PAPER

Movement-Imagery Brain-Computer Interface: EEG Classification of Beta Rhythm Synchronization Based on Cumulative Distribution Function

Teruyoshi SASAYAMA[†,††a)], ***Student Member and*** **Tetsuo KOBAYASHI**[†], ***Member***

SUMMARY We developed a novel movement-imagery-based brain-computer interface (BCI) for untrained subjects without employing machine learning techniques. The development of BCI consisted of several steps. First, spline Laplacian analysis was performed. Next, time-frequency analysis was applied to determine the optimal frequency range and latencies of the electroencephalograms (EEGs). Finally, trials were classified as right or left based on β-band event-related synchronization using the cumulative distribution function of pretrigger EEG noise. To test the performance of the BCI, EEGs during the execution and imagination of right/left wrist-bending movements were measured from 63 locations over the entire scalp using eight healthy subjects. The highest classification accuracies were 84.4% and 77.8% for real movements and their imageries, respectively. The accuracy is significantly higher than that of previously reported machine-learning-based BCIs in the movement imagery task (paired t-test, $p < 0.05$). It has also been demonstrated that the highest accuracy was achieved even though subjects had never participated in movement imageries.

key words: *electroencephalogram (EEG), brain-machine interface (BCI), event-related synchronization (ERS), spline Laplacian, Hilbert transform*

1. Introduction

A brain-computer interface (BCI), also known as a brain-machine interface (BMI), is a system that enables communication using signals from the brain [1]. It is known that α (8–13 Hz) and β (> 16 Hz) rhythms decrease (event-related desynchronization, ERD) and increase (event-related synchronization, ERS) when people execute or imagine movements, e.g., bending the wrist and tapping the foot, and the sources of the rhythms depend on which body part they move or imagine [2].

A number of movement-imagery-based BCIs have been reported [3]–[10]. Most movement-imagery-based BCIs classify trials by extracting the ERD and/or ERS of α and β rhythms during right- and left-movement imageries. However, the change in the α rhythm is relatively slow compared with that in the β rhythm [2], [11]. This may be an obstacle in improving the information transfer rate (ITR).

To classify right- and left-hand-movement imageries based on ERD/ERS in β rhythm, the majority of recent BCIs use machine learning techniques, such as a support vector machine (SVM) [4], [5]. However, the machine-learning-based BCIs require many training data sets to learn, and we cannot obtain sufficient data sets easily due to the exhaustion of subjects during long training sessions. In addition, untrained subjects may not always execute trials properly, especially in the movement imageries. As a result, the classifier may learn inappropriate trials and classify trials improperly.

Common spatial patterns (CSPs) are among the most successful spatial filters for obtaining features of ERD and ERS for machine learning [6]. Although BCIs that are based on CSP achieve high accuracy in some cases, CSPs are disadvantageous in that they require a large number of electrodes to obtain good results [12]. Moreover, they do not work well in the presence of large artifacts [13]. BCIs requiring a smaller number of electrodes can be more robust and allow quicker and easier preparation.

Another method used to classify trials is the extraction of β ERS after movement imagery [3]. This is based on the fact that β rhythm exhibits a larger and faster change after the movements or their imageries in ERS than during the movements or their imageries in ERD [2], [11]. Therefore, we focused only on β ERSs after the movements or their imageries to classify them into right or left. This may allow BCIs in this study to achieve higher values of ITR than those reported previously.

The main objective of this study is to develop movement-imagery-based BCIs for untrained subjects without employing machine learning techniques. Instead, we use the neurophysiological information only where and when β ERS significantly increases. Thresholds for the significant increase were determined based on the cumulative distribution function (CDF) of the measured EEG noise.

2. Methods

2.1 Subjects

Eight right-handed male subjects (aged 21–24 years) participated in the experiments. Informed consent was obtained from all the subjects. The procedures described here were approved by the local institutional ethics committee. Subjects had never participated in movement or movement im-

Manuscript received November 18, 2010.
Manuscript revised June 22, 2011.

[†]The authors are with the Biomedical Function Engineering Laboratory, Department of Electrical Engineering, Graduate School of Engineering, Kyoto University, Kyoto-shi, 615–8510 Japan.

[††]The author is with the Japan Society for the Promotion of Science, Tokyo, 102–8471 Japan.

a) E-mail: sasayama@bfe.kuee.kyoto-u.ac.jp

DOI: 10.1587/transinf.E94.D.2479

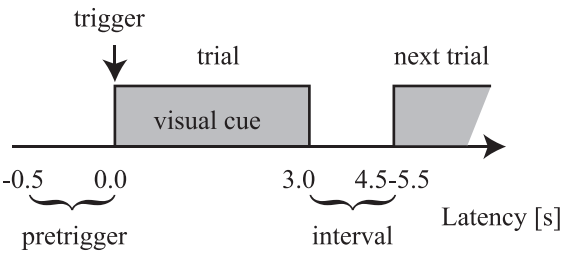

Fig. 1 Timing of movement or its imagery task.

agery experiments.

2.2 Task

The subjects sat in a chair in front of a 21" monitor at a distance of 0.50 m. Figure 1 shows an overview of the task procedure. A fixation point was constantly presented on the monitor and the subjects stared at it. The cues, "right hand" or "left hand," appeared for 3.0 s. After this, the cues disappeared for 1.5–2.5 s. This was defined as one trial and 50 trials was defined as one run. The subjects were instructed to execute or imagine brisk wrist-bending movements following the visual cues indicating the right or left hand. One session consisted of 4 kinds of runs: right hand movement, left hand movement, right hand movement imagery, and left hand movement imagery. The subjects performed two session. Thus, the total number of trials was 400. A short break was given after a run.

2.3 Data Set

In offline analysis, trials were split into two data sets: a training set (first half of trials), which were labeled as left hand and right hand, and an unlabeled test set (last half of trials).

2.4 EEG Recordings

A 96-channel EEG system (Bio-logic, Inc., USA) was used to record EEGs from 63 locations on the entire scalp using eight healthy subjects (Fig. 2). The arrangement of the electrodes followed the international 10–20 electrode system. All electrode impedances were kept below 10 kΩ. Reference electrodes were placed on the right and left earlobes (REF_R, REF_L) and the grounded electrode was placed on the left mastoid. Electro-oculograms (EOGs) were also measured to detect eye blinks with electrodes placed in the vicinity of both eyes (EOGL1, EOGL2, EOGR2, and FP1 in Fig. 2: FP1 was also used as an EEG electrode). The EEGs and EOGs were band-pass filtered (0.1–100 Hz) and sampled at 256 Hz.

2.5 Data Processing

Figure 3 shows the flow chart of data processing and classification. Details are as follows.

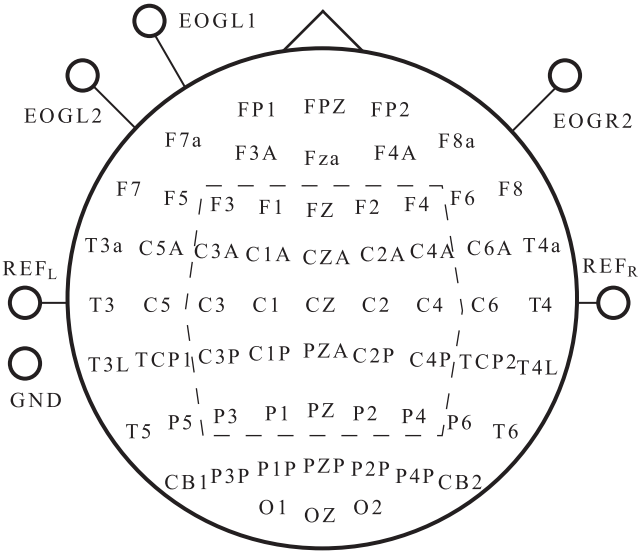

Fig. 2 Arrangement of electrodes. Reference electrodes were placed on the right and left earlobes (REF_R, REF_L) and the grounded electrode (GND) was placed on the left mastoid. EOGL1, EOGL2, EOGR2, and FP1 were EOG electrodes: FP1 was also used as an EEG electrode. The squares show the region for classification. Electrodes, which are used by CSPs, are surrounded with a dashed line.

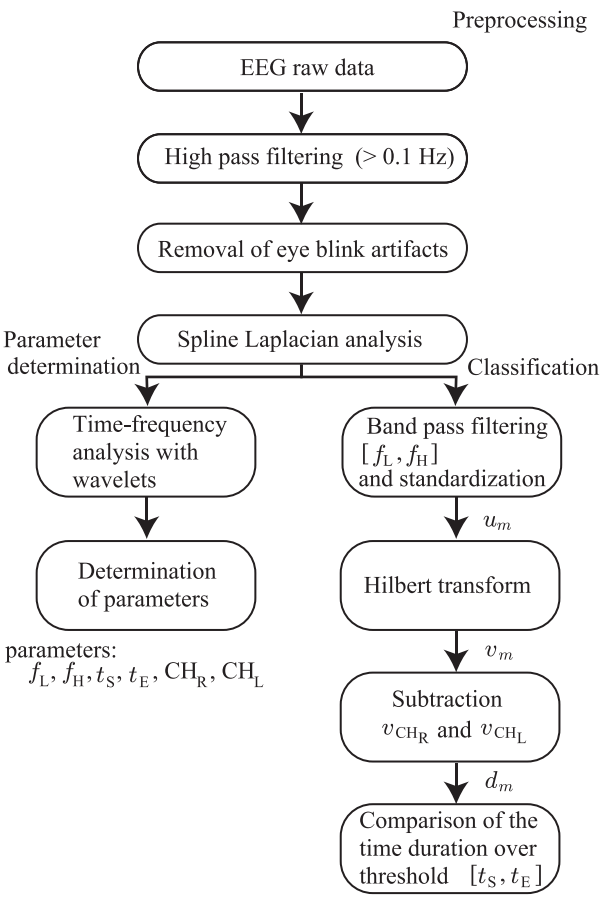

Fig. 3 Flow chart of preprocessing, parameter determination, and classification.

2.5.1 Preprocessing

First, the data were high-pass filtered (0.1 Hz) to eliminate long-lasting drift components. In this study, trials including EOGs over $\pm 150\,\mu$V were assumed to contain eye blink artifacts. Second, the artifacts were eliminated using an automatic elimination method for eye blink artifacts [14], which used both principal component analysis (PCA) and independent component analysis (ICA). Third, we performed spline Laplacian (SL) analysis [15] and obtained signals s_k, where k is a subscript indicating an EEG channel name ($k \in \{\mathrm{FP1}, \mathrm{FPZ}, \cdots, \mathrm{O2}\}$). In SL analysis, the Laplacian of the potentials at all surface locations is estimated by spline functions instead of local potentials. The physical meaning of the Laplacian operation is the calculation of the gradient of the radial surface current density entering the scalp through the skull [16].

2.5.2 Parameter Determination

Time-frequency analysis with wavelets [17] was applied to the preprocessed data of the training set to determine the optimal parameters, such as latency duration $[t_\mathrm{S}, t_\mathrm{E}]$ and frequency range $[f_\mathrm{L}, f_\mathrm{H}]$ of the EEGs. An advantage of wavelet analysis, compared with the short-time fast Fourier transform (FFT), is that each frequency component could be adjusted to its scale. We used the complex Morlet wavelet,

$$\Psi(t, f) = A \cdot \exp\left(-\frac{t^2}{2\sigma_t^2}\right) \times \exp\left(2\pi i f t\right), \tag{1}$$

where $A = \left(\sigma_t \sqrt{\pi}\right)^{-1/2}$, $2\pi\sigma_t \cdot f = 7$, and $i = \sqrt{-1}$. The time-variant energy $E_k(t, f)$ of the signal $s_k(t)$ in a frequency band around f is the squared norm of the convolution of a complex wavelet with the signal

$$E_k(t, f) = |\Psi(t, f) * s_k(t)|^2. \tag{2}$$

The asterisk denotes convolution and $|\cdot|$ denotes the absolute value.

$E_k(t, f)$ were averaged over all trials. The obtained mean value was defined as $\bar{E}_k(t, f)$. We assumed that the EEGs during the pretrigger period were white Gaussian noise and were independent and identically distributed (i.i.d.). On the basis of this assumption, the probability density of $s_k(t)$ was considered to be a normal distribution, and that of $\bar{E}_k(t, f)$ was also a normal distribution if the number of samples compared to the mean was large enough, i.e., if the number of trials was large enough. $\bar{E}_k(t, f)$ was standardized by the mean $\mu_k(f)$ and the standard deviation (S.D.) $\sigma_k(f)$ during the pretrigger duration is

$$z_k(t, f) = \frac{\bar{E}_k(t, f) - \mu_k(f)}{\sigma_k(f)}. \tag{3}$$

In this study, $z_k(t, f)$ was calculated, where $f = 1, 2, \cdots, 60$. The parameters f_L, f_H, t_S, and t_E were determined by calculating the area over the significant increase ($z_k(t, f) > 2.58$, $p < 0.005$) in the z_k maps.

To analyze in a certain band $[f_\mathrm{L}, f_\mathrm{H}]$, we introduced the index

$$Z_k(t; f_\mathrm{L}, f_\mathrm{H}) = \frac{1}{f_\mathrm{H} - f_\mathrm{L} + 1} \sum_{f_\mathrm{L} \le f \le f_\mathrm{H}} z_k(t, f). \tag{4}$$

This index was depicted by a topographical method. The channel that indicated the most significant change was detected from time-frequency analysis. The channels that showed the most significant change during right- and left-hand movements or their mental simulation were defined as $\mathrm{CH_R}$ and $\mathrm{CH_L}$, respectively ($\mathrm{CH_R}, \mathrm{CH_L} \in \{\mathrm{FP1}, \mathrm{FP2}, \cdots, \mathrm{O2}\}$).

2.5.3 Classification

The EEGs of the test data set analyzed with SL analysis were band-pass filtered ($f_\mathrm{L} - f_\mathrm{H}$), and we obtained $\tilde{s}_m(t)$ ($m \in \{\mathrm{CH_R}, \mathrm{CH_L}\}$). We assume $\tilde{s}_m(t)$ in the pretrigger duration was a white Gaussian noise. Each $\tilde{s}_m(t)$ is standardized with the mean, $\tilde{\mu}_m$, and the S.D., $\tilde{\sigma}_m$, in the pretrigger duration as

$$u_m(t) = \frac{\tilde{s}_m(t) - \tilde{\mu}_m}{\tilde{\sigma}_m}, \tag{5}$$

and we obtained $u_m(t)$. After this, their envelopes were calculated using a Hilbert transformation [18], which has widely been used to obtain envelopes in other BCIs [19], [20]. The Hilbert transform $\hat{u}_m(t)$ or $\mathcal{H}$ as a function of time $u_m(t)$ was defined by

$$\mathcal{H}[u_m(t)] = \hat{u}_m(t) = \frac{1}{\pi} \mathcal{P} \int_{-\infty}^{\infty} \frac{u_m(\tau)}{t - \tau}\, \mathrm{d}\tau, \tag{6}$$

where $\mathcal{P}$ denotes the Cauchy principal value. The amplitude of the envelope of $u_m(t)$ is

$$v_m(t) = \sqrt{u_m^2(t) + \hat{u}_m^2(t)}. \tag{7}$$

Next we calculated the differential signals of $v_{\mathrm{CH_R}}(t)$ and $v_{\mathrm{CH_L}}(t)$, and obtained $d_{\mathrm{CH_R}}(t) = v_{\mathrm{CH_R}}(t) - v_{\mathrm{CH_L}}(t)$, $d_{\mathrm{CH_L}}(t) = v_{\mathrm{CH_L}}(t) - v_{\mathrm{CH_R}}(t)$.

Finally, we compared the total durations in which $d_{\mathrm{CH_R}}$ and $d_{\mathrm{CH_L}}$ exceeded their thresholds, as explained in the following section, during the latency $[t_\mathrm{S}, t_\mathrm{E}]$. The channel that indicated a longer duration was defined as the significant change channel. If the significant change channel was $\mathrm{CH_R}$, the trial was assumed to be a right-hand movement or its imagery.

If both signals of $\mathrm{CH_R}$ and $\mathrm{CH_L}$ do not exceed their thresholds, the trial is classified into 'neither'. Therefore, BCI-CDF classify trials into three classes (right, left, neither).

2.5.4 Setting of Thresholds

The probability density function (PDF) of the white Gaussian noise is the normal distribution function, $\mathcal{N}(x\,|\,\mu, \sigma^2)$,

where μ is the mean and σ is the S.D. The PDF of an envelope of the white Gaussian noise ($x \sim \mathcal{N}(0, \sigma^2)$) is the Rayleigh distribution, defined as

$$f_X(x) = \begin{cases} \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right), & \text{for } x \geq 0, \\ 0, & \text{for } x < 0. \end{cases} \quad (8)$$

In what follows, we assumed the signal was standardized, i.e., $\sigma = 1$. We derived the PDF of the differential signal of two envelopes of white Gaussian noises ($x \sim \mathcal{N}(0, 1)$), as follows:

$$\begin{aligned} f_{X-X}(x) &= f_X(x) * f_X(-x) \\ &= \int_{-\infty}^{\infty} f_X(\tau + x) f_X(\tau)\, \mathrm{d}\tau, \end{aligned} \quad (9)$$

and we finally obtain

$$\begin{aligned} f_{X-X}(x) = \exp\left(-\frac{x^2}{4}\right)\Bigg[\frac{|x|}{4}\exp\left(-\frac{x^2}{4}\right) \\ + \left(-\frac{\sqrt{\pi}x^2}{8} + \frac{\sqrt{\pi}}{4}\right)\mathrm{erfc}\left(\frac{|x|}{2}\right)\Bigg]. \end{aligned} \quad (10)$$

The function $\mathrm{erfc}(x)$ is the complementary error function. Refer the Appendix for more details on the derivation of $f_{X-X}(x)$.

We defined the CDFs of $\mathcal{N}(x\,|\,\mu, \sigma^2)$, $f_X(x)$, $f_{X-X}(x)$ as $\Phi(x\,|\,\mu, \sigma^2)$, $F_X(x)$, $F_{X-X}(x)$, respectively. To calculate percentage points, we defined the inverse functions of $\Phi(x\,|\,\mu, \sigma^2)$, $F_X(x)$, $F_{X-X}(x)$ as $\Phi^{-1}(\alpha\,|\,\mu, \sigma^2)$, $F_X^{-1}(\alpha)$, $F_{X-X}^{-1}(\alpha)$, respectively, where α is the significant level and $0 \leq \alpha \leq 1$.

In this study, we developed a BCI, BCI-CDF (BCI using CDFs of differential signals of the envelopes, $d_m(t)$). Because α was set to 0.01, which was chosen from the significant level commonly used in statistics, the thresholds of each signal, $u_m(t)$, $v_m(t)$, and $d_m(t)$, were determined as follows:

(i) $u_m(t)$: $\begin{cases} \Phi^{-1}(1 - \frac{\alpha}{2}\,|\,0, 1) = 2.58 \\ \Phi^{-1}(\frac{\alpha}{2}\,|\,0, 1) = -2.58 \end{cases}$
(ii) $v_m(t)$: $F_X^{-1}(1 - \alpha) = 3.03$
(iii) $d_m(t)$: $F_{X-X}^{-1}(1 - \alpha) = 2.43$

2.6 Comparison with Other Methods

To compare BCI-CDF with previously reported BCIs that used machine learning techniques, we also classified the test data set as follows. First, eye blink artifacts were removed and then band-pass filtering was applied. After this, feature vectors were obtained from the preprocessed test data set using CSPs. In this study, the six most important CSPs were used, that is, six-dimension vectors were obtained using CSPs. Next, the feature vectors were classified using SVM [4] (BCI-SVM). Our implementation of SVM was based on the LIBSVM library [21].

We selected electrodes as shown in Fig. 2 when using CSPs to avoid noisy channels for improving classification accuracy, whereas all EEG electrodes were used in the preprocessing. Weights of the scalp CSP maps were large in the outmost region of the electrode configuration. As described in a previous study [10], this is considered to be neurophysiologically implausible. Actually, the accuracy was chance level when we used all electrodes for CSPs.

Table 1 The channel, latency, and frequency band, where the most significant changes were observed.

Subject	Task	Channel (CH_R , CH_L)	Latency [ms]	Frequency [Hz]
S1	Movement	C1P, C2P	770-3000	16-40
	Imagery	C1P, C2P	850-3000	16-37
S2	Movement	C1P, C2P	930-3000	16-37
	Imagery	C1P, C2P	940-3000	16-37
S3	Movement	C1, C2	960-3000	18-32
	Imagery	C1, C2	2160-3000	16-36
S4	Movement	C1P, C4P	1000-3000	16-34
	Imagery	C1P, C4A	2000-3000	16-34
S5	Movement	C1, C4	990-3000	16-22
	Imagery	C3P, C4	570-3000	16-26
S6	Movement	C3, C2	820-3000	22-26
	Imagery	C3, C2	1720-3000	17-32
S7	Movement	C1P, C4	1290-3000	20-33
	Imagery	CZ, C2P	1290-3000	17-22
S8	Movement	C1P, C2	820-3000	16-37
	Imagery	CZ, C2	620-3000	16-33

Table 2 Number of accurately classified trials (classified trials) and ITR. The unit of ITR is bit per minute.

Subject	Task	BCI-CDF		BCI-SVM	
		Acc.	ITR	Acc.	ITR
S1	Movement	81 (96)	4.32	83 (97)	4.93
	Imagery	63 (81)	2.29	55 (100)	0.09
S2	Movement	71 (94)	2.23	74 (99)	2.19
	Imagery	60 (83)	1.48	46 (100)	0.06
S3	Movement	66 (85)	2.38	56 (100)	0.12
	Imagery	50 (68)	1.36	53 (100)	0.03
S4	Movement	66 (94)	1.37	66 (100)	0.90
	Imagery	53 (79)	0.82	66 (100)	0.90
S5	Movement	43 (65)	0.60	55 (98)	0.13
	Imagery	44 (64)	0.80	43 (100)	0.17
S6	Movement	42 (65)	0.49	51 (98)	0.01
	Imagery	42 (63)	0.62	52 (100)	0.01
S7	Movement	58 (94)	0.45	67 (100)	1.02
	Imagery	40 (69)	0.15	63 (100)	0.59
S8	Movement	57 (90)	0.56	55 (99)	0.09
	Imagery	43 (74)	0.17	52 (100)	0.01

ITR is often used to compare the performances among BCIs and measure system improvements [22]. The ITR in bits per trial is defined as follows:

$$B = \log_2(N_c) + p\log_2(p) + (1 - p)\log_2\frac{1 - p}{N_c - 1}, \quad (11)$$

where p represents the accuracy and N_c is the number of classes (In this study, $N_c = 2$). When we calculate ITR, we assumed that the time of the trial is 5.5 s and that B is 0 for trials that are not classified.

3. Results

Figure 4 illustrates the time-frequency analysis and the SL

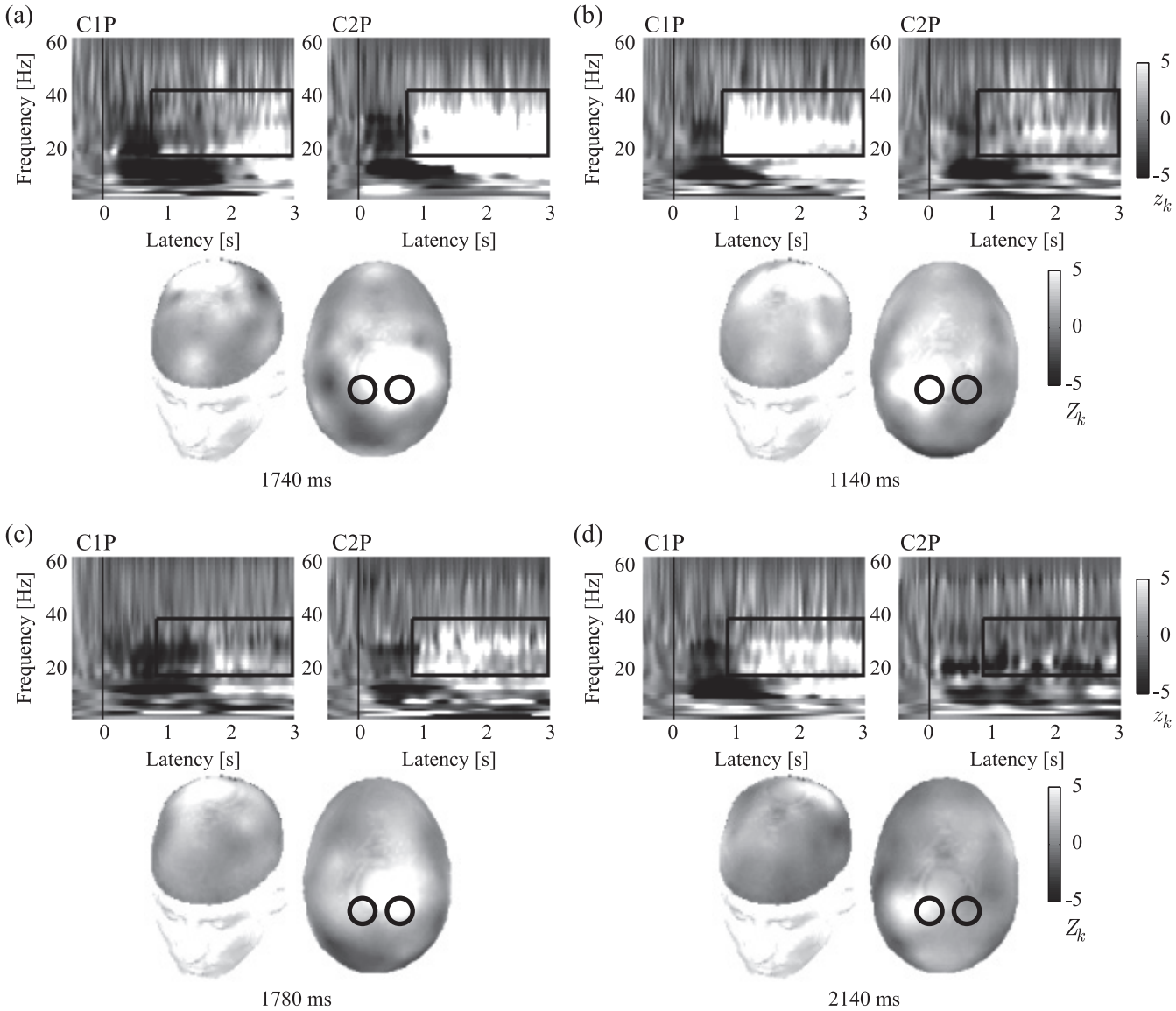

Fig. 4 Time-frequency maps $[z_k(t, f)]$ and the topography of $Z_k(t; f_L, f_H)$ at the peak latency in the condition of (a) left-hand movement, (b) right-hand movement, (c) left-hand-movement imagery, and (d) right-hand-movement imagery in subject 1. The range of latency and frequency for classification is surrounded by a black rectangle. Channels C1P and C2P are each surrounded by a black circle.

maps at the peak latency of the β-band ERS in subject 1. A strong β ERS was observed in the vicinity of the contralateral motor cortex in all the subjects after they executed right/left wrist-bending movements or their imageries. The channels showing the most significant ERS in subject 1 were C1P and C2P in the left and right hemispheres, respectively. Table 1 shows the channels, latency duration, and frequency band where the most significant changes were observed. The channels in the movements and their imageries were the same in subjects 1, 2, 3, and 6. In contrast, the channels in the movements and their imageries were different in subjects 4, 5, 7, and 8.

Time courses of $u_m(t)$, $v_m(t)$, and $d_m(t)$ in a right-hand-movement task in subject 1 are shown in Fig. 5. The total duration in which $u_{\mathrm{CH_R}}(t)$, $v_{\mathrm{CH_R}}(t)$, and $d_{\mathrm{CH_R}}(t)$ exceeded their thresholds was longer than those of $u_{\mathrm{CH_L}}(t)$, $v_{\mathrm{CH_L}}(t)$, and $d_{\mathrm{CH_L}}(t)$, respectively. Therefore, BCI-CDF can classify this trial as "right." Both $u_{\mathrm{CH_R}}(t)$ and $u_{\mathrm{CH_L}}(t)$, or both $v_{\mathrm{CH_R}}(t)$ and $v_{\mathrm{CH_L}}(t)$, exceeded the thresholds, whereas $d_{\mathrm{CH_R}}(t)$ exceeded but $d_{\mathrm{CH_L}}(t)$ did not exceed the threshold. Because the signals are expected to change significantly only at the observed channel (in this case, $\mathrm{CH_R}$), $d_m(t)$ is a better index than $u_m(t)$ or $v_m(t)$ for the classifier.

The classification accuracy using the test data set is shown in Fig. 6. The difference in the mean accuracy of BCI-CDF and BCI-SVM may appear to be small, especially with movement. However, the mean accuracy with BCI-CDF is significantly higher than that with BCI-SVM ($p < 0.05$). The highest accuracy for real movements in all the subjects was 84.4% and that for their imagery was

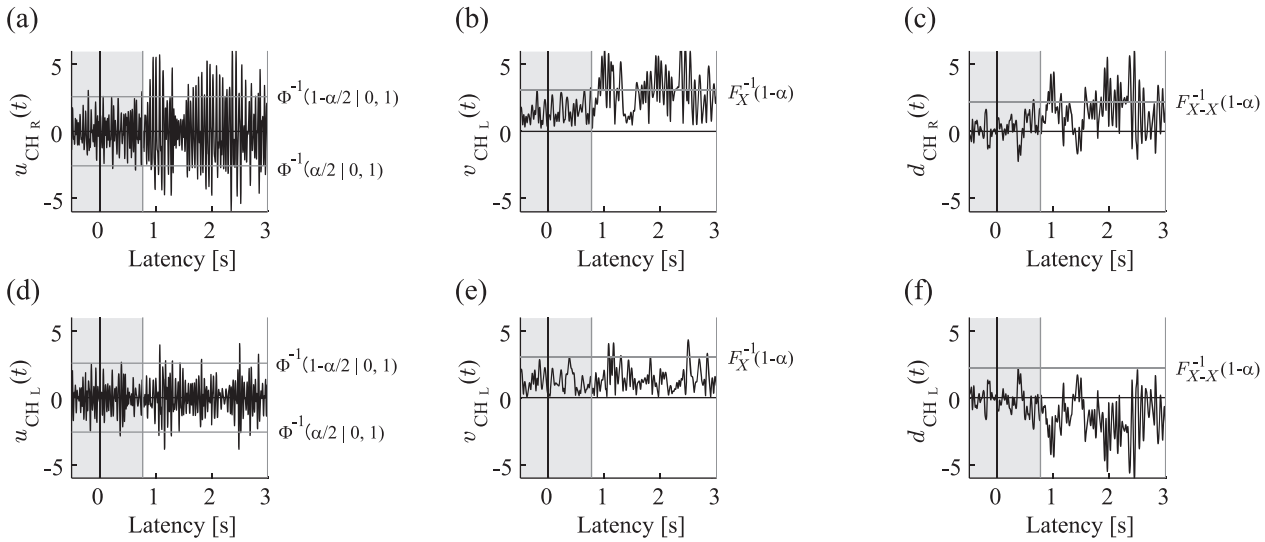

Fig. 5 Time courses of band-pass filtered waves, (a) $u_{\mathrm{CH_R}}(t)$, (d) $u_{\mathrm{CH_L}}(t)$, enveloped waves, (b) $v_{\mathrm{CH_R}}(t)$, (e) $v_{\mathrm{CH_L}}(t)$, and differential waves, (c) $d_{\mathrm{CH_R}}(t)$, (f) $u_{\mathrm{CH_L}}(t)$ in a right-hand movement in subject 1. Horizontal gray lines indicate the threshold ($\alpha = 0.01$). The latency duration $[t_S, t_E]$ is the distance between two vertical gray lines. The total duration in which each of the waves of $\mathrm{CH_R}$ (upper row) exceeds the threshold is longer than that of $\mathrm{CH_L}$ (lower row).

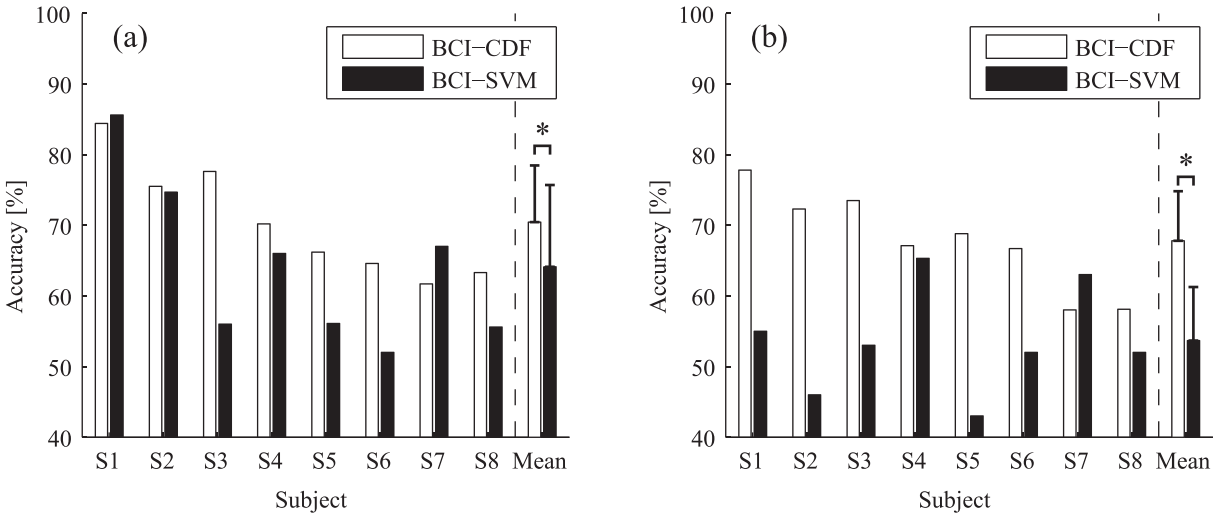

Fig. 6 Means of classification accuracy in (a) real movement and (b) their imageries. An asterisk (*) denotes significant difference (paired t-test, $p < 0.05$).

77.8% with BCI-CDF. The accuracies in subjects 7 and 8 were lower (about 60%) than those in the other three subjects (about 70% or above) in the movement imageries. This suggests that the modulation of the β rhythm in subjects 7 and 8 is smaller than that in other subjects.

Table 2 shows the number of accurately classified trials (classified trials) and the ITR. The BCI-SVM classifies all trials into two classes (the number of classified BCI-SVM trials was not 100 because noisy trials were removed from the classified trials). In contrast, BCI-CDF does not classify all trials. The ITR results are also shown in Table 2. The mean ITR with BCI-CDF is higher than that with BCI-SVM ($p < 0.05$).

4. Discussion

BCI-CDF does not use the absolute value of signals but classifies them on the basis of total duration in which they exceed the thresholds. Even if large spike noises were contained in the EEGs in $\mathrm{CH_R}$ or $\mathrm{CH_L}$, BCI-CDF can classify them precisely. On the other hand, a classifier based on the absolute value of signals may lead to incorrect results when large spike noises occur in the EEGs in $\mathrm{CH_R}$ or $\mathrm{CH_L}$.

Moreover, BCI-CDF considers differential signals, $d_{\mathrm{CH_R}}(t)$ and $d_{\mathrm{CH_L}}(t)$. If the total duration in which long-lasting noise, e.g., eye movement artifacts, exceeds the threshold, a classifier based only on $u_m(t)$ or $v_m(t)$ may

give incorrect results. This is because the total duration in which EEGs from brain activity exceed the threshold becomes shorter than that of noises and actual activity could be ignored. In this situation, the differential signals in two channels allow the noises to cancel each other, and thus the classifier's accuracy may improve.

In this study, although we used only two channels, i.e. CH_R and CH_L, for the classification, BCI-CDF requires many EEG electrodes to eliminate eye-blink artifacts and perform SL analysis. However, a method for eliminating the artifacts in only a small number of EEG electrodes has been reported [23], and the number of electrodes needed to perform SL analysis can be reduced by localizing electrodes in the vicinity of the sensorimotor cortex. BCI-SVM with CSP can mathematically classify trials using two channels. However, as described in the Introduction, CSP requires a large number of electrodes to obtain good results [12]. If BCI-SVM classifies trials using only two channels, the accuracy decreases. In contrast, the accuracy of BCI-CDF is significantly higher than that of BCI-SVM although BCI-CDF classifies trials having fewer channels than BCI-SVM.

BCI-CDF adopts the parameters of optimal frequency band, channels, and latency that were identified from the training data. In this sense, BCI-CDF may be considered for adopting machine learning techniques, especially when the parameters are identified automatically. However, BCI-CDF adopts parameters that correspond to neurophysiological information. In contrast, SVM adopts parameters that do not directly correspond to neurophysiological information, suggesting that we cannot explain the role played by each parameter. This is an obstacle in improving accuracy.

CSP, which is adopted by BCI-SVM for feature extraction, depends heavily on the identified latency because it calculates the covariance matrix during latency. However, for a given subject in different trials, the latency when ERS is observed could differ. This results in misclassification. In contrast, BCI-CDF does not require the latency duration for classification because the threshold is determined such that it does not exceed the maximum value during the latency period in which no significant ERS is observed. In fact, the waves of CH_R and CH_L scarcely exceed the threshold during the latency period that is not identified, as shown in Fig. 5. Although BCI-CDF may perform accurately without a reason to identify latency, BCI-CDF should adopt the identified latency to improve robustness for artifacts.

BCI-SVM classifies all trials into two classes whereas BCI-CDF does not classify all trials. This is an important characteristic of BCI-CDF that they classify trials into three classes (right, left, neither) unlike BCI-SVM into two classes (right, left). In other words, BCI-CDF detects trials that are difficult for classification and do not classify the trials. Therefore, BCI-CDF is useful for subjects to train how to modulate β rhythms for improving accuracy. The thresholds are determined by the noise level only. This suggests that we can extend BCI-CDF to an asynchronous BCI, which does not require any stimuli, e.g., visual cues, for the classification.

BCI-SVM may improve accuracy to output "neither", i.e., to adopt three-class classifier such as BCI-CDF. One possible way to find "neither" trial is to use threshold for posterior probability in each class. However, the issue concerns how to determine the threshold on the posterior probability. In BCI-CDF, the threshold is determined by CDF and the significant level.

One reason for the improved accuracy in BCI-CDF is to select trials for classification. However, note that this is not the only reason for improved accuracy. There are some cases for which the number of accurately classified trials by BCI-CDF exceeds that by BCI-SVM even after removing some of the bad trials. Moreover, when we consider the removal of bad trials, one half of the removed trials must be subtracted from all of the accurately classified trials because the accuracy of the removed trials is assumed to be 50%. However, that is not necessarily the case. Therefore, the accuracy improvement in BCI-CDF was not accomplished solely by the removal of bad trials.

5. Conclusion

Our results demonstrated that the developed BCI (BCI-CDF) achieved higher classification accuracy than that of previously reported machine-learning-based BCI (BCI-SVM), especially for the movement imageries (paired t-test, $p < 0.05$). This may be because the machine-learning-based BCI classified trials into two classes even though improperly executed trials may have existed in the movement imageries. In contrast, BCI-CDF classified trials into three classes.

It demonstrated that the highest accuracy was achieved even though the subjects had never participated in the movement and their imagery experiments. In future work, we will develop a real-time feedback system to achieve high accuracy using our BCI.

Acknowledgment

The authors would like to thank Dr. J. Jung for his technical expertise and suggestions. This work was supported by a Grant-in-Aid for JSPS Fellows ($21 \cdot 5335$).

References

[1] 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, June 2002.
[2] G. Pfurtscheller and F.H. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: Basic principles," Clin. Neurophysiol., vol.110, no.11, pp.1842–1857, Nov. 1999.
[3] G. Pfurtscheller and T. Solis-Escalante, "Could the beta rebound in the EEG be suitable to realize a "brain switch"?," Clin. Neurophysiol., vol.120, no.1, pp.24–29, Jan. 2009.
[4] X. Liao, D. Yao, D. Wu, and C. Li, "Combining spatial filters for the classification of single-trial EEG in a finger movement task," IEEE Trans. Biomed. Eng., vol.54, no.5, pp.821–831, May 2007.
[5] 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, Sept. 2009.

[6] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," Clin. Neurophysiol., vol.110, no.5, pp.787–798, May 1999.
[7] J. Lehtonen, P. Jylanki, L. Kauhanen, and M. Sams, "Online classification of single EEG trials during finger movements," IEEE Trans. Biomed. Eng., vol.55, no.2, pp.713–720, Feb. 2008.
[8] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Müller, V. Kunzmann, F. Losch, and G. Curio, "The Berlin brain–computer interface: EEG-based communication without subject training," IEEE Trans. Neural Syst. Rehabil. Eng., vol.14, no.2, pp.147–152, June 2006.
[9] C. Vidaurre, A. Schlogl, R. Cabeza, R. Scherer, and G. Pfurtscheller, "Study of on–line adaptive discriminant analysis for EEG-based brain computer interfaces," IEEE Trans. Biomed. Eng., vol.54, no.3, pp.550–556, March 2007.
[10] X. Lei, P. Yang, and D. Yao, "An empirical bayesian framework for brain–computer interfaces," IEEE Trans. Neural Syst. Rehabil. Eng., vol.17, no.6, pp.521–529, Dec. 2009.
[11] C. Neuper and G. Pfurtscheller, "Evidence for distinct beta resonance frequencies in human EEG related to specific sensorimotor cortical areas," Clin. Neurophysiol., vol.112, no.11, pp.2084–2097, Nov. 2001.
[12] C. Guger, H. Ramoser, and G. Pfurtscheller, "Real-time EEG analysis with subject-specific spatial patterns for a brain-computer interface (BCI)," IEEE Trans. Rehabil. Eng., vol.8, no.4, pp.447–456, Dec. 2000.
[13] H. Ramoser, J. Müller-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.
[14] Y. Okada, J. Jung, and T. Kobayashi, "An automatic identification and removal method for eye blink artifacts in event-related magnetoencephalographic measurements," Physiol. Meas., vol.28, no.12, pp.1523–1532, Oct. 2007.
[15] S.K. Law, P.L. Nunez, and R.S. Wijesinghe, "High-resolution EEG using spline generated surface Laplacians on spherical and ellipsoidal surfaces," IEEE Trans. Biomed. Eng., vol.40, no.2, pp.145–153, Feb. 1993.
[16] P.L. Nunez, R.B. Siberstein, P.J. Cadusch, R.S. Wijesinghe, A.F. Westdrop, and R. Srinivasan, "A theoretical and experimental study of high resolution EEG based on surface Laplacians and cortical imaging," Electroencephalogr. Clin. Neurophysiol., vol.90, no.1, pp.40–57, Jan. 1994.
[17] J.-P. Lachaux, A. Luts, D. Rudrauf, D. Cosmelli, M.L.V. Quyen, J. Martinerie, and D. Varela, "Estimating the time-course of coherence between single-trial brain signals: An introduction to wavelet coherence," Neurophysiol. Clin., vol.32, no.3, pp.157–174, June 2002.
[18] N. Thrane, "The Hilbert transform," B&K Technical Rewiew, no.3, 1984.
[19] A. Médl, D. Flotzinger, and G. Pfurtscheller, "Hilbert-transform based predictions of hand movements from EEG measurements," Proc. An. Int. Conf. IEEE on Engineering in Medicine and Biology Society, vol.14, pp.2539–2540, Oct. 1992.
[20] J. Mellinger, G. Schalk, C. Braun, H. Preissl, W. Rosenstiel, N. Birbaumer, and A. Kubler, "An MEG-based brain-computer interface (BCI)," Neuro Image, vol.36, no.3, pp.581–593, July 2007.
[21] C.-C. Chang and C.-J. Lin, LIBSVM: A library for support vector machines, 2001, Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm
[22] 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, June 2000.
[23] A. Scholögl, C. Keinrath, D. Zimmermann, R. Scherer, R. Leeb, and G. Pfurtscheller, "A fully automated correction method of EOG artifacts in EEG recordings," Clin. Neurophysiol., vol.118, no.1, pp.98–104, Jan. 2007.

Appendix: The derivative of f_{X-X}

$f_{X-X}(x)$ is an even function. Hence, in what follows, we consider only $x \geq 0$.

$$\begin{aligned} f_{X-X}(x) &= \int_0^\infty (\tau + x) \exp\left(-\frac{(\tau+x)^2}{2}\right) \\ &\quad \cdot \tau \exp\left(-\frac{\tau^2}{2}\right) \mathrm{d}\tau \\ &= \int_{\frac{x}{2}}^\infty \left\{u^2 - \left(\frac{x}{2}\right)^2\right\} \exp\left(-u^2 - \frac{x^2}{4}\right) \mathrm{d}u \\ &= \exp\left(-\frac{x^2}{4}\right) \\ &\quad \cdot \left[\int_{\frac{x}{2}}^\infty u^2 e^{-u^2}\,\mathrm{d}u - \frac{x^2}{4}\int_{\frac{x}{2}}^\infty e^{-u^2}\,\mathrm{d}u\right], \end{aligned}$$

where $\tau + \frac{x}{2}$ is replaced by u. The first term in [] is

$$\begin{aligned} \int_{\frac{x}{2}}^\infty u^2 e^{-u^2}\,\mathrm{d}u &= \int_{\frac{x}{2}}^\infty -\frac{1}{2} u \left(e^{-u^2}\right)'\,\mathrm{d}u \\ &= \frac{x}{4}\exp\left(-\frac{x^2}{4}\right) + \frac{1}{2}\int_{\frac{x}{2}}^\infty e^{-u^2}\,\mathrm{d}u. \end{aligned}$$

Replacing $\int_{\frac{x}{2}}^\infty e^{-u^2}\,\mathrm{d}u$ by $\frac{\sqrt{\pi}}{2}\mathrm{erfc}(x)$, we finally obtain (10).

Teruyoshi Sasayama was born in 1983. He received the M.S. degree in electrical engineering in 2009 from Kyoto University, Japan. Since 2009, he is currently a Ph.D. student in the Department of Electrical Engineering, Kyoto University, and working toward the Ph.D. degree. He is a research fellow of the Japan Society for the Promotion of Science (JSPS) since April, 2009. His research interests include brain-computer interface (BCI).

Tetsuo Kobayashi was born in 1956. He received the Ph.D. degree in electronic engineering in 1984, from Hokkaido University, Sapporo, Japan. He is currently a professor at Department of Electrical Engineering, Kyoto University. He was a visiting faculty member at Department of Electrical Engineering, University of Rochester, Rochester, NY, from 1987 to 1988 and at Brain Behavior Laboratory, Simon Fraser University, BC, Canada, from 1996 to 1997. His research interests include brain mechanisms of binocular rivalry and human neurocortical dynamics.