-B, we introduce the method for deriving a subspace from the adjacency matrix.

A. Adjacency Matrix of Functional Connectivities

The functional connectivity of the brain is an important concept in cognitive neuroscience [36], [37]. In general, functional connectivity represents the synchronization of two different brain regions or time segments. Therefore, functional connectivity can represent the functional structure of the brain. In this study, the functional connectivity between two optimization parameters is obtained from a signal set that is different from the training set. The signal set is obtained from recordings of experimental subjects/sessions that are different from those in the testing set. We call the signal set the "complementary data."

Although the EEG signals are given as a matrix (channel by time), we use the tensor form to represent the signals for generality. The complementary data, composed of N_a signals, are denoted as $\{\mathcal{T}^{(m)} \in \mathbb{R}^{I_1 \times \cdots \times I_N}\}_{m=1}^{N_a}$. We assume that the signals in the complementary data are in the same format (size) as those in the testing set. The ith row and jth column of the adjacency matrix for the nth mode ($n \in \{1, \ldots, N\}$) are defined as

$$[\boldsymbol{A}_n]_{i,j} = \begin{cases} 0, & i = j, \\ C_\chi(\boldsymbol{t}_n^{(i)}, \boldsymbol{t}_n^{(j)}), & \text{otherwise}, \end{cases} \tag{33}$$

for $i, j = 1, \ldots, I_n$, where C_χ is an operator that measures the χ-type connectivity between two input vectors, $\boldsymbol{t}_n^{(l)} \in \mathbb{R}^{N_p}$ is the vector composed of the l elements in the nth mode of the complementary data given by

$$\left[\boldsymbol{t}_n^{(l)}\right]_{(m-1)\bar{N}_n + k} = \left[\boldsymbol{T}_{(n)}^{(m)}\right]_{l,k}, \tag{34}$$

for $k = 1, \ldots, \bar{N}_n$ and $m = 1, \ldots, N_a$, $\bar{N}_n = \prod_{n'=1, n' \neq n}^{N} I_{n'}$, and $N_p = N_p \bar{N}_n$.

We propose two types of operators, $C_{\mathrm{corr}}(\cdot, \cdot)$ and $C_{\mathrm{mi}}(\cdot, \cdot)$ as follows.

1) Normalized Correlation: The operator that estimates the normalized correlation from the input vectors is defined as

$$C_{\mathrm{corr}}(\boldsymbol{q}_1, \boldsymbol{q}_2) = \frac{\tilde{\boldsymbol{q}}_1^\top \tilde{\boldsymbol{q}}_2}{\|\tilde{\boldsymbol{q}}_1\| \|\tilde{\boldsymbol{q}}_2\|}, \tag{35}$$

where $\tilde{\boldsymbol{q}}_1$ and $\tilde{\boldsymbol{q}}_2$ are the centered samples given by $\tilde{\boldsymbol{q}}_i = \boldsymbol{q}_i - \frac{1}{N_p} \sum_{m=1}^{N_p} [\boldsymbol{q}_i]_m$, $i = 1, 2$.

2) Mutual Information: The mutual information is a measure of the joint dependence of two random variables. In this study, we assume Gaussian random variables with the parameters estimated from the input vectors. The definition of the mutual information of two variables is given as

$$I(Q_1; Q_2) = \int_{Q_2} \int_{Q_1} p(q_1, q_2) \log \frac{p(q_1, q_2)}{p(q_1) p(q_2)} dq_1 dq_2. \tag{36}$$

The operator for the mutual information $C_{\mathrm{mi}}(\boldsymbol{q}_1, \boldsymbol{q}_2)$ is given as $I(Q_1; Q_2)$ with the probability distributions estimated from the input vectors, $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$. The joint probability distribution $p(q_1, q_2)$ and the marginal probability distributions, $p(q_1)$ and $p(q_2)$, are assumed to be Gaussian:

$$p(q_1, q_2) = \mathcal{N}(q_1, q_2 \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{37}$$

and

$$p(q_i) = \mathcal{N}(q_i \mid \mu_i, \sigma_i^2), \quad i = 1, 2. \tag{38}$$

The maximum likelihood estimate of the parameters in (37) and (38) is given by

$$\hat{\mu}_i = \frac{1}{N_p} \sum_{m=1}^{N_p} [\boldsymbol{q}_i]_m, \quad i = 1, 2, \tag{39}$$

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{N_p} \sum_{m=1}^{N_p} (\boldsymbol{z}_m - \hat{\boldsymbol{\mu}})(\boldsymbol{z}_m - \hat{\boldsymbol{\mu}})^\top, \tag{40}$$

and $\hat{\sigma}_i^2 = [\hat{\boldsymbol{\Sigma}}]_{i,i}$, where $\hat{\boldsymbol{\mu}} = [\hat{\mu}_1, \hat{\mu}_2]^\top$, and $\boldsymbol{z}_m = [[\boldsymbol{q}_1]_m, [\boldsymbol{q}_2]_m]^\top$. Using the estimated parameters, the operator giving the mutual information is reduced to

$$C_{\mathrm{mi}}(\boldsymbol{q}_1, \boldsymbol{q}_2) = \frac{1}{2} \log \frac{\hat{\sigma}_1^2 \hat{\sigma}_2^2}{|\hat{\boldsymbol{\Sigma}}|}. \tag{41}$$

B. Subspace by a Graph Fourier Basis

An adjacency matrix represents a graph having edges and nodes. The eigenvectors of the graph Laplacian, which can be derived from the adjacency matrix, are called the graph Fourier basis [38]. The ith and jth elements of the eigenvector corresponding to the small eigenvalues take similar values if the ith and jth nodes are strongly connected (the corresponding edge weight is large). These eigenvectors corresponding to the small eigenvalues can be called smooth vectors across the graph [16]. We employ the feature of the graph Fourier basis to obtain the subspaces constraining the optimized parameters that are smooth across the functional connectivities.

The subspace for the nth mode S_n, with the adjacency matrix given by (33), is obtained as follows. The graph Laplacian is defined as

$$\boldsymbol{L}_n = \boldsymbol{D}_n - \boldsymbol{A}_n, \tag{42}$$

where $\boldsymbol{D}_n$ is a diagonal matrix whose diagonal elements are given by $[\boldsymbol{D}_n]_{i,i} = \sum_{k=1}^{I_n} [\boldsymbol{A}_n]_{i,k}$. The normalized graph Laplacian is also defined as

$$\bar{\boldsymbol{L}}_n = \boldsymbol{D}_n^{-\frac{1}{2}} \boldsymbol{L}_n \boldsymbol{D}_n^{-\frac{1}{2}}. \tag{43}$$

The orthonormal basis of S_n satisfies

$$\bar{\boldsymbol{L}}_n \boldsymbol{s}_n^{(i)} = \lambda_i \boldsymbol{s}_n^{(i)}. \tag{44}$$

The indices of the eigenvalues and the corresponding eigenvectors are determined in such a way that $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{I_n}$. Using the eigenvectors, the subspace is designed as

$$S_n = \mathrm{Span}\{\boldsymbol{s}_n^{(1)}, \boldsymbol{s}_n^{(2)}, \ldots, \boldsymbol{s}_n^{(D_n)}\}, \tag{45}$$

where D_n is the dimension of S_n and is a parameter of SMLDA.

VI. Experiments

We conducted classification experiments of single-trial ERPs to evaluate the performance of the proposed SMLDA projection. The SMLDA method was compared with LDA, including a dimensionality reduction by PCA; common spatial pattern (CSP) spatial filter [9]; xDAWN spatial filter [39], LDA with the shrunk covariance matrix, and UMLDA.

A. Data Description

Five datasets were used. The information is summarized in Table I.

1) DATASET-A/B/C: In the BCI experiments, 16 persons participated in the recording of the datasets. These datasets consisted of EEG signals with P300-based BCI responses evoked by auditory, visual, and audiovisual stimuli. Detailed information about these datasets is presented in [40]. The main problem of these datasets has been the classification of the EEG signals as either target (class 1) or nontarget (class 2) stimuli. The electrodes were placed at Cz, CPz, POz, Pz, P1, P2, C3, C4, O1, O2, T7, T8, P3, P4, F3, and F4 in the International 10/20 System, referenced to the active electrodes installed at the earlobes.

2) DATASET-D: In the second series of BCI experiments, 12 people (23.6 ± 1.7 years; two females) participated in the recording of the dataset. The EEG recordings were conducted at Toyohashi University of Technology, Aichi, Japan. The experiment was approved by the Committee for Human Research of Toyohashi University of Technology, and all participants gave written informed consent. The subjects performed the two-choice response time (TCRT) task without receiving feedback about the accuracy of their responses [41]. For visual stimuli, square and circle symbols were presented randomly on an LCD display every 1.5 ± 0.1 s. We instructed the subjects to click the left button when a circle appeared, and to click the right button when a square appeared. For each subject, the button assignment was random. A single session consisted of 310 stimulus presentations. Two sessions were conducted for each subject.

The probability of the occurrence of each symbol varied between the two sessions. In one of the sessions, the probability of presentation of a circle was 70%, and that of the square was 30%. In the other session, the probability of occurrence of the circle was 30%, and that of the square was 70%. The stimuli with probabilities of 70% and 30% were called standard and odd, respectively. The subjects did not know which symbol corresponded to the standard or odd stimulus. The main problem for this dataset was to classify the EEG signals as either standard (class 1) or odd (class 2) stimuli. We recorded the EEG signals using a Biosemi ActiveTwo system. The electrodes covered the whole head (Fig. 1), referenced to the active electrodes installed at the earlobes.

TABLE I
SUMMARY OF DATASETS

	DATASET				
	A	B	C	D	E
Stimulus	Audio	Visual	Audiovisual	Visual	Visual
Task	Spelling	Spelling	Spelling	TCRT[1]	Spelling
# subjects	16	16	16	11	10
# channels	16	16	16	64	8
f_s[2] [Hz]	512	512	512	512	256

[1] TCRT stands for two-choice response task.
[2] f_s stands for the sampling frequency.

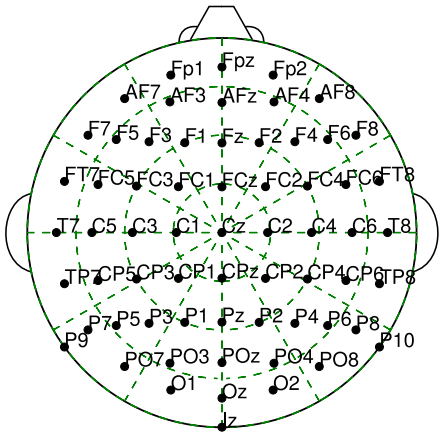


Fig. 1. The electrode arrangement for *DATASET-D*.

3) DATASET-E: As an experiment with a public dataset, the dataset provided by g.tec medical engineering GmbH was used [42]. The dataset consisted of EEG signals with a P300-Speller evoked by visual stimuli. Detailed information about this dataset is presented in [42]. The main problem of this dataset has been the classification of EEG signals as target (class 1) or nontarget (class 2) stimuli. The electrodes were placed at Fz, Cz, P3, Pz, P4, PO7, Oz, and PO8.

B. Classification Procedure

The ERP signal for each trial was segmented from 100 to 800 ms after the stimulus onset to create features. In order to remove trials contaminated by muscular artifacts related to eye blinks or other movement, a threshold of 80 μV was used for the EEG amplitude. After the removal of artifacts, the number of samples for each subject in *DATASET-D* was reduced to less than 100. Therefore, we removed data from the classification experiment, and the count of the subjects for *DATASET-D* was 11. A Butterworth bandpass filter whose passband was 0.1–15 Hz was applied. We down-sampled the ERP feature to an effective sampling frequency of 32 Hz.

As a result of the above preprocessing procedures, the feature set of each BCI trial for *DATASET-A*, *-B*, and *-C* was represented by a matrix with a size of 16×23. In *DATASET-D*, the feature matrix size was 64×23. In *DATASET-E*, the matrix size was 8×23. For the LDA projection, we vectorized the signal matrices. To alleviate the ill-posedness caused by the dimension of the vectors, which were generally larger than the number of the training samples, we applied dimensionality reduction by PCA before obtaining the LDA projection. The projection is denoted by PCA + LDA.

For the LDA projection with the shrunk covariance matrix, the regularization parameter α in (7) was selected empirically. This projection is denoted by sLDA. The extraction of the logarithm signal powers [9] of the CSP-filtered signals, followed by the projection with LDA, is denoted by CSP + LDA. The projection with the xDAWN spatial filter, followed by the projection with LDA, is denoted by xDAWN + LDA. In this projection, the spatial filtered signals were vectorized. For the UMLDA and SMLDA projections, we did not vectorize the feature matrices. The EEG features were processed as two-mode tensors (the first and second modes corresponded to the time and spatial domains (channels), respectively).

The complementary data for SMLDA consisted of the data recorded with subjects that were different from the tested ones. For example, when we tested Subject 1, the complementary data consisted of the data from Subject 2, Subject 3, ..., and Subject 16 in *DATASET-A*. SMLDA with the subspaces using C_{corr} is denoted by SMLDAc, and that using C_{mi} is denoted by SMLDAm.

The parameters for each method were tuned with the training samples. For the tuning, a nested leave-one-out cross-validation was adopted. The reduced dimension in the PCA + LDA projection was chosen out of $\{1, 2, \ldots, 100\}$. The regularization parameter α for sLDA was chosen out of $\{0.01, 0.02, \ldots, 1\}$. The number of CSP filters for CSP + LDA was chosen out of $\{1, 2, \ldots, 50\}$. The number of xDAWN spatial filters for xDAWN + LDA was chosen out of $\{1, 2, 3, 4\}$. The dimensions of the subspaces for the spatial and time domains (D_1 and D_2) in the SMLDA projection were chosen out of $\{1, 2, \ldots, 15\}$. The reduced dimension R in SMLDA and UMLDA was chosen out of $\{1, 2, \ldots, 5\}$.

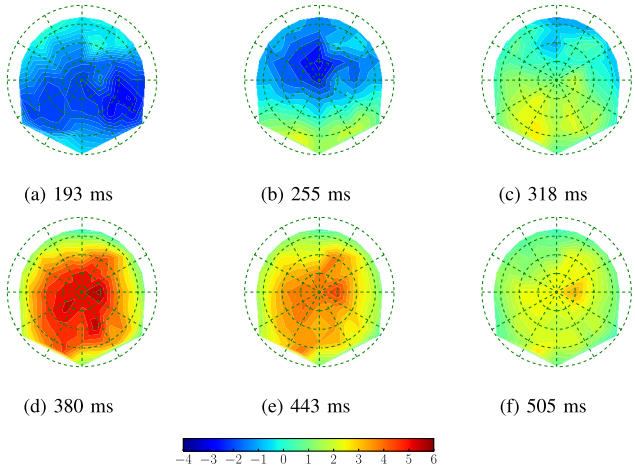


Fig. 2. Grand-averaged EEG potentials [μV] at each time index in *DATASET-D*. The potential decreases as the color changes from red to blue (see the colorbar below the figures).

The vectors projected by LDA, sLDA, UMLDA, and SMLDA were classified by Bayes' rule [43] using a Gaussian distribution:

$$\hat{c} = \arg \max_{c \in \{\text{class 1}, \text{class 2}\}} \mathcal{N}(\boldsymbol{y} \mid \boldsymbol{m}_c, \boldsymbol{\Sigma}_c) p(c), \qquad (46)$$

where $\boldsymbol{y}$ is a projected sample (scalar or vector), $\boldsymbol{m}_c$ and $\boldsymbol{\Sigma}_c$ are estimated as the sample mean and covariance of the projected samples in the training sets belonging to class c, and $p(c)$ is estimated as $p(\text{class } 1) = p(\text{class } 2) = 0.5$.

We randomly removed some samples, so that the numbers of samples for two classes were the same. Because the numbers of trials were different for each subject, we randomly selected 100 samples (50 samples per class) for the training sets, and the remaining samples were used to evaluate the classification accuracy. We validated the classification accuracy by repeating the selection for the training and testing samples 100 times. The classification accuracy rates were given as the average over the repetitions.

C. Results

Fig. 2 shows the EEG potentials in *DATASET-D*, averaged over all trials (grand average), on topographical maps. Potential fluctuations related to the events are observed at 193 ms (Fig. 2(a)) and 380 ms (Fig. 2(d)) in the parietal area. The early component is considered to be N100, reflecting the attention to the event, and the late component is considered to be P300 that is modulated by the probability of the event in the oddball paradigm [44]. The grand-averaged EEG waveforms in *DATASET-D* are shown in Fig. 3. Differences between the classes are observed between 200 ms and 300 ms and, and at 380 ms. The time range from 250 ms to 300 ms shows a signif-

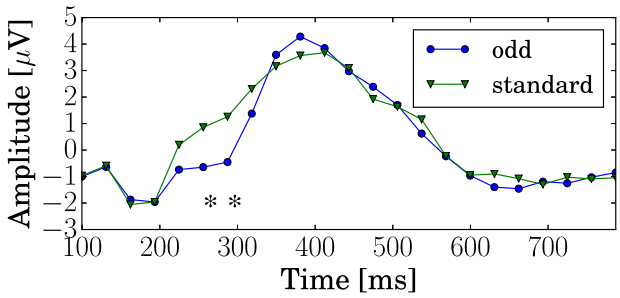


Fig. 3. Grand-averaged EEG potentials observed at Pz in *DATASET-D*. The asterisks in the figure indicate the results of a t-test at each time index ($*$: $p < 0.01$).

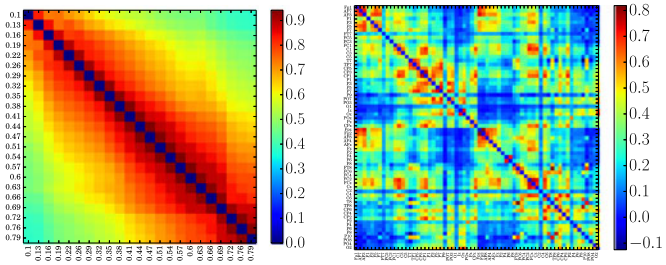


Fig. 4. Adjacency matrices with the normalized correlation operator C_{corr}.

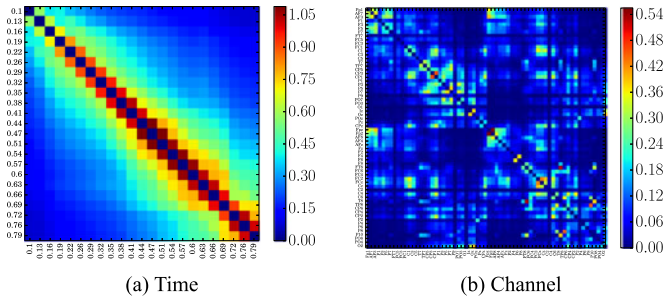


Fig. 5. Adjacency matrices with the mutual information operator C_{mi}.

icant difference ($p < 0.01$) between the two classes. This component is considered to be N200 (or the mismatch negativity) and P300 (especially its subcomponent P3a) [45], [46].

The adjacency matrices are represented in Figs. 4 and 5. They were obtained from all of the experimental recordings (all subjects) of *DATASET-D*. Figs. 4(a) and 5(b) show that the functional connectivities are high if the optimization parameters (time segments) are close in the temporal domain. The functional connectivities in the spatial (channel) domain (Figs. 4(b) and 5(b)) also show that the connectivity is high between pairs of channels of the electrodes that were located at a close distance. Moreover, the connectivity among the electrodes in the frontal and parietal areas are high. Fig. 6 shows the vectors in the graph Fourier basis given by the graph Laplacian of the adjacency matrix shown in Fig. 5(a). These vectors are the eigenvectors corresponding to some of the smallest eigenvalues obtained in (44). The lines in Fig. 6 correspond to the indices of the eigenvalues. Because the functional connectivities between

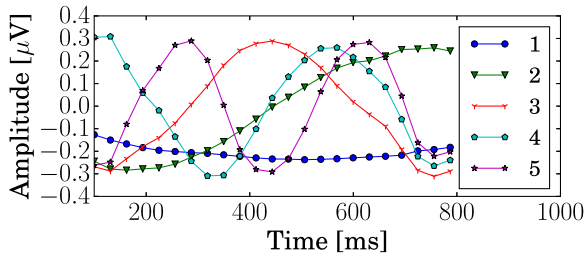


Fig. 6. Representatives of first five bases in the graph Fourier transform with C_{mi} in the temporal mode.

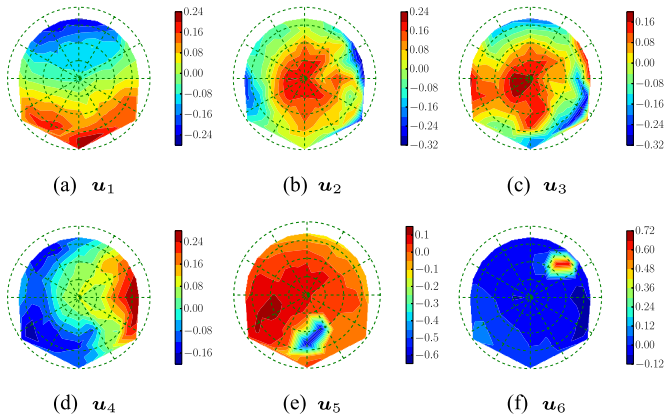
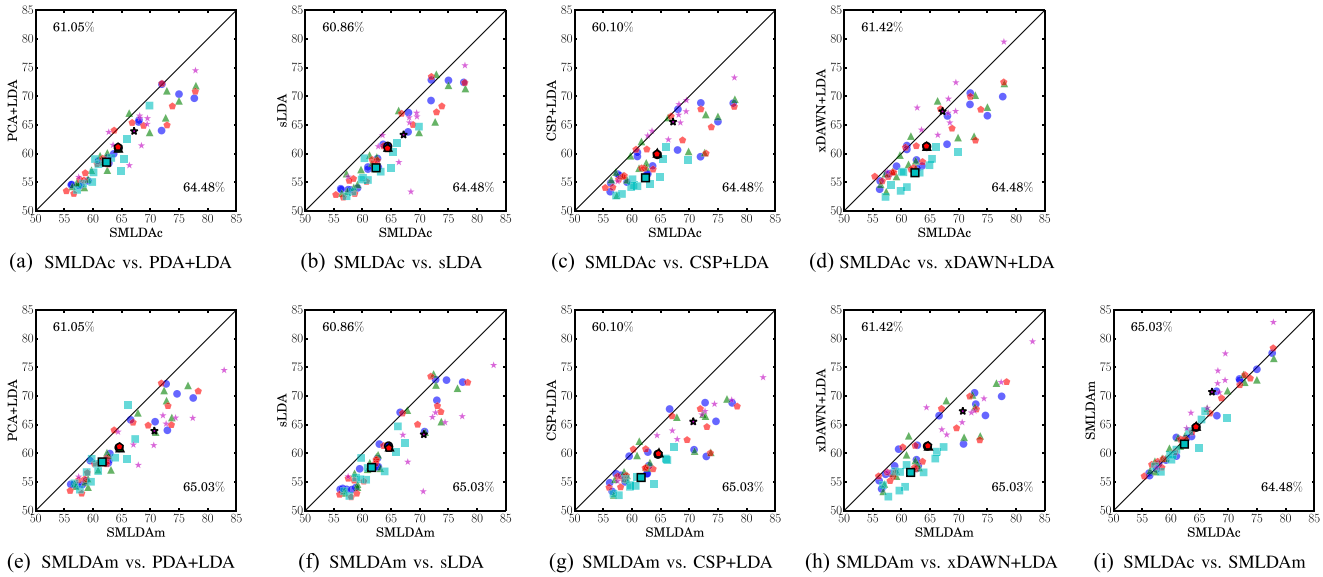


Fig. 7. Representatives of first six bases of the graph Fourier transform with C_{mi} in the spatial mode.

successive time segments are high, the bases are similar to those of Fourier transform. Fig. 7 shows the vectors in the basis of the graph Fourier transform given by the adjacency matrix of Fig. 2. As shown in Fig. 2, the synchronization of the parietal area in the grand-averaged potential would affect the high coefficients of the parietal area in Fig. 7(b) and (c).

The obtained classification accuracies are shown in Fig. 8, and their averages are listed in Table II. The classification performance of UMLDA for *DATASET-A, -B, -C,* and *-D* is lower than that of the other methods discussed in this paper. On the other hand, the classification accuracy for *DATASET-E* with a lower number of channels, is similar to PCA + LDA and sLDA. This means that UMLDA has an overfitting problem for *DATASET-A*, *-B*, *-C*, and *-D*. This suggests that the reduction of the optimization parameters by multilinearization is not good enough to solve the overfitting problem in our experimental setting. By using subspace constraint to solve the problem, the proposed methods (SMLDAc and SMLDAm) increase the classification accuracy by 3–4% on average.

Fig. 9 shows the projection vectors for the first (temporal) mode in the first EMP ($\boldsymbol{u}_1^{(1)}$) of SMLDAm. The projections are given by 100 samples of Subject 1 in *DATASET-D*. The lines in Fig. 9 correspond to the numbers of the dimension of the subspace D_1. We observe that the projection vector is smooth in the temporal domain with a small dimension of the subspace. The coefficients ($D_1 = 6$) around 200 ms are higher



(a) SMLDAc vs. PDA+LDA (b) SMLDAc vs. sLDA (c) SMLDAc vs. CSP+LDA (d) SMLDAc vs. xDAWN+LDA

(e) SMLDAm vs. PDA+LDA (f) SMLDAm vs. sLDA (g) SMLDAm vs. CSP+LDA (h) SMLDAm vs. xDAWN+LDA (i) SMLDAc vs. SMLDAm

Fig. 8. Results of the classification accuracy rates. The x and y axes in each panel show the classification accuracy rates [%] of the tested method. The circles, triangles, pentagons, squares, and stars represent the classification accuracy of *DATASET-A*, *-B*, *-C*, *-D*, and *-E*, respectively. The symbols with the black edges show the averages of the classification accuracy rates of each dataset. The text labels at the top left and bottom right are the averages of the classification accuracy rates of all subjects for each method.

TABLE II
RESULTS OF THE CLASSIFICATION ACCURACY RATES [%] AVERAGED OVER THE SUBJECTS IN EACH DATASET

Method	DATASET					Ave.[1]
	A	B	C	D	E	
PCA + LDA	61.00	60.95	61.15	58.50	63.91	61.05
sLDA	61.28	61.11	60.92	57.53	63.28	60.86
CSP + LDA	59.84	60.09	59.93	55.78	65.53	60.07
xDAWN + LDA	61.23	61.26	61.30	56.66	67.35	61.42
UMLDA	56.69	55.86	56.12	54.33	64.63	57.14
SMLDAc	64.41	64.43	64.36	62.36	67.17	64.48
SMLDAm	64.54	64.72	64.61	61.60	70.70	65.03

[1]This column shows the averages over all subjects in all datasets.

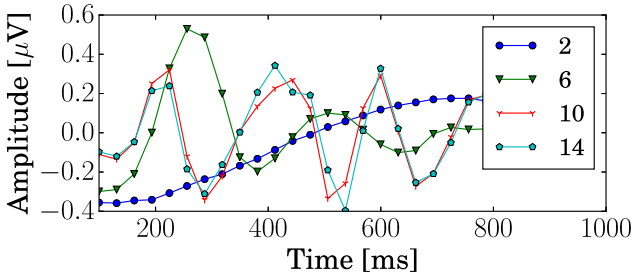
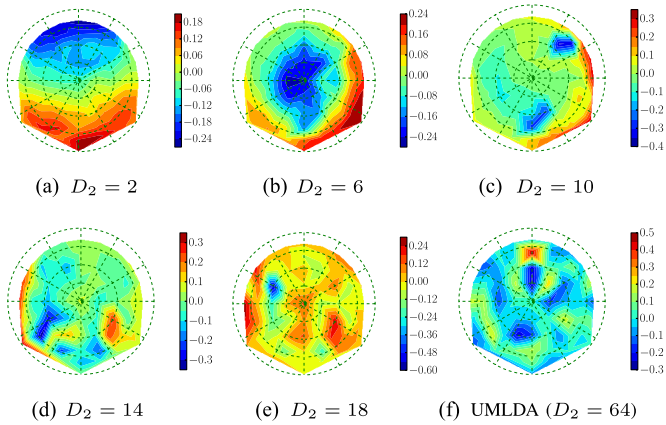


Fig. 9. Representatives of the SMLDAm projections in the temporal mode.

than those for the other time ranges. This feature corresponds to the difference observed in the grand-averaged waveforms in Fig. 3. This suggests that the projection in the temporal domain captures the temporal features.

Fig. 10 shows the projection vectors for the second (spatial) mode in the first EMP ($\boldsymbol{u}_2^{(1)}$) of SMLDAm. In Fig. 10(b), the



(a) $D_2 = 2$ (b) $D_2 = 6$ (c) $D_2 = 10$

(d) $D_2 = 14$ (e) $D_2 = 18$ (f) UMLDA ($D_2 = 64$)

Fig. 10. Representatives of the SMLDAm projections in the spatial mode.

coefficients for the projection have large amplitudes for the channels located in the parietal area of the brain. This suggests that the ERPs are extracted by the projection with the subspace constraints, because the major ERPs obtained in the oddball paradigms for BCIs, such as the P300 and N200 components, are distributed in the parietal cortex [46].

VII. CONCLUSION

We developed a supervised dimensionality reduction method with UMLDA and the subspace constraints for the parameter spaces. For the subspace constraints, we proposed a subspace design method using the functional connectivity estimated with the complementary data recorded in the experimental subjects/sessions. Our results have shown that the proposed method could improve the classification accuracy of ERP-based

BCIs. The proposed subspace constraints may also be applied to other neurotechnology classification problems. As an example, they could be applied to linear classifiers and projections, such as the CSP method [47] that has been widely used in BCI paradigms based on motor imagery [3], [16]. Moreover, SMLDA can be applied to tensors with more than three modes (e.g., the time–frequency spectrum obtained from a wavelet transform [48]). Additionally, the proposed method can be combined with spatial filters, such as the CSP and xDAWN, which return the filtered signals in a matrix form.

One of the limitations of SMLDA is that the number of parameters is large, because it has the parameters of the reduced dimension R and the dimensions of the subspaces for each mode $\{D_n\}_{n=1}^{N}$. However, the results of our experiment suggest that the appropriate parameters can be found by a cross-validation. Moreover, although we assumed that the locations of the electrodes were the same for all subjects and the resulting functional connectivities among the electrodes, the true locations actually differed among the subjects. For this problem, methods for obtaining functional connectivity that do not depend on the precise positions, such as [49], shall be applied.

REFERENCES

[1] G. Muller-Putz *et al.*, "Towards noninvasive hybrid brain–computer interfaces: Framework, practice, clinical application, and beyond," *Proc. IEEE*, vol. 103, no. 6, pp. 926–943, Jun. 2015.

[2] B. He, S. Gao, H. Yuan, and J. Wolpaw, "Brain–computer interfaces," in *Neural Engineering*, B. He, Ed. New York, NY, USA: Springer, 2013, pp. 87–151.

[3] B. He, B. Baxter, B. J. Edelman, C. C. Cline, and W. W. Ye, "Noninvasive brain-computer interfaces based on sensorimotor rhythms," *Proc. IEEE*, vol. 103, no. 6, pp. 907–925, Jun. 2015.

[4] G. Dornhege, J. D. R. Millan, T. Hinterberger, D. McFarland, and K.-R. Muller, Eds., *Toward Brain-Computer Interfacing*. Cambridge, MA, USA: MIT Press, 2007.

[5] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clin. Neurophysiol.*, vol. 113, no. 6, pp. 767–791, 2002.

[6] K. Squires, S. Petuchowski, C. Wickens, and E. Donchin, "The effects of stimulus sequence on event related potentials: A comparison of visual and auditory sequences," *Perception Psychophysics*, vol. 22, no. 1, pp. 31–40, Jan. 1977.

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

[8] D. J. Krusienski *et al.*, "A comparison of classification techniques for the P300 Speller," *J. Neural Eng.*, vol. 3, no. 4, 2006, Art. no. 299.

[9] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K.-R. Müller, "Single-trial analysis and classification of ERP components—A tutorial," *NeuroImage*, vol. 56, no. 2, pp. 814–25, May 2011.

[10] S. Fazli, S. Dahne, W. Samek, F. Bieszmann, and K.-R. Muller, "Learning from more than one data source: Data fusion techniques for sensorimotor rhythm-based brain–computer interfaces," *Proc. IEEE*, vol. 103, no. 6, pp. 891–906, Jun. 2015.

[11] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[12] B. Blankertz, M. Kawanabe, R. Tomioka, F. Hohlefeld, V. Nikulin, and K. R. Müller, "Invariant common spatial patterns: Alleviating nonstationarities in brain-computer interfacing," in *Proc. Adv. Neural Inf. Process. Syst.*, 2008, vol. 20, pp. 113–120.

[13] F. Lotte, "Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain–computer interfaces," *Proc. IEEE*, vol. 103, no. 6, pp. 871–890, Jun. 2015.

[14] 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 Trans. Biomed. Eng.*, vol. 57, no. 12, pp. 2936–2946, Dec. 2010.

[15] H. Higashi and T. Tanaka, "Regularization using similarities of signals observed in nearby sensors for feature extraction of brain signals," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2013, pp. 7420–7423.

[16] H. Higashi, T. Tanaka, and Y. Tanaka, "Smoothing of spatial filter by graph Fourier transform for EEG signals," in *Proc. Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf.*, 2014, pp. 1–8.

[17] W. Samek, M. Kawanabe, and K.-R. Müller, "Divergence-based framework for common spatial patterns algorithms," *IEEE Rev. Biomed. Eng.*, vol. 7, pp. 50–72, 2013.

[18] H. Morioka *et al.*, "Learning a common dictionary for subject-transfer decoding with resting calibration," *NeuroImage*, vol. 111, pp. 167–178, May 2015.

[19] R. Tomioka and K.-R. Müller, "A regularized discriminative framework for EEG analysis with application to brain-computer interface," *NeuroImage*, vol. 49, no. 1, pp. 415–432, 2010.

[20] A. Cichocki "Tensor decompositions for signal processing applications: From two-way to multiway component analysis," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 145–163, Mar. 2015.

[21] F. Cong, Q.-H. Lin, L.-D. Kuang, X.-F. Gong, P. Astikainen, and T. Ristaniemi, "Tensor decomposition of EEG signals: A brief review," *J. Neurosci. Methods*, vol. 248, pp. 59–69, Jun. 2015.

[22] Y. Liu, Q. Zhao, and L. Zhang, "Uncorrelated multiway discriminant analysis for motor imagery EEG classification," *Int. J. Neural Syst.*, vol. 25, no. 4, Jun. 2015, Art. no. 1550013.

[23] Q. Li and D. Schonfeld, "Multilinear discriminant analysis for higher-order tensor data classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 12, pp. 2524–2537, Dec. 2014.

[24] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "MPCA: Multilinear principal component analysis of tensor objects," *IEEE Trans. Neural Netw.*, vol. 19, no. 1, pp. 18–39, Jan. 2008.

[25] H. Higashi and T. Tanaka, "Simultaneous design of FIR filter banks and spatial patterns for EEG signal classification," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 4, pp. 1100–1110, Apr. 2013.

[26] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Uncorrelated multilinear discriminant analysis with regularization and aggregation for tensor object recognition," *IEEE Trans. Neural Netw.*, vol. 20, no. 1, pp. 103–123, Jan. 2009.

[27] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "A survey of multilinear subspace learning for tensor data," *Pattern Recognit.*, vol. 44, no. 7, pp. 1540–1551, Jul 2011.

[28] E. Bullmore and O. Sporns, "Complex brain networks: Graph theoretical analysis of structural and functional systems," *Nature Rev. Neurosci.*, vol. 10, no. 3, pp. 186–198, Mar. 2009.

[29] M. Hassan, O. Dufor, I. Merlet, C. Berrou, and F. Wendling, "EEG source connectivity analysis: From dense array recordings to brain networks," *PLoS One*, vol. 9, no. 8, Aug. 2014, Art. no. e105041.

[30] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 2, pp. 355–362, Feb. 2011.

[31] J. H. Friedman, "Regularized discriminant analysis," *J. Amer. Statist. Assoc.*, vol. 84, no. 405, pp. 165–175, Mar. 1989.

[32] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *J. Multivariate Anal.*, vol. 88, no. 2, pp. 365–411, Feb. 2004.

[33] J. Schäfer and K. Strimmer, "A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics," *Statist. Appl. Genetics Molecular Biol.*, vol. 4, no. 1, pp. 1544–6115, Jan. 2005.

[34] C. Vidaurre, N. Krämer, B. Blankertz, and A. Schlögl, "Time domain parameters as a feature for EEG-based brain computer interfaces," *Neural Netw.*, vol. 22, no. 9, pp. 1313–1319, Nov. 2009.

[35] Z. Jin, J.-Y. Yang, Z.-S. Hu, and Z. Lou, "Face recognition based on the uncorrelated discriminant transformation," *Pattern Recognit.*, vol. 34, no. 7, pp. 1405–1416, Jan. 2001.

[36] B. P. Rogers, V. L. Morgan, A. T. Newton, and J. C. Gore, "Assessing functional connectivity in the human brain by fMRI," *Magn. Resonance Imag.*, vol. 25, no. 10, pp. 1347–1357, Dec. 2007.

[37] J.-M. Schoffelen and J. Gross, "Source connectivity analysis with MEG and EEG," *Human Brain Mapping*, vol. 30, no. 6, pp. 1857–1865, Jun. 2009.

[38] 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 Process. Mag.*, vol. 30, no. 3, pp. 83–98, May 2013.

[39] B. Rivet, A. Souloumiac, V. Attina, and G. Gibert, "xDAWN algorithm to enhance evoked potentials: Application to brain–computer interface," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 8, pp. 2035–2043, Aug. 2009.

[40] M. Chang *et al.*, "Comparison of P300 responses in auditory, visual and audiovisual spatial speller BCI paradigms," in *Proc. 5th Int. Brain-Comput. Interface Meeting*, 2013, Art. no. 156 (2 pages).
[41] M. Kutas, G. McCarthy, and E. Donchin, "Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time," *Science*, vol. 197, no. 4305, pp. 792–795, Aug. 1977.
[42] C. Guger *et al.*, "How many people are able to control a P300-based brain-computer interface (BCI)?" *Neurosci. Lett.*, vol. 462, no. 1, pp. 94–98, Oct. 2009.
[43] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, 3rd ed. New York, NY, USA: Taylor & Francis, 2014.
[44] C.-G. Bénar *et al.*, "Single-trial analysis of oddball event-related potentials in simultaneous EEG-fMRI," *Human Brain Mapping*, vol. 28, no. 7, pp. 602–613, 2007.
[45] R. Näätänen, P. Paavilainen, T. Rinne, and K. Alho, "The mismatch negativity (MMN) in basic research of central auditory processing: A review," *Clin. Neurophysiol.*, vol. 118, no. 12, pp. 2544–2590, Dec. 2007.
[46] J. Polich, "Updating P300: An integrative theory of P3a and P3b," *Clin. Neurophysiol.*, vol. 118, no. 10, pp. 2128–2148, Oct. 2007.
[47] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 441–446, Dec. 2000.
[48] J. Li, L. Zhang, D. Tao, H. Sun, and Q. Zhao, "A prior neurophysiologic knowledge free tensor-based scheme for single trial EEG classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 17, no. 2, pp. 107–115, Apr. 2009.
[49] K. Morishige, T. Yoshioka, D. Kawawaki, N. Hiroe, M. Sato, and M. Kawato, "Estimation of hyper-parameters for a hierarchical model of combined cortical and extra-brain current sources in the MEG inverse problem," *NeuroImage*, vol. 101, pp. 320–336, Nov. 2014.

Hiroshi Higashi (S'10–M'14) received the B.E., M.E., and Ph.D. degrees from Tokyo University of Agriculture and Technology, Tokyo, Japan, in 2009, 2011, and 2013, respectively. From 2011 to 2012, he was a Junior Research Associate with the Laboratory for Advanced Brain Signal Processing, Brain Science Institute, RIKEN, Saitama, Japan. From 2012 to 2014, he was a Research Fellow of the Japan Society for the Promotion of Science. He is currently an Assistant Professor in the Department of Computer Science and Engineering, Toyohashi University of Technology, Aichi, Japan, and a Visiting Researcher in RIKEN Brain Science Institute. His research interests include brain and biomedical signal processing.

Tomasz M. Rutkowski (M'08–SM'12) received the M.Sc. degree in electronics and Ph.D. degree in telecommunications and acoustics from Wroclaw University of Technology, Wroclaw, Poland, in 1994 and 2002, respectively. He received the Postdoctoral training at the Multimedia Laboratory, Kyoto University, and during 2005–2010, he worked as a Research Scientist at RIKEN Brain Science Institute, Japan. During 2011–2016, he served as an Assistant Professor at the University of Tsukuba and as a Visiting Scientist at RIKEN Brain Science Institute. He is currently a Research Fellow at the University of Tokyo and a Visiting Professor at Saitama Institute of Technology, Fukaya, Japan. His research interests include computational neuroscience, especially brain–computer interfacing technologies and sleep studies, computational modeling of brain processes, neurobiological signal and information processing, multimedia interfaces, and intermedia technology design. He received The Annual BCI Research Award 2014 for the best brain–computer interface project and the same award nomination in 2016. He is a Senior Member of the Asia-Pacific Signal and Information Processing Association, where he serves as a BioSiPS Technical Committee Chair. He is a Member of the Editorial Board of Frontiers in Fractal Physiology and serves as a Reviewer for IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: PART B, *Cognitive Neurodynamics*, and the *Journal of Neural Engineering*, *PLOS One*, *Nature Scientific Reports*, *Journal of Neuroscience Methods*, etc.

Toshihisa Tanaka (S'98–M'02–SM'10) received the B.E., the M.E., and the Ph.D. degrees from the Tokyo Institute of Technology, Meguro, Japan, in 1997, 2000, and 2002, respectively. From 2000 to 2002, he was a JSPS Research Fellow. From October 2002 to March 2004, he was a Research Scientist at RIKEN Brain Science Institute. In April 2004, he joined the Department of Electrical and Electronic Engineering, Tokyo University of Agriculture and Technology, where he is currently an Associate Professor. In 2005, he was a Royal Society Visiting Fellow at the Communications and Signal Processing Group, Imperial College London, U.K. From June 2011 to October 2011, he was a Visiting Faculty Member in the Department of Electrical Engineering, University of Hawaii at Manoa.

His research interests include a broad area of signal processing and machine learning including brain and biomedical signal processing, brain–machine interfaces and adaptive systems. He is a coeditor of *Signal Processing Techniques for Knowledge Extraction and Information Fusion* (with Mandic, New York, NY, USA: Springer, 2008).

He served as an Associate Editor and a Guest Editor of special issues in journals including *Neurocomputing* and *IEICE Transactions on Fundamentals*. He currently serves as an Associate Editor of IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, *Computational Intelligence and Neuroscience* (Hindawi), and *Advances in Data Science and Adaptive Analysis* (World Scientific). Furthermore, he serves as a Member-at-Large and Board of Governors of Asia-Pacific Signal and Information Processing Association (APSIPA). He was a Chair of the Technical Committee on Biomedical Signal Processing, APSIPA. He is a Member of the IEICE, APSIPA, and Society for Neuroscience.

Yuichi Tanaka (S'06–M'07) received the B.E., M.E., and Ph.D. degrees in electrical engineering from Keio University, Yokohama, Japan, in 2003, 2005, and 2007, respectively. He was a Postdoctoral Scholar at Keio University, Yokohama, Japan, from 2007 to 2008, and supported by the Japan Society for the Promotion of Science. From 2006 to 2008, he was also a Visiting Scholar at the University of California, San Diego. From 2008 to 2012, he was an Assistant Professor in the Department of Information Science, Utsunomiya University, Tochigi, Japan. Since 2012, he has been an Associate Professor in the Graduate School of BASE, Tokyo University of Agriculture and Technology, Tokyo, Japan. His current research interests include the field of multidimensional signal processing which includes: graph signal processing, image and video processing with computer vision techniques, distributed video coding, objective quality metric, and effective spatial-frequency transform design.

Dr. Tanaka has been an Associate Editor of IEEE TRANSACTIONS ON SIGNAL PROCESSING since 2016 and also *IEICE Transactions on Fundamentals* since 2013. He is currently an Elected Member of the APSIPA Image, Video and Multimedia Technical Committee. He received the Yasujiro Niwa Outstanding Paper Award in 2010, the TELECOM System Technology Award in 2011, and Ando Incentive Prize for the Study of Electronics in 2015. He also received Best Paper Awards in APSIPA ASC 2014 and 2015.