

Received March 12, 2021, accepted March 15, 2021, date of publication March 19, 2021, date of current version March 30, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3067337

# Semi-Supervised Learning for Auditory Event-Related Potential-Based Brain–Computer Interface

MIKITO OGINO<sup>1,2</sup>, SUGURU KANOGA<sup>3</sup>, SHIN-ICHI ITO<sup>4</sup>, (Member, IEEE), AND YASUE MITSUKURA<sup>1</sup>, (Member, IEEE)

<sup>1</sup>School of Integrated Design Engineering, Keio University, Kanagawa 223-8522, Japan

<sup>2</sup>Dentsu ScienceJam Inc., Tokyo 107-0052, Japan

<sup>3</sup>National Institute of Advanced Industrial Science and Technology (AIST), Tokyo 135-0064, Japan

<sup>4</sup>Graduate School of Technology, Tokushima University, Tokushima 770-8501, Japan

Corresponding author: Yasue Mitsukura (mitsukura@keio.jp)

This work was supported in part by the Keio Leading-edge Laboratory (KLL) 2020 Ph.D. Program Research Grant, and in part by Japan Society for the Promotion of Science (JSPS) KAKENHI under Grant 20K19854.

**ABSTRACT** A brain–computer interface (BCI) is a communication tool that analyzes neural activity and relays the translated commands to carry out actions. In recent years, semi-supervised learning (SSL) has attracted attention for visual event-related potential (ERP)-based BCIs and motor-imagery BCIs as an effective technique that can adapt to the variations in patterns among subjects and trials. The applications of the SSL techniques are expected to improve the performance of auditory ERP-based BCIs as well. However, there is no conclusive evidence supporting the positive effect of SSL techniques on auditory ERP-based BCIs. If the positive effect could be verified, it will be helpful for the BCI community. In this study, we assessed the effects of SSL techniques on two public auditory BCI datasets—AMUSE and PASS2D—using the following machine learning algorithms: step-wise linear discriminant analysis, shrinkage linear discriminant analysis, spatial temporal discriminant analysis, and least-squares support vector machine. These backbone classifiers were firstly trained by labeled data and incrementally updated by unlabeled data in every trial of testing data based on SSL approach. Although a few data of the datasets were negatively affected, most data were apparently improved by SSL in all cases. The overall accuracy was logarithmically increased with every additional unlabeled data. This study supports the positive effect of SSL techniques and encourages future researchers to apply them to auditory ERP-based BCIs.

**INDEX TERMS** Auditory stimuli, brain–computer interface, event-related potential, semi-supervised learning, P300.

## I. INTRODUCTION

As a brain-computer interface (BCI) performs communication based on neural activity measurement. It is a useful tool for patients with paralysis who find it difficult to express their feelings and thoughts via body movements [1], [2]. Among devices such as functional near-infrared spectroscopy (fNIRS), functional magnetic resonance imaging (fMRI), and electroencephalogram (EEG) used to measure neuronal activities, EEGs have attracted considerable research attention due to their noninvasive monitoring potential with a high

temporal resolution for capturing information of the neuronal responses [3], [4]. When a patient with paralysis operates an EEG-based BCI, a paradigm evoking event-related potential (ERP) is commonly used [5]–[7]. An ERP is evoked by a visual, auditory, or tactile rare stimulus and contains characteristic positive/negative amplitude peaks with their latencies (e.g., P100, N100, P200, N200, and P300) appearing a few hundred milliseconds after the stimulus. A P300 speller, which mainly uses the temporal features of positive amplitude peaks appearing approximately 300 ms after the aforementioned stimuli is presented, is regarded as the most well-known paradigm in BCI community [8], [9]. In order to use such ERP-based BCI regardless of the

The associate editor coordinating the review of this manuscript and approving it for publication was Tao Zhou<sup>1</sup>.

environment, combining them with wearable devices is a hot research topic for BCI researchers [10]. Some studies developed portable ERP-based BCIs through the use of wearable EEG devices [11]–[13]. These developments make routine use of ERP-based BCIs possible for not only patients with paralysis but also for healthy people.

Classifying ERPs is predicated on machine learning (ML) methods such as linear discriminant analysis (LDA) [14] and support vector machine (SVM) [15]. To classify them, the ML methods consider the temporal features including peak amplitudes and latencies among trials, and spatial features over the channels [16]. Human-specific parameters such as fatigue and concentration change during the operation of the BCI as well as on a day-to-day basis, and the characteristics of the ERP change accordingly. This is one of the central challenges of the ML methods. To deal with this challenge, an increasing number of studies have been conducted on robust universal feature extraction methods such as deep learning [17], [18]. However, if the parameters of the ML-based classifier are fixed after the initial training, it is not possible to mitigate the effects of these occasional changes. This leads to a sudden accuracy drop and results in inhibition of the routine use of ERP-based BCIs.

Semi-supervised learning (SSL) is an effective technique for adapting ML-based trained classifiers to varying patterns owing to the use of unlabeled new data. It has recently been applied to various biological signal-based applications [19]–[21] including visual ERP-based and motor imagery (MI)-based BCIs [22]–[25]. In the visual ERP- and MI-based BCIs, SSL has indicated its efficiency of adapting pre-trained LDA and SVM families: SVM [22], least squares SVM (LS-SVM) [23], spectral regression kernel discriminant analysis (SRKDA) [24], and shrinkage regularized based LDA (SKLDA) [25]. In most cases, an incremental update strategy is naively applied every time a single-trial unlabeled data is recorded. Although SSL has recently been combined with a deep learning method in an MI-based BCI [26], LDA and SVM series are still used/applied in the mainstream. BCI operation protocol generally consists of pre-training phase in which the labeled data is used for training a classifier, and testing phase in which the trained classifier infers the class of unlabeled data for controlling the system and storing the unlabeled data. In SSL, the trained classifier is updated by unlabeled data in the testing phase to tune the parameters optimally, combining older and new data. Thus, the technique can mitigate the effect of changes on human-specific parameters during BCI operation even with a shortened pre-training phase [22], [25].

In this study, we investigate the adaptability of SSL to auditory ERP-based BCIs. In patients with amyotrophic lateral sclerosis (ALS), the final stage of the disease involves the loss of eye-related functions (e.g., manipulation of eyes and opening and closing of the eyelids) [27]. Therefore, auditory paradigms have been gaining attention to meet the needs of BCI for a wider range of patient groups [28], [29]. However, to the best of our knowledge, previously proposed

SSL techniques have not been applied to auditory ERP-based BCIs. In visual and auditory ERP-based BCI, the characteristic peaks of ERPs represented by N200 and P300 differ between auditory and visual stimuli [30], [31]. Moreover, the accuracy of the auditory ERP-based BCI is less than that of the visual one because of its weaker ERP evoked by the auditory stimulus [32]. These differences may cause a negative effect for the application of SSL techniques to auditory ERP-based BCIs. In contrast, if the positive effect could be verified in the auditory ERP-based BCIs, it will be helpful for BCI communities.

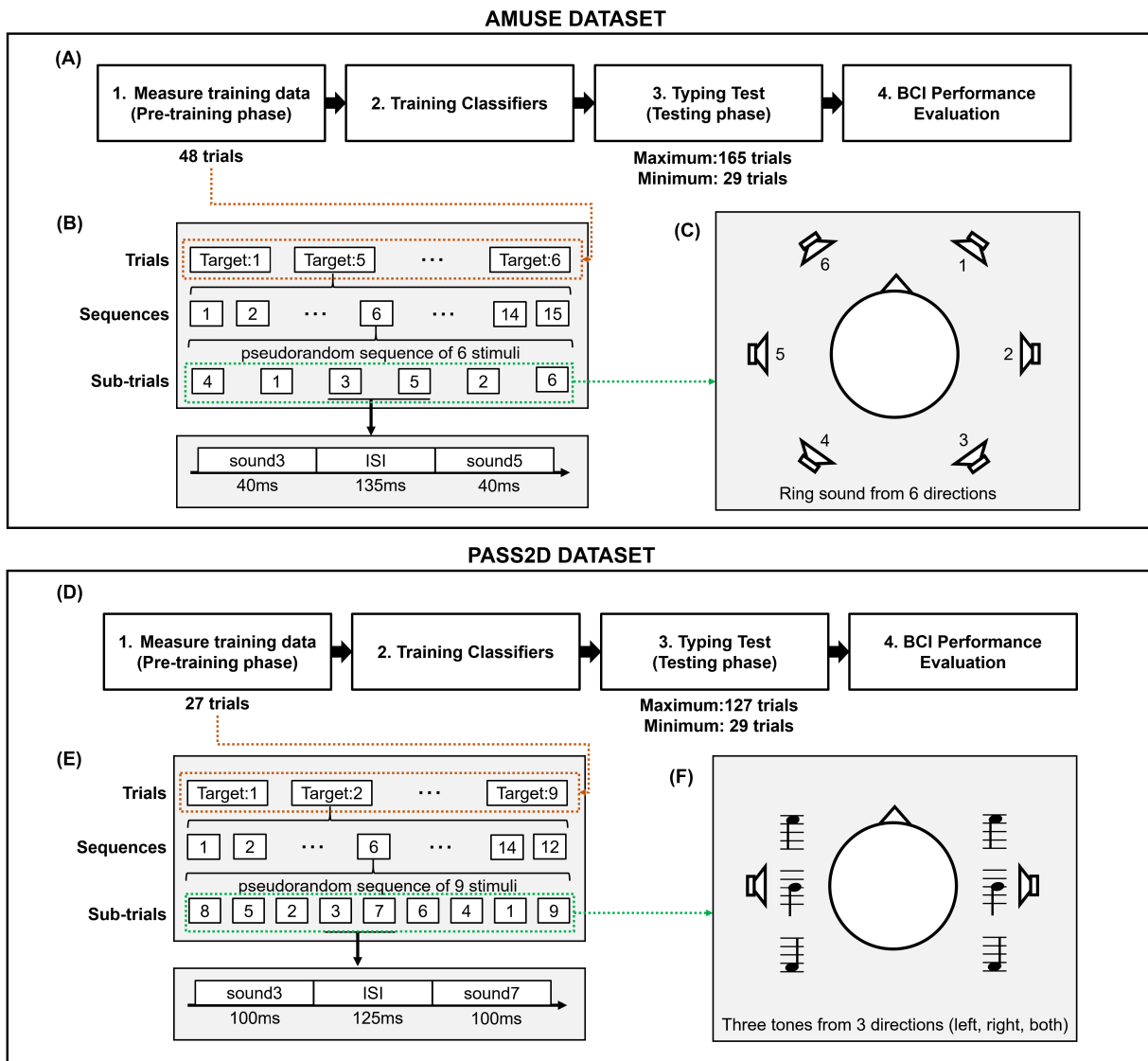
To evaluate the effect of SSL for auditory ERP-based BCI, we used multiple public datasets (AMUSE and PASS2D) and traditional ML algorithms for ERP-based BCIs as backbone classifiers: step-wise LDA (SWLDA) [33] and LS-SVM [23]. In addition, two advanced ML algorithms named SKLDA [34] and spatial-temporal discriminant analysis (STDA) [35] were also investigated as backbone classifiers. Employing the two datasets and four ML algorithms, we evaluate the effect of SSL for auditory ERP-based BCI. We expect that the obtained results will encourage future studies related to auditory ERP-based BCIs to use SSL techniques for their paradigms.

This paper is organized as follows. Section II details the datasets, preprocessing methods, classification algorithms including SSL techniques, and evaluation metrics. Section III presents the results and discussions of the experiments; the supervised learning techniques and SSL techniques using four ML algorithms are compared. Section IV presents the conclusions of our study.

## II. MATERIALS AND METHODS

Public datasets enable us to compare the obtained results with those of previous studies. For this study, we used two public auditory ERP-based BCI datasets: auditory multiclass spatial ERP (AMUSE) [36], [37] and predictive auditory spatial speller with two-dimensional cues (PASS2D) [38]. These datasets were collected by Neurotechnology Group, Technische Universität Berlin, Germany (available from BNCI Horizon 2020 database [39]) and they have been used in many studies [40], [41]. Therefore, we utilize them for comparing our results with those of other studies, enabling future researches to compare their results with ours.

The presented auditory stimuli must be easily recognizable by all users with minimal or no loss of cognitive function. The elements of auditory stimulus are broadly divided into two factors: (1) timbre and (2) sound direction. Several studies proposed methods for differentiating recognizable timbres using synthetic beeps [37], natural speech [42], and natural sounds (i.e., ducks, birds, and frogs) [33]. Furthermore, some other studies proposed methods to differentiate the directionality of recognizable sounds by physically placing loudspeakers evenly around the user or by varying the interaural level difference or interaural time difference when using headphones [36], [38]. AMUSE and PASS2D datasets present paradigms of auditory stimuli that can recognize both



**FIGURE 1.** (A) Experimental flow of AMUSE dataset. Forty eight trials were conducted in pre-training phase. After training classifiers, 29 to 165 trials were conducted in testing phase. (B) Stimulus selection flow of trials. Each trial consisted of 15 sequences, and each sequence consisted of 6 sub-trials. Sound duration was 40 ms, and ISI was 135 ms. (C) Explanation of stimulus sounds. Six ring sounds were presented from each six directions. (D) Experimental flow of PASS2D dataset. Twenty seven trials were conducted in pre-training phase. After training classifiers, 29 to 127 trials were conducted in testing phase. (E) Stimulus selection flow of trials. Each trial consisted of 12 sequences, and each sequence consisted of 9 sub-trials. Sound duration was 100 ms, and ISI was 125 ms. (F) Explanation of stimulus sounds. Three different tones were presented from each of the three directions (left, right, and both sides).

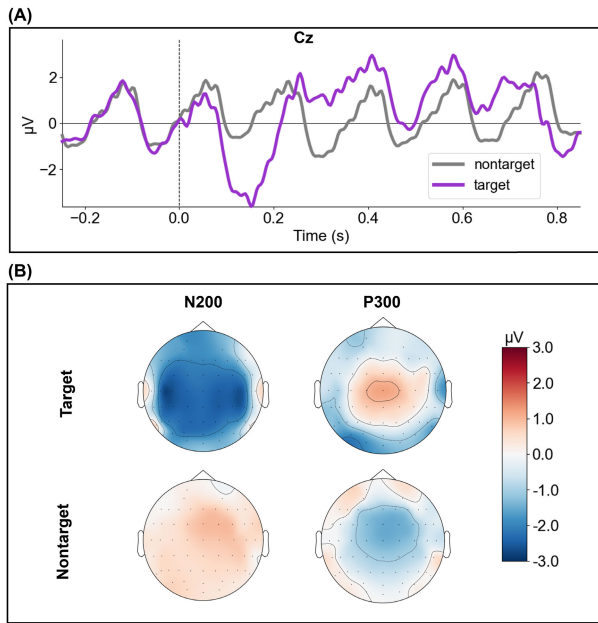
differences based on combinations of the aforementioned methods. Both datasets are provided in the binary data container formats used in MATLAB (.mat).

**A. AMUSE DATASET**

The paradigm for the AMUSE dataset was proposed by Schreuder *et al.* [36] and published by Höhne *et al.* [37] as a dataset with 21 healthy subjects. EEG data were recorded with 60 Ag/AgCl electrodes based on the placement of the International 10-20 System. Electrooculogram (EOG) data were recorded with two bipolar channels over the eyes; the reference channel was placed on the nose. These data were

amplified using a Brain Products amplifier (Brain Products Co., Munich, Germany) with a sampling rate of 1000 Hz, and filtered through an analog bandpass filter (between 0.1 and 250 Hz; the publicly available data are distributed in a downsampled form to 250 Hz).

The experimental design of the AMUSE dataset is shown in Fig. 1 (A)–(C). Firstly, training data for constructing classifiers were measured in pre-training phase. The pre-training phase consists of 48 trials. Secondly, classifiers were trained to classify target and nontarget ERP features, as shown in Fig. 2. Thirdly, subjects typed digital letters as they liked using the trained classifier and the BCI system in



**FIGURE 2.** (A) Ground average of ERPs on Cz channel from subject “fcb” of AMUSE dataset. Purple line denotes ERP when target sound presented. Gray line denotes ERP when nontarget sound presented. (B) Topological map of primary features (N200 and P300) of target and nontarget ERPs of subject “fcb” of AMUSE dataset.

testing phase. The trained classifiers inferred which letter the subject wanted to select. The minimum and maximum number of trials in the testing phase were 29 and 165, respectively. Lastly, the BCI performance was evaluated. The trials were composed of sequences and sub-trials shown in Fig. 1 (B). The target direction was visually and aurally instructed prior to each trial. The number of sub-trials in each sequence was equal to the kinds of sound stimuli. Discriminable synthetic beep stimuli were randomly presented from six audio speakers evenly spaced around the subject (i.e., audio speakers were arranged every  $60^\circ$ ) as shown in Fig. 1 (C). Based on the idea of a classic oddball paradigm, any of the six directions could be a target (probability 16.7%), which leaves the others as nontargets (probability 83.3%). Each stimulus was presented for 40 ms. The inter-stimulus interval (ISI) was 135 ms. Furthermore, the stimulus onset asynchrony (SOA) or inter-stimulus onset interval, which is the sum of the stimulus presentation time and interval time for sound stimuli, was 175 ms.

## B. PASS2D DATASET

The paradigm for this dataset was proposed and published by Höhne *et al.* [38] as a dataset with 12 healthy subjects. EEG data were recorded using a Fast’n Easy Cap (Easy-Cap GmbH, Inning, Germany) with 63 Ag/AgCl electrodes based on the placement of the International 10-20 System. The reference channel was placed on the nose. These data were amplified using a Brain Products amplifier (Brain Products Co., Munich, Germany) with a sampling rate of 1000 Hz and filtered through an analog bandpass filter

(between 0.1 and 250 Hz; the publicly available data are distributed in a downsampled form to 250 Hz.).

The experimental design of the PASS2D dataset is shown in Fig. 1 (D)–(F). Similar to AMUSE dataset, the flow consists of pre-training phase, training the classifiers, testing phase, and BCI performance evaluation. The pre-training phase consisted of 27 trials. In testing phase, subjects typed virtual pre-defined letters until all selections were correctly performed. The minimum and maximum number of trials in the testing phase were 29 and 127, respectively. The trials were composed of sequences and sub-trials shown in Fig. 1 (E). The target direction was visually and aurally instructed prior to each trial. The number of sub-trials in each sequence was equal to the kinds of sound stimuli. Discriminable synthetic beep stimuli were randomly presented from headphones and were generated by three kinds of pitch (high/medium/low) and location (left/middle/right). Based on the idea of a classic oddball paradigm, any of the three pitches and directions could be a target (probability 11.1%), leaving the others as nontargets (probability 88.9%). The current target cue was presented to the subject three times before starting each pre-training recording trial, and the corresponding number on a  $3 \times 3$  grid was further highlighted on the screen. Each stimulus was presented for 100 ms. The ISI and SOA were 125 ms and 225 ms, respectively.

## C. PREPROCESSING

To compare the effects of SSL techniques for traditional and advanced ML algorithms, we performed the same preprocessing approach on the two datasets.

Each EEG signal was low-pass filtered with 40 Hz using a fourth-order Butterworth filter. Furthermore, the filtered signal was segmented into epochs from -150 ms to 800 ms based on the stimulus onset. If the peak-to-peak voltage difference of any channel was greater than  $100 \mu\text{V}$ , the epoch was considered contaminated with some artifacts (e.g., motion and ocular artifacts), and it was discarded [38]. Then, the remaining epochs were downsampled to 50 Hz. Finally, we performed baseline correction using the average values of the first 150 ms data. Since the amplitude values can be used as temporal features in ERP analysis, baseline corrected 800 ms epoch (40 time points) after the onset were flattened in the channel direction and used as a feature vector  $\mathbf{x} \in \mathbb{R}^D$ , where  $D$  denotes the dimensionality of the features (the number of channels and time points).

## D. CLASSIFICATION

In this study, for all classification algorithms, a binary classifier to determine whether an auditory stimulus was a target or nontarget was generated based on a previous study [37]. The typical target and nontarget ERPs are described in Fig. 2. There are differences of peak amplitudes, the latencies (Fig. 2 (A)), and spatial patterns over the channels (Fig. 2 (B)) between target and nontarget ERPs. The binary classifiers were trained to recognize such differences. It should be noted

that the binary classification problem swaps the definitions of class labels, depending on the target sounds. If there are  $N_s$  types of target sounds,  $N_s$  binary classifiers will be trained ( $k = 1, \dots, N_s$ ). In the AMUSE and PASS2D datasets, six and nine types of target sounds were presented, respectively. Thus, six binary classifiers for the AMUSE ( $N_s = 6$ ) and nine binary classifiers for the PASS2D ( $N_s = 9$ ) were trained.

The feature vectors in the pre-training phase  $\mathbf{X}_{tr} = \{\mathbf{x}_i\}_{i=1}^{N_L} \in \mathbb{R}^{D \times N_L}$ , where  $N_L$  indicates the number of all pre-training trials (4,320 or 2,916), were used to train the classifiers. We did not emphasize the ERP waveform of training epochs by an averaging process. To evaluate the performance of trained classifiers, we used the averaged feature vectors over sequences in the testing phase  $\mathbf{X}_{te} = \{\mathbf{x}_j\}_{j=1}^{N_U} \in \mathbb{R}^{D \times N_U}$ , where  $N_U$  indicates the number of averaged featured vectors. In this case, an averaging process was applied to a group of data corresponding to the stimulus sound.

### 1) SWLDA

SWLDA is commonly used in auditory ERP-based BCI studies as a classifier, including for feature selection functions [33], [43]. The classification module of SWLDA is the same as that of Fisher LDA (FLDA), and it learns the decision boundaries for binary labeled (target and nontarget) training feature vectors  $\mathbf{X}_{tr}$ . In the dataset for the  $k$ -th auditory sound,  $N_{L_1}$  trials in class 1 are denoted by  $\mathcal{X}_{L_1}$ , and  $N_{L_2}$  trials in class 2 are denoted by  $\mathcal{X}_{L_2}$  ( $N_L = N_{L_1} + N_{L_2}$ ).

The FLDA finds the class separability in the direction  $\mathbf{w} \in \mathbb{R}^D$  by maximizing,

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}}. \quad (1)$$

where  $\mathbf{S}_B \in \mathbb{R}^{D \times D}$ ,  $\mathbf{S}_W \in \mathbb{R}^{D \times D}$ , and  $(\cdot)^T$  are the between-class and within-class scatter matrices, and the transpose operator, respectively. The between-class and within-class scatter matrices can be written as,

$$\mathbf{S}_B = \frac{1}{N_L} \sum_{c=1}^2 N_{L_c} (\bar{\mathbf{x}}_{L_c} - \bar{\mathbf{x}})(\bar{\mathbf{x}}_{L_c} - \bar{\mathbf{x}})^T, \quad (2)$$

$$\mathbf{S}_W = \frac{1}{N_L} \sum_{c=1}^2 \sum_{\mathbf{x} \in \mathcal{X}_{L_c}} (\mathbf{x} - \bar{\mathbf{x}}_{L_c})(\mathbf{x} - \bar{\mathbf{x}}_{L_c})^T, \quad (3)$$

where  $\bar{\mathbf{x}} \in \mathbb{R}^D$  is the overall mean of  $\mathbf{X}_{tr}$ , and  $\bar{\mathbf{x}}_{L_c} \in \mathbb{R}^D$  is the mean of  $\mathcal{X}_{L_c}$ .

In the testing phase, the likelihood  $p_{k,j}$  for the  $k$ -th auditory sound to the  $j$ -th testing data can be obtained by linearly multiplying the  $k$ -th learned weight vector  $\mathbf{w}_k$  by the averaged feature vector  $\mathbf{x}_j$  as,

$$p_{k,j} = \mathbf{w}_k^T \mathbf{x}_j. \quad (4)$$

The label corresponding to the sound stimulus that presented the highest likelihood is set to 1, and the others are set to 0 to determine the prediction vector  $\hat{\mathbf{y}}_j \in \mathbb{R}^{N_s}$  to the

$j$ -th testing data.

$$\hat{y}_{k,j} = \begin{cases} 1 & (p_{k,j} = \max(\mathbf{p}_j)) \\ 0 & (\text{otherwise}). \end{cases} \quad (5)$$

The feature selection module of SWLDA involves a step-wise approach that selects important feature variables automatically with a combination of forward and backward steps [44]. The  $p$ -values are calculated using least-squares regression of the feature variables. The feature variable set has a lower  $p$ -value compared to the input threshold ( $p$ -value  $< 0.10$ ), and it is pooled into the FLDA module. The unpooled feature variables are added to the feature set one by one, and the  $p$ -value is updated in the FLDA module. Each time the  $p$ -value is updated, the feature variable is removed from the variable set if the  $p$ -value is greater than the threshold value ( $p$ -value  $> 0.15$ ). This loop is repeated until the feature variables no longer meet the input or output criteria. Since ERPs have subject-specific shapes, the feature variables were selected for each subjects.

### 2) SKLDA

FLDA estimates between-class and within-class covariance matrices  $\mathbf{S}_b, \mathbf{S}_w$ . Generally, these matrices are not biased and have acceptable properties. However, when the training data is high-dimensional, the covariance matrix cannot be estimated accurately because the estimates of the maximum and minimum eigenvalues of the original covariance matrix deviate significantly.

In order to deal with this problem, SKLDA was proposed to estimate the covariance matrices more precisely through adjusting the extreme eigenvalues of the covariance matrix towards the average eigenvalue [16], [45]. To adjust the estimation error, the new covariance matrix  $\tilde{\mathbf{S}}$  is calculated as follows,

$$\tilde{\mathbf{S}}(\gamma) = (1 - \gamma)\mathbf{S} + \gamma v \mathbf{I}, \quad (6)$$

where  $\gamma$  is a tuning parameter  $\gamma \in [0, 1]$ ,  $v$  is an average eigenvalue of  $\mathbf{S}$ , and  $\mathbf{I}$  is the identity matrix. In this study, we used optimal tuning parameter calculated using the following equation [34],

$$\gamma^* = \frac{N_L}{(N_L - 1)^2} \frac{\sum_{i,j=1}^D \text{var}_n(z_{ij}(n))}{\sum_{i \neq j} s_{ij}^2 + \sum_i (s_{ii} - v)^2}, \quad (7)$$

where  $s_{ij}$  is the element in the  $i$ -th row and  $j$ -th column of covariance vector  $\mathbf{S}$  and  $z_{ij}(n)$  is computed as follows.

$$z_{ij}(n) = ((\mathbf{x}_n)_i - (\bar{\mathbf{x}})_i)((\mathbf{x}_n)_j - (\bar{\mathbf{x}})_j), \quad (8)$$

where  $(\mathbf{x}_n)_i$  and  $(\bar{\mathbf{x}})_i$  are the  $i$ -th element of the feature vector  $\mathbf{x}_n$  and mean vector  $\bar{\mathbf{x}}$ ,

To train a weight vector, SKLDA adopts vectorized (one-way) training samples with  $(M \times P)$  dimensions where  $M$  denotes the number of channels and  $P$  denotes the temporal points of each channel. Similar to that in LDA, the training data is high dimensional.

### 3) STDA

In traditional classifiers for auditory ERP-based BCI such as SWLDA and LS-SVM,  $D(= MP)$ -dimensional samples where each sample is the concatenation of  $P$  temporal features from each of  $M$  channels have been used as training and testing data. Therefore, the dimensionality tends to be very high.

STDA adopts spatial-temporal (two-way) training samples  $\mathbf{X}_i \in \mathbb{R}^{M \times P}$  ( $i = 1, 2, \dots, N$ ) to separately find the two projection matrices from spatial and temporal dimensions and effectively reduces the original feature dimension for discriminant analysis [35]. The spatial and temporal projection matrices are expressed as,  $\mathbf{W}_1 \in \mathbb{R}^{M \times L}$  and  $\mathbf{W}_2 \in \mathbb{R}^{P \times L}$  respectively. Here,  $L$  should be lower than  $M$  and  $P$ . The projection matrices are optimized using traditional LDA algorithm. After the projection matrices are optimized as  $\tilde{\mathbf{W}}_1$  and  $\tilde{\mathbf{W}}_2$ , a training data is transformed to a lower dimensional sample as:

$$\mathbf{f}_i = \text{vec}(\mathbf{X}_i \times_1 \tilde{\mathbf{W}}_1^T \times_2 \tilde{\mathbf{W}}_2^T), \quad (9)$$

where  $\mathbf{f}_i \in \mathbb{R}^{L^2}$  and  $\text{vec}(\cdot)$  denotes the vectorization operator.  $\times_1$  and  $\times_2$  denote multiplication of matrices in a particular dimension. The projected samples are used for training an LDA classifier.

### 4) SSL WITH SWLDA, SKLDA, AND STDA

In SSL, the backbone pre-trained SWLDA, SKLDA, and STDA update their weight vectors using the unlabeled data measured in testing phase. These updates are based on incremental update strategy that updates with each one-trial testing data. We refer to it as SSL in this study. Each testing dataset inferred as to which class it belongs to by the classifier trained until the previous trial. The estimated labels are not true labels but called pseudo-labels and used for re-training the classifiers. SSL uses training data and pseudo-labeled "original" testing data. Here, original testing data indicates that the data are not subjected to averaging process across trials.

The three LDA-based classifiers commonly use Eqs. (1), (4), and (5) for testing data. From Eq. (5), we can determine the stimulus-induced ERPs contained in the averaged testing data  $\mathbf{x}_j$ . We assign binary pseudo-labels (i.e., target and nontarget) to all original testing data based on the estimation result for the averaged testing data, merge a set of original testing data and pseudo labels into the original training data, and retrain the model. Not only the classification module but also the important feature variables obtained in the step-wise approach and projected samples in STDA projection approach are retrained by SSL.

### 5) LS-SVM

SVM constructs a hyperplane or a set of hyperplanes using the first layer of vectors closest to the plane, called support vectors. This classification algorithm is frequently used in BCI communities [15], [46]. Instead of solving

computationally intensive quadratic programming of traditional SVM, LS-SVM, which uses a set of linear equations for training, has been proposed [23], [47]. The optimization problem of LS-SVM is:

$$\begin{aligned} \min_{\mathbf{w}, b, \mathbf{e}} \quad & \|\mathbf{w}\|^2 + \gamma \|\mathbf{e}\|^2, \\ \text{s.t.} \quad & \mathbf{w}^T \varphi(\mathbf{x}_i) + b = y_i + e_i, \end{aligned} \quad (10)$$

where  $\varphi(\mathbf{x}_i)$  denotes a mapping function from the original space to a high-dimensional space,  $e_i$  represents an error for the  $i$ -th feature,  $\gamma$  denotes the regularization term ( $\gamma \geq 0$ ), and  $b$  represents the bias term. The labels here are the binaries of  $-1$  and  $1$  (i.e.,  $y_i \in \{-1, 1\}$ ).

Introducing a positive-definite kernel  $\mathbf{K} \in \mathbb{R}^{N_L \times N_L}$ , the problem in Eq. (10) can be represented as a set of linear equations:

$$\begin{bmatrix} \mathbf{K} + \gamma^{-1} \mathbf{I} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ b \end{bmatrix} = \begin{bmatrix} \mathbf{y} \\ 0 \end{bmatrix}, \quad (11)$$

where  $\mathbf{y} \in \mathbb{R}^{N_L}$  denotes the label vector; all elements in vector  $\mathbf{1} \in \mathbb{R}^{N_L}$  are 1.  $\mathbf{I} \in \mathbb{R}^{N_L \times N_L}$  denotes an identity matrix, and  $\mathbf{a} \in \mathbb{R}^{N_L}$  represents the dual variable vector.  $\mathbf{K}_{i'i'} = K(\mathbf{x}_i, \mathbf{x}_{i'}) = \varphi(\mathbf{x}_i)^T \varphi(\mathbf{x}_{i'})$  is the  $(i, i')$ -th element of  $\mathbf{K}$ . The optimal parameters for  $\mathbf{a}$  and  $b$  are calculated by,

$$\mathbf{a} = \mathbf{H}^{-1}(\mathbf{y} - b\mathbf{1}), \quad (12)$$

$$b = \mathbf{1}^T \mathbf{H}^{-1} \mathbf{y} (\mathbf{1}^T \mathbf{H}^{-1} \mathbf{1})^{-1}, \quad (13)$$

where the model matrix  $\mathbf{H} \in \mathbb{R}^{N_L \times N_L}$  is represented as,

$$\mathbf{H} = \mathbf{K} + \gamma^{-1} \mathbf{I}. \quad (14)$$

The decision function of LS-SVM can be expressed as,

$$f(\mathbf{x}_j) = \text{sgn}\left(\sum_{i=1}^N a_i y_i K(\mathbf{x}_j, \mathbf{x}_i) + b\right). \quad (15)$$

In this study, we used the linear kernel. LS-SVM was applied for each sound stimulus; thus,  $N_s$  decision functions were calculated and the auditory sound tied to the highest-value function was selected as the estimated target for the testing data  $\mathbf{x}_j$ .

### 6) SEMI-SUPERVISED LS-SVM

Semi-supervised SVM for P300 BCI was proposed by Li *et al.* [22]; Subsequently, Gu *et al.* [23] reduced the computational cost by introducing LS-SVM, and they developed the application by sequentially updating the self-training algorithm in an online P300 BCI called SUST-LSSVM. By adding new unlabeled feature vectors in the testing phase  $\{\mathbf{x}_j\}_{j=1}^{N_U}$  to the original feature vectors  $\{\mathbf{x}_i\}_{i=1}^{N_L}$ , the model matrix  $\mathbf{H}$  can be updated using,

$$\mathbf{H}_{N_L+N_U} = \begin{bmatrix} \mathbf{H}_{N_L} & \mathbf{U}_{N_U} \\ \mathbf{U}_{N_U}^T & \mathbf{V}_{N_U} \end{bmatrix}, \quad (16)$$

where  $U_{N_U} \in \mathbb{R}^{N_L \times N_U}$ , with  $u_{ij} = K(\mathbf{x}_i, \mathbf{x}_j)$ , and  $V_{N_U} \in \mathbb{R}^{N_U \times N_U}$  as,

$$V_{N_U} = \begin{bmatrix} K(\mathbf{x}_1, \mathbf{x}_1) & \cdots & K(\mathbf{x}_1, \mathbf{x}_{N_U}), \\ \vdots & \ddots & \vdots \\ K(\mathbf{x}_{N_U}, \mathbf{x}_1) & \cdots & K(\mathbf{x}_{N_U}, \mathbf{x}_{N_U}), \end{bmatrix} + \gamma^{-1} \mathbf{I}. \quad (17)$$

To update the parameters in SUST-LSSVM, the predicted labels for new data  $\{y_j\}_{j=1}^{N_U}$  obtained using the initial pre-trained LS-SVM classifier were merged with the initial labels  $\{y_i\}_{i=1}^{N_L}$ . Similar to the semi-supervised SWLDA, the important feature variables found by a step-wise approach with LDA were applied in SUST-LSSVM.

**E. EVALUATION**

The effectiveness of the SSL techniques was evaluated based on two metrics: BCI accuracy and information transfer rate (ITR). The BCI accuracy indicates how accurately the user can select the desired command, which purely evaluates the performance of the supervised/semi-supervised classification algorithms. ITR is proportional to BCI accuracy; however, it evaluates the performance of the BCI on a timescale that considers the number of commands per minute, thereby allowing a direct comparison of performance with previous studies that use the same/other BCI frameworks, unlike BCI accuracy.

For each test data, the number of data in the pre-training phase is validated in an unchanging state (i.e., the trained classifier remains unchanged over all testing data); however, in SSL, each test data can be added to the training dataset. Thus, we incrementally added test data to training datasets and retrained the classifier and plotted the changes in BCI accuracy and ITR.

**1) BCI ACCURACY**

Considering the application for the BCI system, the BCI accuracy should be defined as the accuracy of selecting a correct command or a character connected to one sound stimulus in each trial. For each trial of the testing data, we intentionally defined one sound stimulus as the target and the other (i.e.,  $N_s - 1$ ) sound stimuli were classified as nontargets. Since only one auditory sound is identified as a target label per trial, if a nontarget is targeted and misclassified, the target label is automatically misclassified as a nontarget as well. Under this unbalanced data scenario, the correct classification rate of the target data is of paramount importance when assessing the performance of the BCI. For example, subject “ja” included in the PASS2D contains 72-trial data (i.e., it contains 72 target data). If all target label data are classified correctly, the BCI accuracy will be  $72/72 = 100\%$ . This value was used as a metric of the accuracy of the BCI as with other BCI studies [5], [22].

**2) ITR**

A well-known metric in BCI studies is ITR, which can fit letter choices, typing speed, and classification accuracy into a single value [2], [48]. It is important to evaluate from a perspective other than classification accuracy because the classification accuracy tends to increase with a reduction in the number of classes and an increase in the number of sequences. However, such paradigm manipulation lowers the number of choices of letters and typing speed. ITR allows us to evaluate the amount of information per unit time (second) of an implemented BCI system quantitatively, independent of the paradigm and the dataset. It can be calculated as,

$$ITR = R \times \frac{60}{T}, \quad (18)$$

$$R = (\log_2 N + P \cdot \log_2 P + (1 - P) \cdot \log_2 \frac{1 - P}{N - 1}), \quad (19)$$

where  $T$  denotes the average time for a command selection,  $N$  denotes the number of selectable commands, and  $P$  represents the BCI accuracy ( $0 \leq P \leq 1$ ). The first term in Eq. (19)— $\log_2 N$ —denotes the information for  $N$ ; the second term— $P \cdot \log_2 P$ —denotes the information entropy defined by Shanon [49], and the last term denotes the information for the misclassification rate.

**3) CHANGES IN BCI PERFORMANCE THROUGH SSL**

First, we designed a backbone supervised learning classifier using all labeled data (pre-training phase data) and then plotted BCI performance (BCI accuracy and ITR) of SSL for each trial to evaluate the change in BCI accuracy and ITR when unlabeled data (testing phase data) incrementally provided pseudo-labels. Each plot denotes five trials, as displaying all plots would be too detailed. Note that because the number of trials in the testing phase varies between subjects, the length of the plot also varies between subjects.

**III. RESULTS AND DISCUSSIONS**

**A. BCI PERFORMANCES OVER SUBJECTS**

BCI accuracy and ITR with subject average for supervised and semi-supervised classifiers in AMUSE and PASS2D datasets are shown in Figs. 3 and 4. The SSL techniques improved averaged BCI accuracy and ITR of auditory ERP-based BCIs for all datasets and classifiers.

In the AMUSE dataset, the four classifiers (SWLDA, SKLDA, STDA, and LS-SVM) showed the averaged BCI accuracy as  $84.03 \pm 11.88$ ,  $83.86 \pm 12.01$ ,  $82.64 \pm 13.96$ ,  $84.10 \pm 12.54\%$  and ITR as  $6.40 \pm 2.12$ ,  $6.37 \pm 2.10$ ,  $6.23 \pm 2.37$ ,  $6.44 \pm 2.23$  bits/min, respectively. Combining SSL with the classifiers improved averaged BCI accuracy as,  $86.28 \pm 11.47$ ,  $88.13 \pm 10.31$ ,  $86.34 \pm 11.52$ , and  $86.90 \pm 12.09\%$  and ITR as  $6.82 \pm 2.12$ ,  $7.15 \pm 1.98$ ,  $6.84 \pm 2.14$ , and  $6.97 \pm 2.15$  bits/min. The improvements in averaged BCI accuracy owing to SSL were 2.25, 4.27, 3.70, and 2.80% for SWLDA, SKLDA, STDA, and LS-SVM respectively. In addition, the improvements in averaged ITR owing to

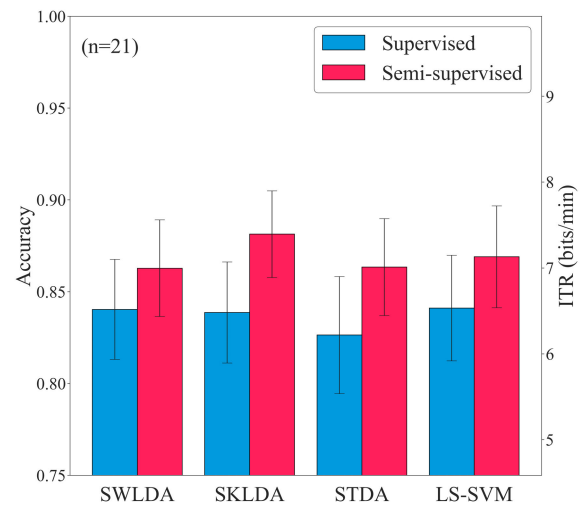
SSL were 0.42, 0.78, 0.61, and 0.53 bits/min for SWLDA, SKLDA, STDA, and LS-SVM respectively.

To evaluate the effect of the two different categorical independent variables (classifier type and learning method), we used a two-way analysis of variance (ANOVA), which allows us to evaluate the main effect of each independent variable and the interaction between them. We noted that there was no interaction between the factors of classifier and learning method ( $F(3, 21) = 0.05, p = 0.98$ ). There was no significant difference for the main effect of the classifiers ( $F(3, 21) = 0.11, p = 0.96$ ). The main effect of learning method was probably significant considering that the  $p$ -value is under 0.1 ( $F(1, 21) = 2.80, p < 0.10$ ) [50], [51]. Therefore, regardless of the classifiers, SSL techniques probably result in positive effect for AMUSE dataset.

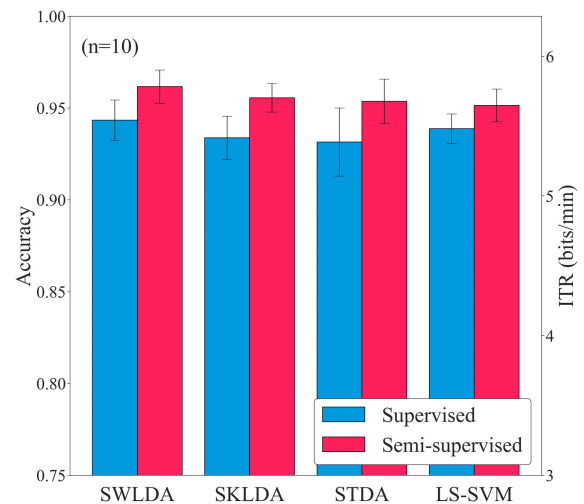
In the PASS2D dataset, the four classifiers (SWLDA, SKLDA, STDA, and LS-SVM) showed the averaged BCI accuracy as  $94.34 \pm 3.13$ ,  $93.38 \pm 3.37$ ,  $93.14 \pm 5.30$ ,  $93.87 \pm 2.30\%$  and ITR as  $5.36 \pm 0.43$ ,  $5.22 \pm 0.45$ ,  $5.24 \pm 0.70$ ,  $5.28 \pm 0.32$  bits/min, respectively. Combining SSL with the classifiers improved averaged BCI accuracy as  $96.16 \pm 2.59$ ,  $95.55 \pm 2.22$ ,  $95.36 \pm 3.45$ , and  $95.14 \pm 2.57\%$  and ITR as  $5.62 \pm 0.39$ ,  $5.53 \pm 0.35$ ,  $5.52 \pm 0.51$ , and  $5.46 \pm 0.37$  bits/min. The improvements in averaged BCI accuracy owing to SSL were 1.82, 2.17, 2.22, and 1.27% for SWLDA, SKLDA, STDA, and LS-SVM respectively. In addition, the improvements in averaged ITR owing to SSL were 0.26, 0.31, 0.28 and 0.18 bits/min for SWLDA, SKLDA, STDA, and LS-SVM respectively.

The two-way ANOVA revealed that there was no interaction between the factors of classifier and learning method ( $F(3, 10) = 0.07, p = 0.97$ ). There was no significant difference for main effect of the classifiers ( $F(3, 10) = 0.29, p = 0.83$ ). On the other hand, there was significant difference due to the main effect of learning method ( $F(1, 10) = 5.42, p = 0.02$ ). For PASS2D dataset, the positive effect of SSL technique was apparently confirmed.

The overall accuracy of AMUSE dataset was lower than that of PASS2D dataset. SKLDA classifiers with SSL showed the best performance in AMUSE dataset (see Fig. 3). In addition, the improvement rate for SKLDA classifiers was the highest among four classifiers. It is known that SKLDA is effective when the number of samples is extremely lower compared to the feature dimensionality [16]. However, the number of training samples in the dataset is considered to be sufficient. Therefore, we speculate that the characteristics of the ERPs were changed during the BCI experiments, and the ideal sample for calculating the exact covariance matrix was insufficient for each trial compared to the feature dimensionality. Using SKLDA, the covariance matrices could be regularized and adjusted as ideal and would lead to the high improvement rate of SSL. SWLDA classifiers showed the best performance in PASS2D dataset (see Fig. 4). It is assumed that the ERPs are more easily classified in PASS2D dataset. In this case, SWLDA classifiers are probably ideal for sophisticated data in which the ERPs tend to be



**FIGURE 3.** BCI accuracy and ITR with subject average for supervised and semi-supervised classifiers in AMUSE dataset. Bar plots denote the subject average, and error bars denote standard errors of all subjects.



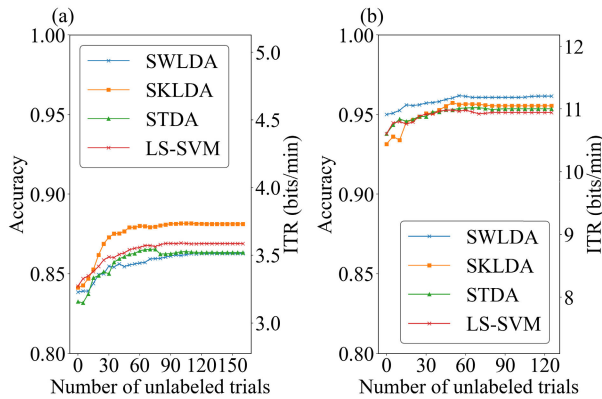
**FIGURE 4.** BCI accuracy and ITR with subject average for supervised and semi-supervised classifiers in PASS2D dataset. Bar plots denote subject average, and the error bars denote standard errors of all subjects.

clearly classified. It is difficult to choose the best classifier for auditory ERP-based BCI because the superiority or inferiority can be changed by the BCI paradigms and proficiency of subjects in controlling the BCI. However, it is clear that SSL could significantly improve the performance of auditory ERP-based BCIs.

### B. CHANGES IN BCI PERFORMANCE THROUGH SSL

To investigate the effect of SSL in more detail, we plotted the changes in BCI accuracy and ITR when the number of unlabeled data in testing phase for re-training increased. Fig. 5 shows the BCI accuracy and ITR averaged across subjects. Since the number of testing data varied from subject to subject, the results of other subjects were supplemented according to the subject with the largest number of tests





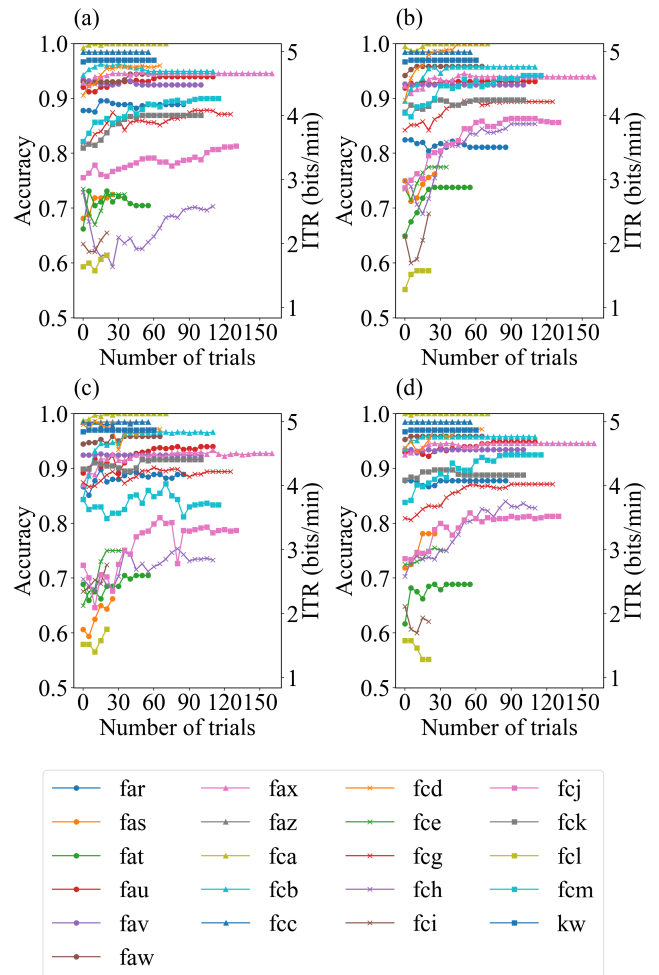
**FIGURE 5.** The subject average changes of BCI accuracy and ITR when the number of unlabeled data is incrementally increased by re-training. The x-label denotes the number of unlabeled trials measured in testing phase added to the data in pre-training phase. (a) AMUSE dataset, (b) PASS2D dataset.

(i.e., if 150 trials was the maximum, the results of the subject who has only 120 trials were concatenated with the same results as 120-th trial from 121-st to 150-th one). The vertical axis denotes accuracy and ITR. The horizontal axis is the number of trials of unlabeled data added in SSL; thus, the leftmost points are the results of classifying testing data using supervised learning.

Similar to previous studies, the accuracy was logarithmically increased by SSL [22], [24]. The tendency was more remarkable for AMUSE dataset because of its lower overall accuracy (i.e., there is still room to grow). From these results, it is clear that SSL has the potential to improve the performance of auditory ERP-based BCIs, especially with a small number of unlabeled data.

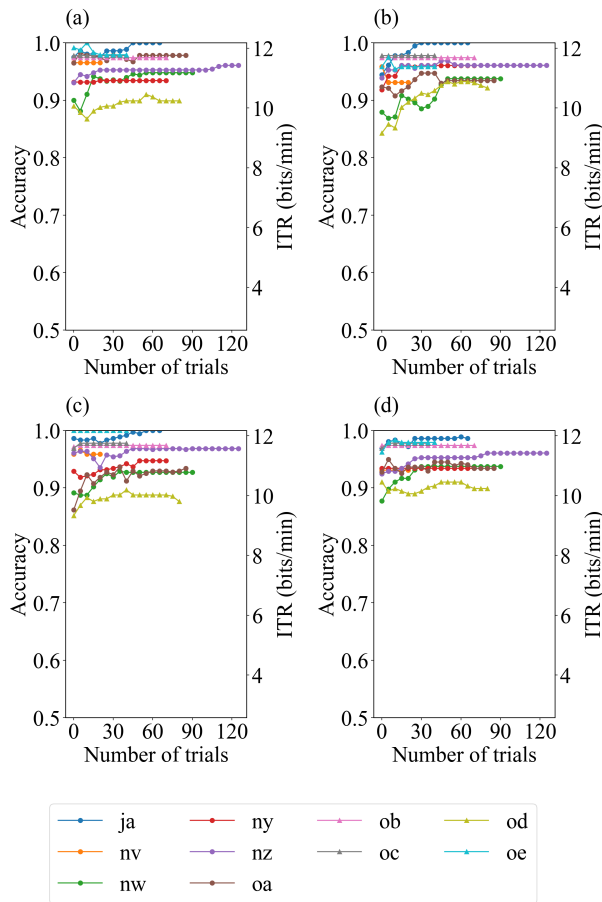
**C. DIFFERENCES OF EFFECT OF SSL AMONG SUBJECTS**

Figs. 6 and 7 plot the changes in BCI accuracy and ITR of each subject as the number of unlabeled data added for training is increased incrementally in the AMUSE and PASS2D datasets, respectively. The horizontal axis is the number of trials of unlabeled data added in SSL; thus, the leftmost points are the results of supervised learning. Unlike the previous subsection, the number of unlabeled data is shown uncorrected in order to make comparisons for each individual. In both figures, each graph denotes a classifier ((a) SWLDA, (b) SKLDA, (c) STDA, and (d) LS-SVM). Although the BCI performances of most subjects were incrementally improved by the four classifiers, for each subject, each classifier showed different tendencies. For example, the accuracy of subject “fch” in Fig. 6 was incrementally improved by LS-SVM. However, other three classifiers decreased the accuracy in the middle or end of the incremental updating. On the other hand, the accuracy of “fcl” seems to be incrementally increased by the SWLDA, SKLDA, and STDA, but not by the LS-SVM. In previous studies, the dimension of features and parameters of classifiers are commonly optimized for each subject [52], [53]. This results suggest that we will need to select the best classifier for each subject.



**FIGURE 6.** BCI accuracy and ITR changes for AMUSE dataset when the number of unlabeled data is incrementally increased for re-training. The x-label denotes the number of unlabeled trials measured in testing phase added to the data in pre-training phase. (a) SWLDA, (b) SKLDA, (c) STDA, and (d) LS-SVM.

To analyze the effect of SSL from the viewpoints of spatial patterns of ERPs, the topological maps of P300 and weight vectors  $w$  of the SKLDA classifier are shown in Fig. 8. The improvement of BCI performance was remarkable for subject “fcj” by SKLDA classifier in AMUSE dataset; thus, we selected this setting for the visualization. In addition, the dimensions of the weight vectors of SWLDA, STDA, and LS-SVM were reduced from the  $MP$ -dimensions, but the weight vectors of SKLDA remained full (2900 for AMUSE dataset and 3150 for PASS2D dataset), hence the weights could be visualized as topography. The feature dimensions selected by the SWLDA was 972 for AMUSE dataset and 1108 for PASS2D dataset averaged across subjects and trials. The selected dimensions were also used for LS-SVM. STDA projected the feature vectors from the full dimensionality to four dimensions ( $L = 2$ ) after finding two projection matrices. The color gradation of left four maps in the figure indicates the strength of peak amplitude of P300. In contrast to the color gradation of nontarget which is not extremely



**FIGURE 7.** BCI accuracy and ITR changes for PASS2D dataset when the number of unlabeled data is incrementally increased for re-training. The x-label denotes the number of unlabeled trials measured in testing phase added to the data in pre-training phase. (a) SWLDA, (b) SKLDA, (c) STDA, and (d) LS-SVM.

different between pre-training and testing phase, the strength of target P300 was apparently increased in parietal area in testing phase compared to pre-training phase. Following the changes of P300, the value of weight vectors in position FCz was increased in testing phase compared to pre-training phase (see right maps in Fig. 8). Halder *et al.* showed that there is training effect in auditory ERP-based BCI by which the peaks of target P300 in parietal become stronger as testing sessions are increased [54]. It suggests that one of the reasons why the SSL improved the accuracy is, the updating weight vectors followed the increase of peak strength of P300 in parietal area.

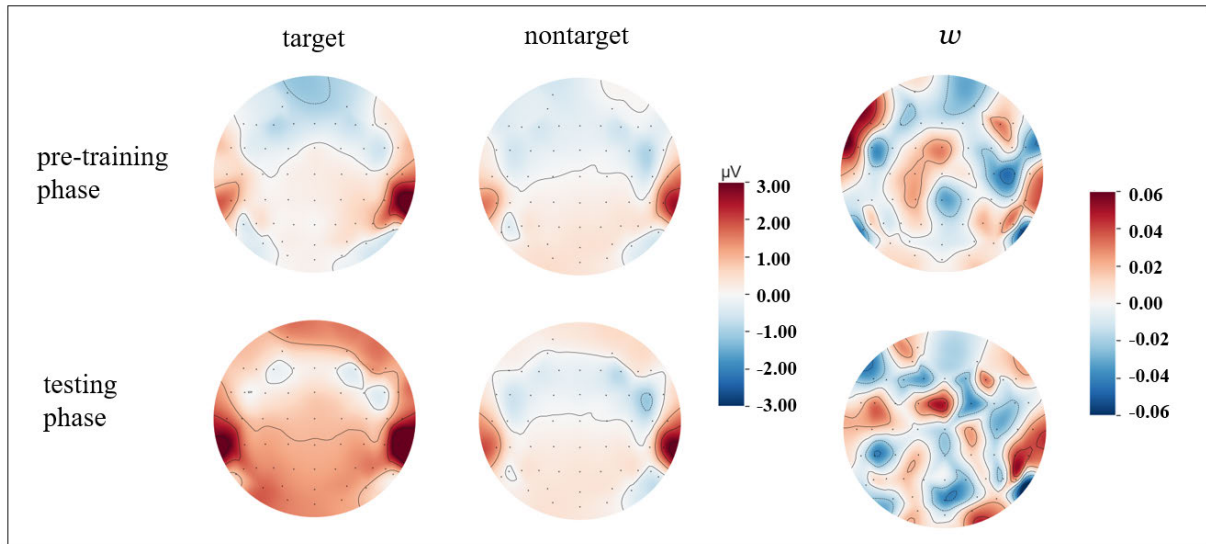
The changes of topographic maps of P300 and weight vectors of the SKLDA classifier for subject “nw” of PASS2D dataset is shown in Fig. 9. We assessed the reason of accuracy decreasing in SSL during testing phase by checking topological maps. The leftmost topographic maps show the weight vectors of SKLDA and target P300 in pre-training phase. The other maps show the sequential changes of updated weight vectors and P300. The subject “nw” has 96 trials in testing phase, thus 95-th trial is the last update of weight vector. As shown in Fig. 7, the accuracy was decreased

while updating weight vectors by SKLDA with SSL. The reason can be explained by the Fig. 9. The strength of peaks of P300 in parietal area was decreased around 24-th trial and increased again after 48-th trial. This changes indicated that the weight vectors trained in pre-training phase is more suitable for the P300 after 48-th trial than updated weight vectors at 24-th trial. During the testing phase, subjects were forced to be more focused. Changing from the pre-training phase to the actual BCI test task may cause some subjects to feel fatigue and emotional changes after the pre-training phase. They indicated that the BCI causes considerable fatigue, and would not prefer to use it very often [6]. Therefore, it is difficult to increase the peak of P300 linearly and it probably leads to the negative effect of SSL. Interestingly, the decrease of BCI performance was mitigated when the STDA and LS-SVM classifiers were applied. The projected features of STDA and support vectors of LS-SVM probably prevented the reduction. It suggests that combining feature extraction methods and the SKLDA would be effective for solving the accuracy reduction issue. Furthermore, applying an SSL technique after collecting some trial unlabeled data instead of every trial would be useful.

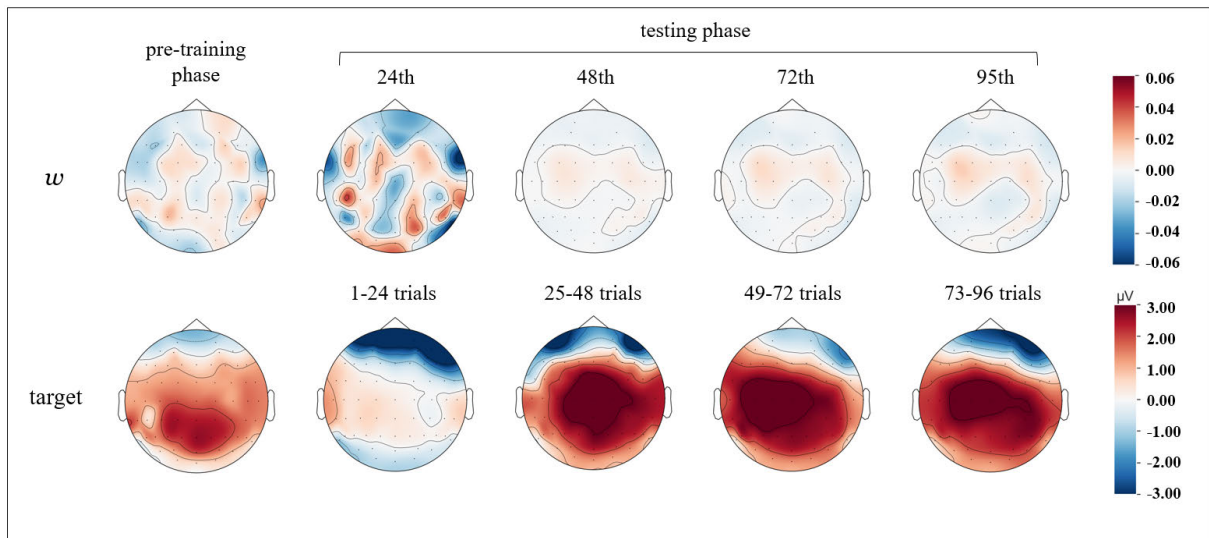
Tables 1 and 2 summarise individual accuracy and ITR of classifiers when all unlabeled data were used for SSL and were not used for re-training classifiers. From subsections III-A and III-B, SSL has been proven to improve the averaged BCI performance of auditory ERP-based BCI. However, when we focus on each individual, some subjects’ accuracy decreased by SSL. In addition, different subjects resulted in the decline of BCI performance depending on the classifiers. It would be necessary to optimize the selection of classifiers for each individual as well. It has also been confirmed in previous studies that the trend of improvement in BCI accuracy stops and the accuracy stabilizes as the number of additional unlabeled data increases [22], [55]. This means that the more the unlabeled data increased, the weaker the effect of SSL becomes. The used datasets included a sufficient number of pre-training data. Therefore, the parameters of classifiers would have already been sophisticated before re-training and the SSL not changing the parameter significantly would lead to the stable BCI accuracy.

#### D. COMPARISON WITH PREVIOUS STUDIES

We calculated ITR to compare our results with other previous researches, even if their paradigms are different. A previous study using SKLDA showed that an SSL technique improved the ITR of MI-based BCI from 0.32 bits/min to 0.84 bits/min [25]. Our results using SKLDA classifiers showed that the ITRs were improved from 6.37 bits/min to 7.15 bits/min in AMUSE dataset, and from 5.22 bits/min to 5.53 bits/min in PASS2D dataset. Furthermore, there has been a previous study on visual ERP-based BCI using SVM classifiers. Li *et al.* showed that the ITR improved from 9.12 bits/min to 14.03 bits/min [22]. Compared to the study, our results showed less improvements of ITRs using LS-SVM classifiers (from 6.44 bits/min to 6.97 bits/min in AMUSE



**FIGURE 8.** The topographic maps for subject “fcj” in AMUSE dataset. Left maps show the peak of P300 for target and nontarget recorded in pre-training phase and testing phase. Right maps show the strength of weight vectors for P300 trained by SKLDA.



**FIGURE 9.** The topographic maps of weight vectors and P300 for subject “nw” in PASS2D dataset. The leftmost topographic maps show the weight vectors and P300 in pre-training phase. Other maps show the change of updated weight vectors and P300 in testing phase.

dataset and from 5.28 bits/min to 5.46 bits/min in PASS2D dataset). It might be attributed to the difference of number of data used for training backbone classifiers. In the previous study, there were only three trials of training data. Thus, the effect of the new data and the range of improvement in accuracy would have been greater. Although it is not possible to show a consistent trend because the experimental setting is not the same across studies, this study could show that SSL is effective for the four mentioned classifiers in auditory ERP-based BCIs.

**E. CALCULATION TIME**

In the case of the setting mentioned in the previous section, the learning process of the SSL techniques is performed on

a trial-by-trial basis in the testing phase by adding testing data to the data obtained in the pre-training phase. Here, the high computational cost of the incremental re-training makes command selection time consuming, and it has a negative impact on the performance of the BCI. Thus, it is important to clarify the relationship between the addition of unlabeled data and the computational cost of re-training. Fig. 10 shows the computation time for re-training four classifiers by SSL techniques. To determine the computation time, a laptop with 64-bit Windows 10 was used for the simulation. The CPU was an Intel Core i7-10710U, 16 GB of RAM, and data were recorded and read on a 512 GB SSD. The analysis code was implemented in Python 3.6.

**TABLE 1.** BCI accuracy for backbone classifiers (SWLDA, SLDA, STDA, and LS-SVM) and SSL combined techniques. The top table and the bottom table present the results for AMUSE dataset and PASS2D dataset, respectively.

BCI accuracy of AMUSE (%)								
Sub	SWLDA	Semi-supervised SWLDA	SKLDA	Semi-supervised SKLDA	STDA	Semi-supervised STDA	LS-SVM	Semi-supervised LS-SVM
far	87.78	<b>88.89</b>	<b>83.33</b>	81.11	86.67	<b>88.89</b>	87.78	87.78
fas	71.88	<b>75.00</b>	78.12	78.12	53.12	<b>65.62</b>	75.00	<b>78.12</b>
fat	65.57	<b>70.49</b>	63.93	<b>73.77</b>	65.57	<b>70.49</b>	59.02	<b>68.85</b>
fau	91.38	<b>93.97</b>	88.79	<b>93.10</b>	89.66	<b>93.97</b>	92.24	<b>94.83</b>
fav	<b>93.40</b>	92.45	<b>94.34</b>	92.45	92.45	92.45	93.40	93.40
faw	93.06	<b>94.44</b>	93.06	<b>95.83</b>	94.44	<b>95.83</b>	94.44	<b>95.83</b>
fax	93.33	<b>94.55</b>	92.12	<b>93.94</b>	84.85	<b>92.73</b>	90.91	<b>94.55</b>
faz	98.44	98.44	98.44	98.44	98.44	98.44	98.44	98.44
fca	100.00	100.00	100.00	100.00	98.68	<b>100.00</b>	100.00	100.00
fcb	92.37	<b>94.92</b>	86.44	<b>95.76</b>	84.75	<b>96.61</b>	92.37	<b>95.76</b>
fcc	98.44	98.44	98.44	98.44	98.44	98.44	98.44	98.44
fed	90.00	<b>97.14</b>	87.14	<b>100.00</b>	97.14	97.14	95.71	<b>97.14</b>
fce	<b>75.00</b>	72.50	75.00	<b>77.50</b>	62.50	<b>75.00</b>	70.00	<b>75.00</b>
fcg	81.82	<b>87.12</b>	83.33	<b>89.39</b>	83.33	<b>89.39</b>	82.58	<b>87.12</b>
fch	<b>77.59</b>	70.69	73.28	<b>85.34</b>	64.66	<b>73.28</b>	69.83	<b>82.76</b>
fei	65.52	65.52	68.97	68.97	68.97	<b>72.41</b>	<b>68.97</b>	62.07
fej	75.54	<b>81.29</b>	73.38	<b>85.61</b>	76.26	<b>78.42</b>	72.66	<b>81.29</b>
fck	81.31	<b>86.92</b>	85.98	<b>89.72</b>	90.65	<b>91.59</b>	87.85	<b>88.79</b>
fcl	55.17	<b>62.07</b>	51.72	<b>62.07</b>	58.62	<b>62.07</b>	<b>58.62</b>	55.17
fem	81.67	<b>90.00</b>	88.33	<b>94.17</b>	<b>89.17</b>	83.33	82.50	<b>92.50</b>
kw	95.45	<b>96.97</b>	96.97	96.97	96.97	96.97	95.45	<b>96.97</b>
Ave	84.03	<b>86.28</b>	83.86	<b>88.13</b>	82.64	<b>86.34</b>	84.10	<b>86.90</b>
SD	11.88	11.47	12.01	10.31	13.96	11.52	12.54	12.09
BCI accuracy of PASS2D (%)								
ja	94.44	<b>100.00</b>	94.44	<b>100.00</b>	98.61	<b>100.00</b>	97.22	<b>98.61</b>
nv	96.55	96.55	<b>96.55</b>	93.10	93.10	93.10	93.10	93.10
nw	89.58	<b>94.79</b>	90.62	<b>93.75</b>	90.62	<b>92.71</b>	89.58	<b>93.75</b>
ny	92.11	<b>93.42</b>	90.79	<b>96.05</b>	93.42	<b>94.74</b>	93.42	93.42
nz	92.13	<b>96.06</b>	93.70	<b>96.06</b>	96.06	<b>96.85</b>	92.13	<b>96.10</b>
oa	96.70	<b>97.80</b>	90.11	<b>93.41</b>	83.52	<b>93.41</b>	92.31	<b>93.41</b>
ob	97.40	97.40	97.40	97.40	97.40	97.40	97.40	97.40
oc	97.78	97.78	97.78	97.78	95.56	<b>97.78</b>	95.56	<b>97.78</b>
od	88.76	<b>89.89</b>	86.52	<b>92.13</b>	83.15	<b>87.64</b>	<b>92.13</b>	89.89
oe	97.92	97.92	95.83	95.83	100.00	100.00	95.83	<b>97.92</b>
Ave	94.34	<b>96.16</b>	93.38	<b>95.55</b>	93.14	<b>95.36</b>	93.87	<b>95.14</b>
SD	3.13	2.59	3.37	2.22	5.30	3.45	2.30	2.57

During the actual ERP-based BCI operation, there is a possibility that re-training for each command selection may be required in time. In the BCI system, the presentation of the whole auditory stimuli is followed by estimation of the target sound. Therefore, there is one trial between the completion of the command selection and the start of the next command estimation. If the calculation time for re-training is

less than the time length of the trial, the negative effect of SSL techniques on BCI performance can be avoided. One trial of sound stimulation required 15.62 s for the AMUSE dataset and 30.25 s for the PASS2D dataset in the testing phase. They are shown in Fig. 10 as purple dash line.

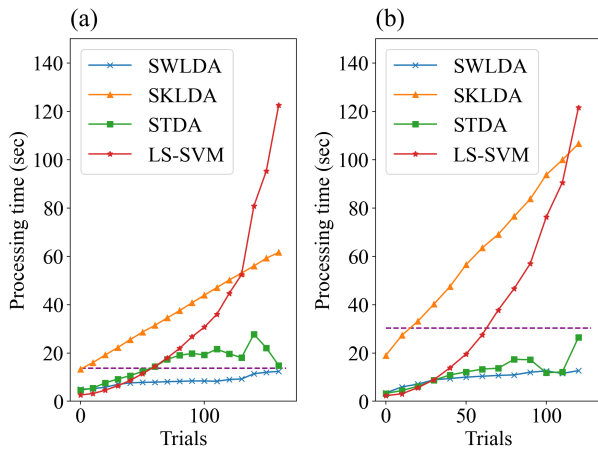
Among the four classifiers, SWLDA had the lowest computational cost, and the re-training time did not exceed the

**TABLE 2.** ITR for backbone classifiers (SWLDA, SLDA, STDA, and LS-SVM) and SSL combined techniques. The top table and the bottom table present the results for AMUSE dataset and PASS2D dataset, respectively.

ITR of AMUSE (bits/min)								
Sub	SWLDA	Semi-supervised SWLDA	SKLDA	Semi-supervised SKLDA	STDA	Semi-supervised STDA	LS-SVM	Semi-supervised LS-SVM
far	6.78	<b>7.01</b>	<b>5.95</b>	5.56	6.57	<b>7.01</b>	6.78	6.78
fas	4.13	<b>4.58</b>	5.07	5.07	1.92	<b>3.30</b>	4.58	<b>5.07</b>
fat	3.29	<b>3.94</b>	3.09	<b>4.40</b>	3.29	<b>3.94</b>	2.52	<b>3.72</b>
fau	7.54	<b>8.13</b>	6.99	<b>7.93</b>	7.17	<b>8.13</b>	7.73	<b>8.34</b>
fav	<b>7.99</b>	7.78	<b>8.22</b>	7.78	7.78	7.78	7.99	7.99
faw	7.91	<b>8.25</b>	7.91	<b>8.60</b>	8.25	<b>8.60</b>	8.25	<b>8.60</b>
fax	7.98	<b>8.27</b>	7.70	<b>8.12</b>	6.22	<b>7.84</b>	7.43	<b>8.27</b>
faz	9.35	9.35	9.35	9.35	9.35	9.35	9.35	9.35
fca	9.93	9.93	9.93	9.93	9.43	<b>9.93</b>	9.93	9.93
fcf	7.76	<b>8.36</b>	6.52	<b>8.58</b>	6.20	<b>8.81</b>	7.76	<b>8.58</b>
fcc	9.35	9.35	9.35	9.35	9.35	9.35	9.35	9.35
fed	7.24	<b>8.96</b>	6.66	<b>9.93</b>	8.96	8.96	8.57	<b>8.96</b>
fce	<b>4.58</b>	4.22	4.58	<b>4.97</b>	2.92	<b>4.58</b>	3.87	<b>4.58</b>
fcg	5.68	<b>6.65</b>	5.95	<b>7.11</b>	5.95	<b>7.11</b>	5.81	<b>6.65</b>
fch	<b>4.98</b>	3.96	4.33	<b>6.32</b>	3.18	<b>4.33</b>	3.85	<b>5.85</b>
fei	3.29	3.29	3.73	3.73	3.73	<b>4.21</b>	<b>3.73</b>	2.87
fej	4.67	<b>5.59</b>	4.35	<b>6.37</b>	4.78	<b>5.12</b>	4.24	<b>5.59</b>
fck	5.59	<b>6.61</b>	6.44	<b>7.18</b>	7.38	<b>7.58</b>	6.80	<b>6.99</b>
fcl	2.12	<b>2.87</b>	1.79	<b>2.87</b>	2.48	<b>2.87</b>	<b>2.48</b>	2.12
fcm	5.66	<b>7.24</b>	6.89	<b>8.18</b>	<b>7.06</b>	5.95	5.80	<b>7.79</b>
kw	8.50	<b>8.91</b>	8.91	8.91	8.91	8.91	8.50	<b>8.91</b>
Ave	6.40	<b>6.82</b>	6.37	<b>7.15</b>	6.23	<b>6.84</b>	6.44	<b>6.97</b>
SD	2.12	2.12	2.10	1.98	2.37	2.14	2.23	2.15
ITR of PASS2D (bits/min)								
ja	5.34	<b>6.29</b>	5.34	<b>6.29</b>	6.00	<b>6.29</b>	5.76	<b>6.00</b>
nv	5.65	5.65	<b>5.65</b>	5.16	5.16	5.16	5.16	5.16
nw	4.71	<b>5.39</b>	4.84	<b>5.25</b>	4.84	<b>5.11</b>	4.71	<b>5.25</b>
ny	5.03	<b>5.20</b>	4.86	<b>5.58</b>	5.20	<b>5.38</b>	5.20	5.20
nz	5.03	<b>5.58</b>	5.24	<b>5.58</b>	5.58	<b>5.70</b>	5.03	<b>5.58</b>
oa	5.68	<b>5.85</b>	4.78	<b>5.20</b>	4.03	<b>5.20</b>	5.05	<b>5.20</b>
ob	5.79	5.79	5.79	5.79	5.79	5.79	5.79	5.79
oc	5.85	5.85	5.85	5.85	5.50	<b>5.85</b>	5.50	<b>5.85</b>
od	4.61	<b>4.75</b>	4.35	<b>5.03</b>	3.99	<b>4.48</b>	<b>5.03</b>	4.75
oe	5.87	5.87	5.54	5.54	6.29	6.29	5.54	<b>5.87</b>
Ave	5.36	<b>5.62</b>	5.22	<b>5.53</b>	5.24	<b>5.52</b>	5.28	<b>5.46</b>
SD	0.43	0.39	0.45	0.35	0.70	0.51	0.32	0.37

stimulus time for both datasets. i.e., even a paradigm other than an ERP-based BCI paradigm could use SWLDA for semi-supervised BCI without loss of interval training time. The plots show that it is easy to infer class label of testing data within the acceptable processing time for SWLDA, SKLDA, and LS-SVM, but difficult for STDA. For SWLDA and SKLDA, re-training time increases linearly. Therefore, the

number of additional unlabeled data affecting the BCI performance can be predicted easily. The calculation cost of LS-SVM increases with the number of total training samples (the complexity order is  $O(N^3)$  [56]). The time length was initially less than one second, but it increased drastically following the increasing of trials and exceeded the time limit for selecting one command in the AMUSE and PASS2D datasets.



**FIGURE 10.** Calculation time for re-training the four classifiers by SSL techniques. The vertical axis denotes number of incremental learning in testing phase. (a) AMUSE dataset and (b) PASS2D dataset. The purple dash line denotes the time length of one trial in testing phase.

The estimation of the processing time of STDA is more difficult. STDA optimizes two projection matrices, and transforms the original samples into low-dimensional features, and re-trains new classifier. The more the data is accumulated, the more difficult it is to optimize the projection matrices. Such a method with multiple processes seems to be incompatible with SSL.

In this study we updated the classifiers in every trial in testing phase. However, updating in every trial is not necessary. Updating classifiers every few trials is also expected to raise the BCI performance. Therefore, we should determine the timing of applying SSL in actual online BCI applications considering its processing time. We hope that our results concerning processing time will help the future designs.

#### F. LIMITATIONS

There are three limitations in this study:

(1) Lack of consideration of subjects' motivation: The simulation results do not consider using an SSL technique at the time of measurement. If the subject participated in the experimental task, the online testing phase shows the results on the screen or speaker and notifies the subject regarding the command that was selected. This feedback effect raises the motivation of subjects and improves the BCI accuracy [31], [57]. If more accurate feedback is returned for SSL, subjects may be more motivated and focused on the task, which can result in more discrimination of unlabeled data. Therefore, it is necessary to prepare a BCI that incorporates SSL and validates the case of online testing and measurement simultaneously.

(2) Variation in the number of trials during the testing phase: The number of trials of the labeled data (in the pre-training phase) was the same across subjects in both datasets (AMUSE dataset: 48 trials, PASS2D dataset: 27 trials). However, the number of trials in the testing phase varied across subjects, thereby making it impossible to assess the

same number of trials. In particular, a few subjects had extremely low numbers of unlabeled data compared to other subjects (see Figs. 6 and 7). Since matching the number of unlabeled data to the lowest state resulted in a very small number of trials that could be validated, the equalization process was discarded in this study, and the incremental update strategy was conducted using the number of trials conducted for each subject. One future work is to investigate the effect of longer-term SSL with some amount of data for each subject.

(3) Data reliability: All data, including the original training data and the newly added unlabeled data, were treated uniformly. Therefore, the quality of the data is not considered, and there is a concern that the effect of poor quality data may be significant. Furthermore, when the estimated labels of unlabeled data in the testing phase are different from the true labels, the accuracy of the classification model will decrease. The problem is related to user's motivation or classifier's incompleteness; if the user is not focused on the selection of commands or is disheartened while playing BCI, ERPs will not occur and all data will be classified as nontarget. In fact, immature classification models often assign incorrect labels to unlabeled data. Our evaluation showed that the accuracy worsened with SSL based on a low-accuracy classifier before re-training started (see Tables. 1 and 2). To avoid such a decrease in accuracy, it may be useful to calculate measures such as data reliability and reflect those values as constraints in SSL. This method using unlabeled data more intelligently is called safe SSL [58], [59], and it has the potential to improve our results in future studies.

#### IV. CONCLUSION

This study investigated the adaptability of SSL to auditory ERP-based BCIs. Four classifiers (SWLDA, SKLDA, STDA, and LS-SVM) and their SSL expansion versions were applied to two public datasets (AMUSE and PASS2D).

As subject averages, the BCI accuracies and ITRs for all four classifiers were improved by combining SSL techniques into the backbone classifiers. Thus, we verified the positive effect of SSL on auditory ERP-based BCIs regardless of traditional and advanced ML classifiers. However, the sequential plots, which showed the changes in BCI accuracy with incremental updates of the classifiers and the changes in EEG topographical maps, suggested that the BCI accuracy was sometimes deteriorated by SSL especially in the beginning of testing phase for a few subjects because of weakening of the ERP. In addition, it was revealed that except SWLDA, the remaining three classifiers cannot be applied for updating parameters in every trial of auditory ERP-based BCIs because of their high computational cost. Thus, auditory ERP-based BCI designers should update backbone classifiers by SSL techniques when some trials were finished and should make SSL techniques more stable. We expect that our study will encourage future researchers working on auditory paradigms to use SSL techniques and that it will help improve the performance of their BCI frameworks.

## REFERENCES

- [1] N. Birbaumer, N. Ghanayim, T. Hinterberger, I. Iversen, B. Kotchoubey, A. Kübler, J. Perelmouter, E. Taub, and H. Flor, "A spelling device for the paralysed," *Nature*, vol. 398, no. 6725, pp. 297–298, Mar. 1999.
- [2] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, D. J. McFarland, P. H. Peckham, G. Schalk, E. Donchin, L. A. Quatrano, C. J. Robinson, and T. M. Vaughan, "Brain-computer interface technology: A review of the first international meeting," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 2, pp. 164–173, Jun. 2000.
- [3] R. Leeb, F. Lee, C. Keinrath, R. Scherer, H. Bischof, and G. Pfurtscheller, "Brain-computer communication: Motivation, aim, and impact of exploring a virtual apartment," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 15, no. 4, pp. 473–482, Dec. 2007.
- [4] B. He, L. Yang, C. Wilke, and H. Yuan, "Electrophysiological imaging of brain activity and connectivity—Challenges and opportunities," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 7, pp. 1918–1931, Apr. 2011.
- [5] S. Halder, E. M. Hammer, S. C. Kleih, M. Bogdan, W. Rosenstiel, N. Birbaumer, and A. Kübler, "Prediction of auditory and visual P300 brain-computer interface aptitude," *PLoS ONE*, vol. 8, no. 2, Feb. 2013, Art. no. e53513.
- [6] I. Käthner, A. Kübler, and S. Halder, "Comparison of eye tracking, electrooculography and an auditory brain-computer interface for binary communication: A case study with a participant in the locked-in state," *J. Neuroeng. Rehabil.*, vol. 12, no. 1, 2015, Art. no. 76.
- [7] C. Guger, R. Spataro, F. Pellas, B. Z. Allison, A. Heilinger, R. Ortner, W. Cho, R. Xu, V. La Bella, G. Edlinger, J. Annen, G. Mandalá, C. Chatelle, and S. Laureys, "Assessing command-following and communication with vibro-tactile P300 brain-computer interface tools in patients with unresponsive wakefulness syndrome," *Frontiers Neurosci.*, vol. 12, p. 423, Jun. 2018.
- [8] L. A. Farwell and E. Donchin, "Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalogr. Clin. Neurophysiol.*, vol. 70, no. 6, pp. 510–523, Dec. 1988.
- [9] T. W. Picton, "The P300 wave of the human event-related potential," *J. Clin. Neurophysiol.*, vol. 9, no. 4, pp. 456–479, Oct. 1992.
- [10] P. Sawangjai, S. Hompoonsup, P. Leelaarporn, S. Kongwudhikunakorn, and T. Wilaiprasitporn, "Consumer grade EEG measuring sensors as research tools: A review," *IEEE Sensors J.*, vol. 20, no. 8, pp. 3996–4024, Apr. 2020.
- [11] M. Ogino, S. Kanoga, M. Muto, and Y. Mitsukura, "Analysis of prefrontal single-channel EEG data for portable auditory ERP-based brain-computer interfaces," *Frontiers Hum. Neurosci.*, vol. 13, p. 250, Jul. 2019.
- [12] I. Käthner, C. A. Ruf, E. Pasqualotto, C. Braun, N. Birbaumer, and S. Halder, "A portable auditory P300 brain-computer interface with directional cues," *Clin. Neurophysiol.*, vol. 124, no. 2, pp. 327–338, Feb. 2013.
- [13] M. De Vos, K. Gandras, and S. Debener, "Towards a truly mobile auditory brain-computer interface: Exploring the P300 to take away," *Int. J. Psychophysiol.*, vol. 91, no. 1, pp. 46–53, Jan. 2014.
- [14] I. Käthner, A. Kübler, and S. Halder, "Rapid P300 brain-computer interface communication with a head-mounted display," *Frontiers Neurosci.*, vol. 9, p. 207, Jun. 2015.
- [15] S. Kundu and S. Ari, "P300 based character recognition using sparse autoencoder with ensemble of SVMs," *Biocybern. Biomed. Eng.*, vol. 39, no. 4, pp. 956–966, Oct. 2019.
- [16] 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. 814–825, May 2011.
- [17] A. Dithaporn, N. Banluesombatkul, S. Ketrat, E. Chuangsuwanich, and T. Wilaiprasitporn, "Universal joint feature extraction for P300 EEG classification using multi-task autoencoder," *IEEE Access*, vol. 7, pp. 68415–68428, 2019.
- [18] J. Lian, L. Bi, and W. Fei, "A novel event-related potential-based brain-computer interface for continuously controlling dynamic systems," *IEEE Access*, vol. 7, pp. 38721–38729, 2019.
- [19] L. Bi, J. Zhang, and J. Lian, "EEG-based adaptive driver-vehicle interface using variational autoencoder and PI-TSVM," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 10, pp. 2025–2033, Oct. 2019.
- [20] K. Chen, L. Yao, D. Zhang, X. Chang, G. Long, and S. Wang, "Distributionally robust semi-supervised learning for people-centric sensing," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 3321–3328.
- [21] S. Kanoga, T. Hoshino, and H. Asoh, "Subject transfer framework based on source selection and semi-supervised style transfer mapping for semg pattern recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 1349–1353.
- [22] Y. Li, C. Guan, H. Li, and Z. Chin, "A self-training semi-supervised SVM algorithm and its application in an EEG-based brain computer interface speller system," *Pattern Recognit. Lett.*, vol. 29, no. 9, pp. 1285–1294, Jul. 2008.
- [23] Z. Gu, Z. Yu, Z. Shen, and Y. Li, "An online semi-supervised brain-computer interface," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 9, pp. 2614–2623, Sep. 2013.
- [24] L. F. Nicolas-Alonso, R. Corralejo, J. Gomez-Pilar, D. Álvarez, and R. Hornero, "Adaptive semi-supervised classification to reduce intersession non-stationarity in multiclass motor imagery-based brain-computer interfaces," *Neurocomputing*, vol. 159, pp. 186–196, Jul. 2015.
- [25] A. Schwarz, J. Brandstetter, J. Pereira, and G. R. Müller-Putz, "Direct comparison of supervised and semi-supervised retraining approaches for co-adaptive BCIs," *Med. Biol. Eng. Comput.*, vol. 57, no. 11, pp. 2347–2357, Nov. 2019.
- [26] J. Chen, Z. Yu, and Z. Gu, "Semi-supervised deep learning in motor imagery-based brain-computer interfaces with stacked variational autoencoder," in *J. Phys. Conf. Ser.*, vol. 1631, Mar. 2020, Art. no. 012007.
- [27] A. R. Murguialday, J. Hill, M. Bensch, S. Martens, S. Halder, F. Nijboer, B. Schoelkopf, N. Birbaumer, and A. Gharabaghi, "Transition from the locked in to the completely locked-in state: A physiological analysis," *Clin. Neurophysiol.*, vol. 122, no. 5, pp. 925–933, May 2011.
- [28] D. S. Klobassa, T. M. Vaughan, P. Brunner, N. E. Schwartz, J. R. Wolpaw, C. Neuper, and E. W. Sellers, "Toward a high-throughput auditory P300-based brain-computer interface," *Clin. Neurophysiol.*, vol. 120, no. 7, pp. 1252–1261, Jul. 2009.
- [29] A. M. Brouwer and J. Van Erp, "A tactile P300 brain-computer interface," *Frontiers Neurosci.*, vol. 10, p. 19, May 2010.
- [30] M. Falkenstein, N. A. Koshlykova, V. N. Kiroj, J. Hoormann, and J. Hohnsbein, "Late ERP components in visual and auditory Go/Nogo tasks," *Electroencephalogr. Clin. Neurophysiol./Evoked Potentials Sect.*, vol. 96, no. 1, pp. 36–43, Jan. 1995.
- [31] F. Nijboer, A. Furdea, I. Gunst, J. Mellinger, D. J. McFarland, N. Birbaumer, and A. Kübler, "An auditory brain-computer interface (BCI)," *J. Neurosci. Methods*, vol. 167, no. 1, pp. 43–50, 2008.
- [32] X. An, J. Höhne, D. Ming, and B. Blankertz, "Exploring combinations of auditory and visual stimuli for gaze-independent brain-computer interfaces," *PLoS ONE*, vol. 9, no. 10, Oct. 2014, Art. no. e111070.
- [33] N. Simon, I. Käthner, C. A. Ruf, E. Pasqualotto, A. Kübler, and S. Halder, "An auditory multiclass brain-computer interface with natural stimuli: Usability evaluation with healthy participants and a motor impaired end user," *Frontiers Hum. Neurosci.*, vol. 8, p. 1039, Dec. 2015.
- [34] J. Schäfer and K. Strimmer, "A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics," *Stat. Appl. Genet. Mol. Biol.*, vol. 4, no. 1, pp. 1–32, Jan. 2005.
- [35] Y. Zhang, G. Zhou, Q. Zhao, J. Jin, X. Wang, and A. Cichocki, "Spatial-temporal discriminant analysis for ERP-based brain-computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 2, pp. 233–243, Mar. 2013.
- [36] M. Schreuder, B. Blankertz, and M. Tangermann, "A new auditory multiclass brain-computer interface paradigm: Spatial hearing as an informative cue," *PLoS ONE*, vol. 5, no. 4, Apr. 2010, Art. no. e9813.
- [37] J. Höhne, M. Schreuder, B. Blankertz, K. R. Mueller, and M. Tangermann, "Novel paradigms for auditory P300 spellers with spatial hearing: Two online studies," in *Frontiers Comput. Neurosci.*, vol. 44, p. 55, Mar. 2010.
- [38] J. Höhne, "A novel 9-class auditory ERP paradigm driving a predictive text entry system," *Frontiers Neurosci.*, vol. 5, p. 99, Oct. 2011.
- [39] (2020). *BNCI Horizon*. [Online]. Available: <http://bnci-horizon-2020.eu/>
- [40] Q. Xu, H. Zhou, Y. Wang, and J. Huang, "Fuzzy support vector machine for classification of EEG signals using wavelet-based features," *Med. Eng. Phys.*, vol. 31, no. 7, pp. 858–865, Sep. 2009.
- [41] A. Schlögl, F. Lee, H. Bischof, and G. Pfurtscheller, "Characterization of four-class motor imagery EEG data for the BCI-competition 2005," *J. Neural Eng.*, vol. 2, no. 4, pp. L14–L22, Dec. 2005.
- [42] M. A. Lopez-Gordo, E. Fernandez, S. Romero, F. Pelayo, and A. Prieto, "An auditory brain-computer interface evoked by natural speech," *J. Neural Eng.*, vol. 9, no. 3, Jun. 2012, Art. no. 036013.
- [43] A. Onishi and S. Nakagawa, "How does the degree of valence influence affective auditory P300-based BCIs?" *Frontiers Neurosci.*, vol. 13, p. 45, Feb. 2019.
- [44] D. Krusienski, E. Sellers, F. Cabestaing, S. Bayouth, D. McFarland, T. Vaughan, and J. Wolpaw, "A comparison of classification techniques for the P300 speller," *J. Neural Eng.*, vol. 3, no. 4, p. 299, 2006.

- [45] Y. Zhang, G. Zhou, J. Jin, Q. Zhao, X. Wang, and A. Cichocki, "Aggregation of sparse linear discriminant analyses for event-related potential classification in brain-computer interface," *Int. J. Neural Syst.*, vol. 24, no. 01, Feb. 2014, Art. no. 1450003.
- [46] A. Rakotomamonjy and V. Guigue, "BCI competition III: Dataset II-ensemble of SVMs for BCI P300 speller," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 3, pp. 1147–1154, Mar. 2008.
- [47] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Process. Lett.*, vol. 9, no. 3, pp. 293–300, Jun. 1999.
- [48] J. R. Wolpaw, H. Ramoser, D. J. McFarland, and G. Pfurtscheller, "EEG-based communication: Improved accuracy by response verification," *IEEE Trans. Rehabil. Eng.*, vol. 6, no. 3, pp. 326–333, Dec. 1998.
- [49] C. E. Shannon, "A mathematical theory of communication," *ACM SIG-MOBILE Mobile Comput. Commun. Rev.*, vol. 5, no. 1, pp. 3–55, 2001.
- [50] C. S. Nam, J. Woo, and S. Bahn, "Severe motor disability affects functional cortical integration in the context of brain-computer interface (BCI) use," *Ergonomics*, vol. 55, no. 5, pp. 581–591, May 2012.
- [51] J. Jin, E. W. Sellers, Y. Zhang, I. Daly, X. Wang, and A. Cichocki, "Whether generic model works for rapid ERP-based BCI calibration," *J. Neurosci. Methods*, vol. 212, no. 1, pp. 94–99, Jan. 2013.
- [52] M. Huang, J. Jin, Y. Zhang, D. Hu, and X. Wang, "Usage of drip drops as stimuli in an auditory P300 BCI paradigm," *Cognit. Neurodyn.*, vol. 12, no. 1, pp. 85–94, Feb. 2018.
- [53] N. Kaongoen and S. Jo, "A novel hybrid auditory BCI paradigm combining ASSR and P300," *J. Neurosci. Methods*, vol. 279, pp. 44–51, Feb. 2017.
- [54] S. Halder, I. Käthner, and A. Kübler, "Training leads to increased auditory brain-computer interface performance of end-users with motor impairments," *Clin. Neurophysiol.*, vol. 127, no. 2, pp. 1288–1296, Feb. 2016.
- [55] J. Long, Y. Li, and Z. Yu, "A semi-supervised support vector machine approach for parameter setting in motor imagery-based brain computer interfaces," *Cognit. Neurodyn.*, vol. 4, no. 3, pp. 207–216, Sep. 2010.
- [56] L. Ding and S. Liao, "Approximate model selection for large scale LSSVM," in *Proc. Asian Conf. Mach. Learn.*, 2011, pp. 165–180.
- [57] T. Hinterberger, N. Neumann, M. Pham, A. Kübler, A. Grether, N. Hofmayer, B. Wilhelm, H. Flor, and N. Birbaumer, "A multimodal brain-based feedback and communication system," *Exp. Brain Res.*, vol. 154, no. 4, pp. 521–526, Feb. 2004.
- [58] Q. She, B. Hu, H. Gan, Y. Fan, T. Nguyen, T. Potter, and Y. Zhang, "Safe semi-supervised extreme learning machine for EEG signal classification," *IEEE Access*, vol. 6, pp. 49399–49407, 2018.
- [59] Y.-F. Li and D.-M. Liang, "Safe semi-supervised learning: A brief introduction," *Frontiers Comput. Sci.*, vol. 13, no. 4, pp. 669–676, Aug. 2019.



**SUGURU KANOGA** received the M.E. and Ph.D. degrees in engineering from Keio University, Japan, in 2014 and 2016, respectively. From 2015 to 2017, he was a Research Fellow with the Japan Society for Promotion of Science, Japan. He is currently a Researcher with the Artificial Intelligence Research Center, AIST, Tokyo, Japan. His research interests include the development of machine learning techniques and the implementation of human-machine interfaces based on machine learning techniques.



**SHIN-ICHI ITO** (Member, IEEE) received the B.E. and M.E. degrees from Tokushima University, in 2002 and 2004, respectively, and the D.E. degree from the Tokyo University of Agriculture and Technology, in 2007. He has worked with Japan Gain the Summit Company Ltd., and Tokyo University of Agriculture and Technology, as a System Engineer and a specially appointed Assistant Professor, in 2004 and 2007, respectively. From 2009 to 2019, he was an Assistant Professor

with Tokushima University. Since 2020, he has been an Associate Professor with Tokushima University. His current research interests include EEG analysis, bio-signal processing, and information visualization. He is a member of IEICE, JSMBE, and IEEJ.



**YASUE MITSUKURA** (Member, IEEE) received the Dr.Eng. degree from the University of Tokushima and the Dr.Med. degree from Keio University, Kanagawa, Japan. From 2007 to 2010, she was an Associate Professor with the Tokyo University of Agriculture and Technology. Since 2011, she has been an Associate Professor, and is currently a Professor with Keio University. Her research interests include bio-signal analysis (EEG, EMG, EOG, ECG, GSR, and body temperature, breath, salivary amylase, NIRS, and fMRI) and brain-computer interfaces. She is a member of SICE, IEEJ, and RISP.

...



**MIKITO OGINO** received the M.E. degree from Keio University, Japan, in 2015, where he is currently pursuing the Ph.D. degree in engineering. From 2015 to 2017, he was a Software Engineer with NEC Corporation. He is currently the Chief Scientist of Dentsu ScienceJam Inc., Tokyo, Japan. His research interests include auditory brain-computer interfaces, estimation of human emotions and status from EEG, and development of practical and portable applications.