

## Research Article

# Sequential Probability Ratio Testing with Power Projective Base Method Improves Decision-Making for BCI

Rong Liu,<sup>1</sup> Yongxuan Wang,<sup>2</sup> Xinyu Wu,<sup>3</sup> and Jun Cheng<sup>3</sup>

<sup>1</sup>Biomedical Engineering Department, Dalian University of Technology, Dalian, Liaoning 116024, China

<sup>2</sup>Affiliated Zhongshan Hospital of Dalian University, Dalian, Liaoning 116001, China

<sup>3</sup>Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

Correspondence should be addressed to Yongxuan Wang; [wyx8904@mail.dlut.edu.cn](mailto:wyx8904@mail.dlut.edu.cn)

Received 30 June 2017; Accepted 11 September 2017; Published 14 November 2017

Academic Editor: Fei Chen

Copyright © 2017 Rong Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Obtaining a fast and reliable decision is an important issue in brain-computer interfaces (BCI), particularly in practical real-time applications such as wheelchair or neuroprosthetic control. In this study, the EEG signals were firstly analyzed with a power projective base method. Then we were applied a decision-making model, the sequential probability ratio testing (SPRT), for single-trial classification of motor imagery movement events. The unique strength of this proposed classification method lies in its accumulative process, which increases the discriminative power as more and more evidence is observed over time. The properties of the method were illustrated on thirteen subjects' recordings from three datasets. Results showed that our proposed power projective method outperformed two benchmark methods for every subject. Moreover, with sequential classifier, the accuracies across subjects were significantly higher than that with nonsequential ones. The average maximum accuracy of the SPRT method was 84.1%, as compared with 82.3% accuracy for the sequential Bayesian (SB) method. The proposed SPRT method provides an explicit relationship between stopping time, thresholds, and error, which is important for balancing the time-accuracy trade-off. These results suggest SPRT would be useful in speeding up decision-making while trading off errors in BCI.

## 1. Introduction

Noninvasive brain-computer interface (BCI) based on the electroencephalogram (EEG) offers a new means of communication to locked-in or paralyzed patients [1, 2] and controlling a prosthesis [3, 4] without reliance on the usual neuromuscular pathways. The critical challenge of BCI technology is to classify the brain signals and mental tasks accurately and fast. However, the EEG recorded from the scalp has the characteristics of low strength, low SNR (signal noise ratio), and the EEG difference under different mental tasks is not significant. Therefore, various pattern recognition algorithms were used in BCI system to extract and classify EEG features.

Event-related desynchronization/synchronization (ERD/ERS) patterns of motor imagery are effective features for EEG-based BCI systems. The experiments show that the phenomenon of ERD/ERS varies among individuals. Therefore, a pattern recognition algorithm should be used to

facilitate decoding “motor intent,” both to find subject-specific EEG features that maximize the separation between the patterns generated by executing the mental tasks and to train classifiers that minimize the classification error rates of these specific patterns. Currently, feature extraction for discrimination of left- and right-hand motor imagery EEG is usually based on EEG band power (BP). For example, autoregression (AR) model [5], discrete Fourier transformation (DFT) [6], and wavelet transforms (WT) [7] have been used to extract EEG features for classification. The wavelet method is one of the most effective algorithms. However, the success of wavelet application greatly depends on the proper selection of subject-specific parameters. Actually, the wavelet transform can be considered as projecting the EEG onto a wavelet basis and the band power as the modulus values of projective coefficients. Inspired by the wavelet method, we introduce a new feature extraction method based on power projective bases to classify EEGs without constrain of wavelet forms.

Moreover, the ability to make rapid decisions based on transient stimuli is a unique aspect of our brains' capacity to process information. Broadly speaking, signal detection theory (SDT) and sequential analysis (SA) are two branches of mathematical models that provide a theoretical framework for understanding how decisions are made [8]. SDT converts a single observation into a categorical choice. According to different decision rules, there are different testing approaches to this problem [9]. For example, Bayesian decision theory is derived by minimizing the posterior expected loss, while Neyman-Pearson (NP) criterion seeks to find the best error probability ( $\alpha$ ) level test. Like most statistical classification methods, for example, linear discriminant analysis (LDA) and support vector machines (SVM), the classification error is the only characteristic of the SDT decision strategies. The necessary number of observation samples determined by the criteria could be very large, which is especially impractical for BCI applications. To control brain-actuated devices, such as robotics and neuroprostheses, both fast decision-making and a stable control signal with a minimal error rate are important [10, 11]. Therefore, recent attentions have been paid to the variable-length sequential sampling model.

A systematic theory of optimal stopping emerged with the work by Wald on the optimality of the sequential probability ratio test (SPRT) [12]. The SPRT achieves a desired error rate with the smallest number of samples, on average. Therefore, in this paper, we introduce a new feature extraction method based on power projective base to classify the EEGs by combining the sequential probability ratio test (SPRT) approach to obtain a continuous dynamic estimate of brain state with accuracy and decision speed balance.

## 2. Methodology

**2.1. Data Description.** The EEG data used in this work were obtained from thirteen subjects from BCI Competitions II, III, and IV. The task was performed based on left- and right-hand motor imagination.

**2.1.1. Dataset III from BCI Competition II.** This dataset contains EEG data from one subject (S1) [13]. The data were recorded from three channels (C3, Cz, and C4) and sampled at 128 Hz. The data consist of 140 labelled and 140 unlabelled trials with an equal number of left- and right-hand trials. Each trial has a duration of 9 s, where a visual cue (arrow) is presented pointing to the left or the right after 3 s preparation period followed by a 6 s motor imagery (MI) task.

**2.1.2. Dataset IIIb from BCI Competition III.** The second dataset contains EEG data recorded over the channels C3 and C4 from three subjects (S2, S3, S4) with some corrections [14]. The data were sampled at 125 Hz. Training and testing sets were available for each subject. Except for the subject O3 that has only just 320 trials for each set, the subjects S4 and X11 contain 540 labelled and 540 unlabelled trials. Each trial has duration of 7 s which consists of 3 s for preparation period and 1 s for a visual cue presentation, followed by another 3 s for the imagination task.

**2.1.3. Dataset IIb from BCI Competition IV.** This dataset contains EEG data from nine subjects (S5–S13) [15]. The data were recorded from three bipolar channels (C3, Cz, and C4) and 3 EOG channels. The sample frequency was 250 Hz. Training and testing set was available for each subject. Each subject participated in two screening sessions without feedback and three online feedback sessions with smiley feedback. The trials without feedback had duration of 7 s, and a visual cue was presented for 1.25 s followed by another 4 s for the imagination task. The trials with feedback had duration of 7.5 s, and a visual cue was presented for 4.5 s until the end of motor imagination.

**2.2. Feature Extraction Method Based on Power Projective Base.** Motor imagery can be regarded as mental rehearsal of a motor act without any obvious motor output. Recent studies show that when performing motor imagination,  $\mu$  (8–13 Hz) and  $\beta$  (18–30 Hz) rhythms are found to reveal event-related desynchronization and synchronization (ERD/ERS) over sensorimotor cortex just as when one performs motor tasks. Due to nonstationary effects having often been observed in brain signals, we proposed a power projective base method to extract classification features from C3 and C4 channels. This method improves the classification accuracy by maximizing the difference of the average projective power between two-class signals. Specifically, the solution of the projective bases can be achieved by generalized eigenvalue decomposition for each subject.

Let  $\mathbf{X}_k \in R^{M \times N_k}$  be the training dataset from one channel, where  $k \in \{L, R\}$  denotes the left- or right-hand motor imagery tasks,  $M$  denotes the sampling points, and  $N_k$  is the number of trials. Moreover, let  $\mathbf{u} \in R^M$  be the projective basis and  $\|\mathbf{u}\| = 1$ . The projective power of signal  $\mathbf{x}_{kj}$ ,  $j = 1, 2, \dots, N_k$  on the projective basis  $\mathbf{u}$  is

$$z_{kj}(\mathbf{u}) = (\mathbf{u}^T \mathbf{x}_{kj}) \cdot (\mathbf{x}_{kj}^T \mathbf{u}). \quad (1)$$

So the mean projective power  $\bar{z}_k(\mathbf{u})$  can be calculated as

$$\begin{aligned} \bar{z}_k(\mathbf{u}) &= \frac{1}{N_k} \sum_{j=1}^{N_k} (\mathbf{u}^T \mathbf{x}_{kj}) \cdot (\mathbf{x}_{kj}^T \mathbf{u}) = \mathbf{u}^T \cdot \left( \frac{\mathbf{X}_k \mathbf{X}_k^T}{N_k} \right) \cdot \mathbf{u} \\ &= \mathbf{u}^T \mathbf{R}_k \mathbf{u}, \end{aligned} \quad (2)$$

where  $\mathbf{R}_k = \mathbf{X}_k \mathbf{X}_k^T / N_k$  is the autocorrelation matrix and it is usually positively definite.

To formulate the objective function to be the ratio of the two-class average projective powers,

$$F(\mathbf{u}) = \frac{\bar{z}_L}{\bar{z}_R} = \frac{\mathbf{u}^T \mathbf{R}_L \mathbf{u}}{\mathbf{u}^T \mathbf{R}_R \mathbf{u}}. \quad (3)$$

By maximizing or minimizing  $F(\mathbf{u})$  to be  $F_{\max}$  or  $F_{\min}$ , the corresponding eigenvector  $\mathbf{u}_{\max}$  or  $\mathbf{u}_{\min}$  is the optimal projective base to be solved. The optimization of (3) could be solved by taking a generalized eigenvalue decomposition

method. First of all, we can get the following decomposition as

$$\mathbf{T}^T \mathbf{R}_L \mathbf{T} = \mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_M),$$

$$\lambda_i \geq \lambda_j > 0 \quad (\forall i < j), \quad (4)$$

$$\mathbf{T}^T \mathbf{R}_R \mathbf{T} = \mathbf{I},$$

where  $\mathbf{T} = [\mathbf{t}_1, \dots, \mathbf{t}_M]$  is the generalized eigenvector matrix and  $\lambda_i$  is the generalized eigenvalue. Therefore, the ratio of mean projection power  $F(\mathbf{u})$  turns to

$$F(\mathbf{u}) = \frac{\mathbf{u}^T \mathbf{T}^{-T} \mathbf{\Lambda} \mathbf{T}^{-1} \mathbf{u}}{\mathbf{u}^T \mathbf{T}^{-T} \mathbf{T}^{-1} \mathbf{u}} = \frac{(\mathbf{u}^T \mathbf{T}^{-T}) \cdot \mathbf{\Lambda} \cdot (\mathbf{T}^{-1} \mathbf{u})}{(\mathbf{u}^T \mathbf{T}^{-T}) \cdot (\mathbf{T}^{-1} \mathbf{u})} \quad (5)$$

$$\triangleq \frac{\mathbf{v}^T \mathbf{\Lambda} \mathbf{v}}{\mathbf{v}^T \mathbf{v}} = \frac{\lambda_1 v_1^2 + \dots + \lambda_M v_M^2}{v_1^2 + \dots + v_M^2},$$

where  $\mathbf{v} = \mathbf{T}^{-1} \mathbf{u}$ . Since

$$\frac{\lambda_1 v_1^2 + \dots + \lambda_M v_M^2}{v_1^2 + \dots + v_M^2} \leq \frac{\lambda_1 v_1^2 + \dots + \lambda_1 v_M^2}{v_1^2 + \dots + v_M^2} = \lambda_1, \quad (6)$$

$$\frac{\lambda_1 v_1^2 + \dots + \lambda_M v_M^2}{v_1^2 + \dots + v_M^2} \geq \frac{\lambda_M v_1^2 + \dots + \lambda_M v_M^2}{v_1^2 + \dots + v_M^2} = \lambda_M$$

$F(\mathbf{u})$  has the maximum value  $F_{\max} = \lambda_1$  and minimum value  $F_{\min} = \lambda_M$ . Obviously, the corresponding vectors  $\mathbf{v}$  can be obtained by

$$\mathbf{v}_{\max} = [1, 0, \dots, 0]^T, \quad (7)$$

$$\mathbf{v}_{\min} = [0, \dots, 0, 1]^T.$$

Then, we have

$$\mathbf{u}_{\max} = \mathbf{T} \cdot [1, 0, \dots, 0]^T, \quad (8)$$

$$\mathbf{u}_{\min} = \mathbf{T} \cdot [0, \dots, 0, 1]^T$$

which means that  $\mathbf{u}_{\max}$  and  $\mathbf{u}_{\min}$  are the first column  $\mathbf{u}_1$  and last column  $\mathbf{u}_M$  of  $\mathbf{T}$ , respectively. Choosing  $\mathbf{u}_{\max}$  or  $\mathbf{u}_{\min}$  as the projective base depends on which is larger between  $\lambda_1$  and  $1/\lambda_M$ . The projective power for the signals of the channels C3 and C4 onto their own projective bases is then stacked together into a 2-dimensional feature vector  $\mathbf{z} = [e_{C3}, e_{C4}]^T$ .

**2.3. SPRT Classification Method.** Sequential analysis is a statistical decision model that assumes decisions are formed by continuously sampling information until the response criterion is satisfied. Once a boundary has been reached, the decision process is concluded and a response is elicited. The number of observations needed for a decision is not determined in advance of the experiment, but by the observations obtained during the test. The data should be fed to the SPRT algorithm sequentially, so we divide each trial into segments with overlap and each one has the same length as that of the projective base used in feature extraction.

Taking into account the nonstationarity of the EEG sampling information, we assume that the probability distribution of the  $i$ th segment feature  $\mathbf{z}_i$  for class  $k$  is  $g_k(\mathbf{z}_i)$ ,  $k \in \{L, R\}$ , and then the probability ratio for the  $i$ th segment is

$$l_i(\mathbf{z}_i) = \frac{g_{Li}(\mathbf{z}_i)}{g_{Ri}(\mathbf{z}_i)} \quad (9)$$

and the evidence accumulation turns to

$$l_q(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_q) = \frac{\prod_{i=1}^q g_{Li}(\mathbf{z}_i)}{\prod_{i=1}^q g_{Ri}(\mathbf{z}_i)}, \quad (10)$$

where  $q$  is the number of accumulated segments. Assuming the segments are independent for computation convenience [12], we have

$$l_q(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_q) = \frac{g_L(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_q)}{g_R(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_q)}, \quad (11)$$

where  $g_k(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_q)$ ,  $k \in \{L, R\}$  is the joint probability distribution of  $q$  dimensional vector  $(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_q)$ . This assumption is violated by our data in practice.

The decision rule with two thresholds  $\eta_L$  and  $\eta_R$  is

$$\begin{aligned} & \text{if } l_q(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_q) \geq \eta_L, \quad \hat{k} = L \\ & \text{if } l_q(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_q) \leq \eta_R, \quad \hat{k} = R \\ & \text{if } \eta_R < l_q(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_q) < \eta_L, \quad \text{compute } l_{q+1}. \end{aligned} \quad (12)$$

With two thresholds, we have the option to increase  $\eta_L$  or decrease  $\eta_R$  which will increase the probability to make a correct decision (by waiting to accumulate more data or evidence) but decrease the probability of making a wrong decision (by delaying the decision). The error probabilities are defined as

$$p_{L|R} = P(\hat{k} = L | k = R), \quad (13)$$

$$p_{R|L} = P(\hat{k} = R | k = L).$$

If  $l_q(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_q) \geq \eta_L$  is satisfied, we define the corresponding space of vector  $(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_q)$  to be  $Z_L$ . With (11), we have

$$g_L(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_q) \geq \eta_L \cdot g_R(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_q). \quad (14)$$

Equation (14) is then integrated in  $Z_L$  yielding

$$\int_{Z_L} g_L(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_q) \geq \eta_L \cdot \int_{Z_L} g_R(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_q). \quad (15)$$

That is,

$$1 - p_{R|L} \geq \eta_L \cdot p_{L|R}. \quad (16)$$

Analogous reasoning for  $l_q(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_q) \leq \eta_R$  yields

$$p_{R|L} \leq \eta_R \cdot (1 - p_{L|R}). \quad (17)$$

Thus, the two detection thresholds  $\eta_L$  and  $\eta_R$  are related to the error probabilities by

$$p_{L|R} \leq \frac{1 - p_{R|L}}{\eta_L} \leq \frac{1}{\eta_L}, \quad (18)$$

$$p_{R|L} \leq (1 - p_{L|R}) \cdot \eta_R \leq \eta_R.$$

The two kinds of error probabilities can be lowered by either increasing  $\eta_L$  or decreasing  $\eta_R$ . However, due to the limited number of segments, the indecision ratio will increase as the error probability  $p_{L|R}$  or  $p_{R|L}$  is decreased. Therefore, we may not obtain the optimal result by simply increasing  $\eta_L$  or decreasing  $\eta_R$ . The suitable  $\eta_L$  and  $\eta_R$  could be achieved by the following optimization criteria.

Under the assumption that the features follow a Gaussian distribution, we can take logarithm on both sides of (10) to obtain the log probability ratio (log PR), which leads a sequential probability ratio test (SPRT) as

$$L_q = \sum_{i=1}^q \frac{1}{2} \left( \log \frac{|\mathbf{S}_{Ri}|}{|\mathbf{S}_{Li}|} + D_M^2(\mathbf{z}_i, \mathbf{m}_{Ri}) - D_M^2(\mathbf{z}_i, \mathbf{m}_{Li}) \right) \triangleq \sum_{i=1}^q J_i, \quad (19)$$

where  $D_M(\mathbf{z}_i, \mathbf{m}_{ki}) = \sqrt{(\mathbf{z}_i - \mathbf{m}_{ki})^T \mathbf{S}_{ki}^{-1} (\mathbf{z}_i - \mathbf{m}_{ki})}$  is the Mahalanobis distance and  $J_i$  is the log PR of the  $i$ th segment.

We can derive the average  $J_i$  for each class:

$$E(J_i | k = L) = \frac{1}{2} \left( \log \frac{|\mathbf{S}_{Ri}|}{|\mathbf{S}_{Li}|} + \text{tr} \left( (\mathbf{S}_{Li} + (\mathbf{m}_{Li} - \mathbf{m}_{Ri}) \cdot (\mathbf{m}_{Li} - \mathbf{m}_{Ri})^T) \cdot \mathbf{S}_{Ri}^{-1} - \mathbf{I} \right) \right), \quad (20)$$

$$E(J_i | k = R) = \frac{1}{2} \left( \log \frac{|\mathbf{S}_{Ri}|}{|\mathbf{S}_{Li}|} - \text{tr} \left( (\mathbf{S}_{Ri} + (\mathbf{m}_{Ri} - \mathbf{m}_{Li}) \cdot (\mathbf{m}_{Ri} - \mathbf{m}_{Li})^T) \cdot \mathbf{S}_{Li}^{-1} - \mathbf{I} \right) \right).$$

For any given threshold pair  $\eta_L$  and  $\eta_R$ , the number of accumulated segments to make a correct decision, that is, stopping time, for class  $k$ , is  $E(q | k)$  which satisfies

$$E(q | k = L) = \inf_n \left( \sum_{i=1}^n E(J_i | k = L) \geq (1 - \eta_R) \cdot \log \eta_L + \eta_R \cdot \log \eta_R \right)$$

$$E(q | k = R) = \inf_n \left( \sum_{i=1}^n E(J_i | k = R) \geq \left( 1 - \frac{1}{\eta_L} \right) \cdot \log \eta_R + \frac{1}{\eta_L} \cdot \log \eta_L \right), \quad (21)$$

where  $\inf_n(\cdot)$  is the minimum element of set of  $n$ . Generally,  $E(q | k = L)$  and  $E(q | k = R)$  may be different. Since the stopping time is a key point in the sequential analysis, we constrain the two thresholds by unifying  $E(q | k)$  of two classes to be equal, that is,  $E(q | k = L) = E(q | k = R) \triangleq q_E$ . Then the thresholds are given by

$$\xi_L = \sum_{i=1}^{q_E} E(J_i | k = L), \quad (22)$$

$$\xi_R = \sum_{i=1}^{q_E} E(J_i | k = R).$$

For any given stopping time  $q_E$ , there is a corresponding threshold pair  $\xi_L$  and  $\xi_R$ . The decision rule with two thresholds  $\xi_L$  and  $\eta_R$  is

$$\begin{aligned} & \text{if } L_q \geq \xi_L, \quad \hat{k} = L, \\ & \text{if } L_q \leq \xi_R, \quad \hat{k} = R, \\ & \text{if } \xi_R < L_q < \xi_L, \quad \text{compute } L_{q+1}. \end{aligned} \quad (23)$$

From this decision policy, we can see that other than assigning one of the two classes  $L$  and  $R$ , the decision functions may still be undecided and continue testing to the next observation. The ‘‘undecided’’ response keeps the number of errors (false positives or false negatives) low, which is useful for avoiding making excessive mistakes to speed up decisions, for example, a BCI control wheelchair running into an obstacle [16]. In addition, when it is still undecided when reaching to the stopping time  $q_E$ , we specify that when  $q = q_E$ , the decision rule is

$$\begin{aligned} & \text{if } L_q > 0, \quad \hat{k} = L, \\ & \text{if } L_q < 0, \quad \hat{k} = R. \end{aligned} \quad (24)$$

Till now, with the above decision rules, the consequent results, such as accuracy, mutual information (MI) [17], the steepness of MI [18], and average decision time, will only rely on the stopping time  $q_E$  and the data to be analyzed. Depending on the actual specific needs, we can set the accuracy, MI, the steepness of MI, and average decision time as the optimization target, respectively, to determine the optimal stopping time  $q_{\text{opt}}$ . At the same time, the two thresholds are determined.

### 3. Results

*3.1. Feature Extraction.* To evaluate the performance of our method, we tested it using BCI Competition Datasets II and

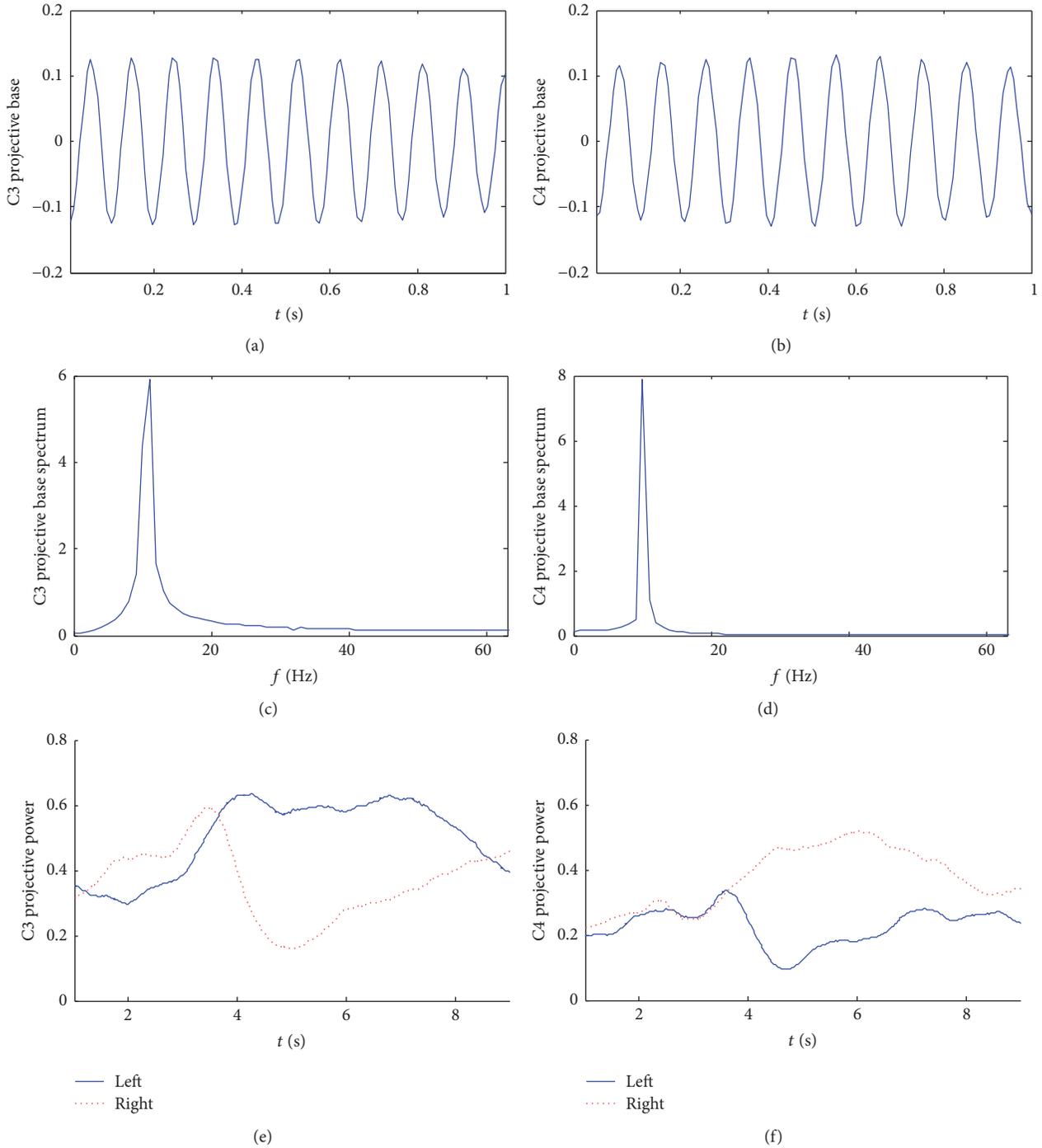


FIGURE 1: The power projective base of the subject S1. (a) C3 projective base; (b) C4 projective base; (c) C3 projective base spectrum; (d) C4 projective base spectrum; (e) the time-varied projective power of C3; (f) the time-varied projective power of C4.

Dataset IIIb from BCI Competition III. They were obtained from four subjects, denoted as S1–S4. The task performed was based on left and right-hand motor imagination.

The dimension of the projective base, that is, the length of the sliding window, is set to be 1 s. The time-domain waveforms of the optimal projective base of the two channels for subject S1 are shown in Figures 1(a) and 1(b). The corresponding frequency spectra are shown in Figures 1(c)

and 1(d). The average projective power time courses during the right-hand (dash line) and left-hand (solid line) imagined movement for the C3 and C4 are displayed in Figures 1(e) and 1(f). From this figure, we can see that the projective bases are similar to modulated sine signals and the spectra have band-pass characteristics which are similar to that of wavelet base. For this subject, the projective power dominates in the  $\mu$  rhythm. During the first 3.5 s (0.5 s after cue presentation),

TABLE 1: A comparison of ACC and MI for three methods.

Method	Criterion	S1	S2	S3	S4	Avg.
DFT	Acc (%)	88.6	86.3	76.0	77.9	82.2
	MI (B)	0.488	0.424	0.205	0.238	0.339
WT	Acc (%)	89.3	88.4	79.2	81.1	84.5
	MI (B)	0.509	0.482	0.262	0.301	0.389
PP	Acc (%)	92.1	90.6	83.1	84.4	87.6
	MI (B)	0.601	0.550	0.345	0.375	0.468

the projective power curves under two conditions are close; after 3.5 s, distinct difference in the projective power can be observed which provides a good classification feature. The projective bases for subjects S2 and S3 are similar to that of subject S1.

The power projective base of subject S4 is shown in Figure 2. In contrast with subject S1, the reactivity patterns of the projective bases of this subject are quite different. The waveforms of the projective bases are oscillating faster. Obviously, the frequency of projective power is higher than that of subject S1. As seen in Figures 2(c) and 2(d), the projective bases display a peak in  $\beta$  rhythms. Moreover, the results of the average projective power time courses demonstrate that the patterns of ERD/ERS subject S4 are quite different. That means this projective power method is subject-adaptive as well as avoiding the parameter setup in advance.

**3.2. Classification Results.** Two kinds of experiments were performed to evaluate the performance of the proposed machine learning method. One is to evaluate the projective power feature extraction method and the other is to evaluate the SPRT classification performance. In the first one, for the purpose of benchmarking, we compared the classification accuracy (ACC), the mutual information (MI) with two benchmark feature extraction algorithms, and DFT and WT based on the sequential Bayesian classifier [17, 18]. These methods were also applied to the data of subjects S1–S4. The classification accuracy (ACC) and mutual information (MI) of the three methods in consideration are listed in Table 1, where Avg. denotes the averaged indexes over all four subjects. The results of WT are derived from Lemm’s method which won the BCI competitions in 2003 and 2005 for motor imagery datasets. From Table 1, we can see that the proposed power projective method outperforms two benchmark methods for every subject. Compared with wavelet method, the average ACC of our method increased from 84.5% to 87.6% and MI increased to 0.468 bits. Furthermore, a paired  $t$ -test analysis was used to compare the classification accuracies of the three methods. The paired  $t$ -test result confirms that with projective base method the ACC and MI across subjects are significantly higher than that with WT ( $t = 15.1$ ,  $p < 0.01$ ) and with DFT ( $t = 8.432$ ,  $p < 0.01$ ).

Then we compared the ACC and MI between several state-of-the-art nonsequential classifiers used in the BCI community and their sequential ones with the projective base feature. Those methods were LDA, SVM, and Bayesian,

and the corresponding sequential methods were denoted as SLDA, SSVM, and SB. The tenfold cross-validation was carried out for classification tests of EEG data in this study; that is, the datasets for each subject were divided into ten subsets, and the following procedure was repeated ten times. Each time, one of the ten subsets was used as the test set and the other nine were used as the training set. The average recognition rate was evaluated across all ten folds. These methods were applied to the three BCI competition datasets described in Section 2.1 from thirteen subjects S1–S13.

Table 2 shows the classification accuracy of all the methods on the competition test data, where “Avg.” denotes the averaged results over all the subjects. By applying the obtained effective features with the power projective method on the nonsequential methods, a classification accuracy of 76.0%, 71.6%, and 75% was achieved by LDA, SVM, and Bayesian, respectively. The accuracies achieved by all of the corresponding sequential classifiers were greater than the nonsequential one. The paired  $t$ -test result confirms that with sequential classifier the accuracies across subjects are significantly higher than that with nonsequential ones ( $t = 6.15$ ,  $p < 0.01$ ). Overall, the classification accuracy is higher when the sequential scheme is adopted.

Additionally, it is also important to consider the stopping time for speedy BCI applications without sacrificing accuracy. Therefore, we further analyze the classification accuracy, the MI, and stopping time of the SPRT and compared the time-accuracy trade-off between SPRT and SB. The classification results of the SPRT are provided in Table 3. By applying the obtained effective features with the power projective method on the SPRT, an average classification accuracy of 84.1% was achieved. The proposed SPRT classifier with the projective base features outperforms the SB classifier with the same feature extraction method in terms of classification accuracy across all the subjects except for subject S2. In terms of average accuracy, SPRT outperforms SB classifier. Analyzing the overall results, it can be concluded that SPRT with the projective base feature extraction method outperforms the state-of-the-art methods on the standard datasets.

The resulting time courses of the accumulative classification information, the so-called accumulative evidence, of our SPRT method and SB method are shown in Figure 3. Figure 3(a) shows that the SB classification method gains information from around 4 s (3 s preparation period and 1 s window length). The cumulative Bayesian posterior probabilities reach the extremum at around 5.5 s, indicating a peak decision confidence at this time. However, the accumulative information falls down at the end of trial. The result shows that the effective control takes place during the middle of a trial. More evidence could not increase the classification performance any further.

Compared with the SB method, the cumulative process of SPRT has monotonicity (shown in Figure 3) which makes it possible to improve the accuracy with more evidence available. During the SPRT classification process, once the accumulative evidence exceeds one of the thresholds, an immediate decision will be given. Moreover, the thresholds can be adjusted to change the decision time. That is to say, with two broader threshold choices, a larger number

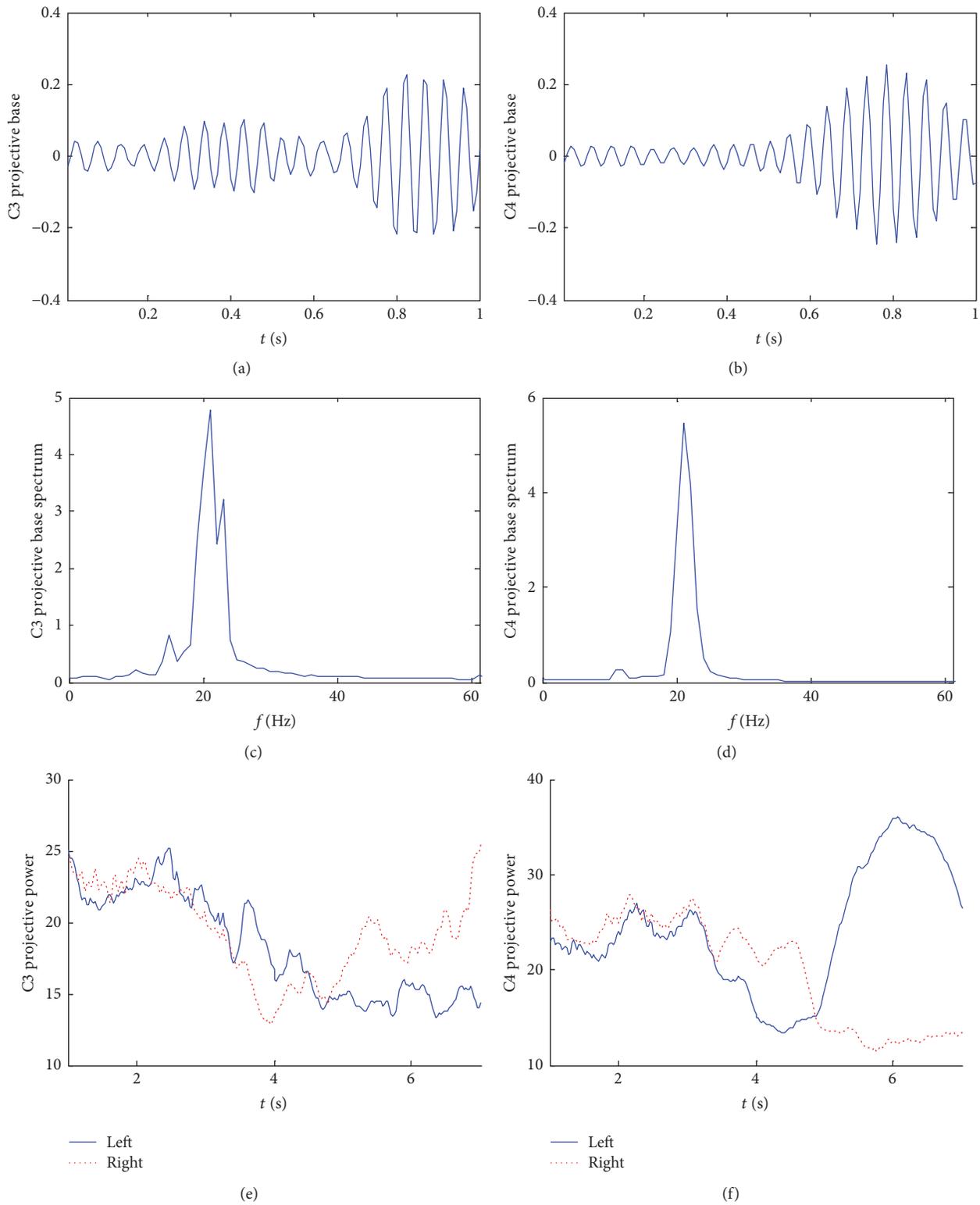


FIGURE 2: The power projective base of the subject S4. (a) C3 projective base; (b) C4 projective base; (c) C3 projective base spectrum; (d) C4 projective base spectrum; (e) the time-varied projective power of C3; (f) the time-varied projective power of C4.

TABLE 2: Accuracy (%) of different classifiers on the BCI competition datasets.

Subject	LDA	SLDA	SVM	SSVM	Bayesian	SB
S1	87.8	90.0	87.1	92.1	86.4	92.1
S2	74.6	88.3	74.6	85.7	76.1	90.6
S3	64.4	68.2	64.4	65.4	62.2	83.1
S4	71.8	77.1	66.4	66.4	71.2	84.4
S5	64.0	64.0	60.9	63.1	61.4	66.6
S6	61.2	61.2	59.1	57.1	60.4	64.1
S7	61.1	55.0	61.1	57.6	59.8	65.0
S8	95.7	98.0	87.3	96.0	97.1	98.1
S9	78.2	83.6	73.7	84.2	77.9	85.9
S10	77.5	86.4	66.1	73.5	72.1	80.5
S11	78.8	80.1	76.7	83.1	78.2	79.7
S12	90.4	92.4	80.5	88.4	90.8	92.8
S13	82.8	89.8	72.6	87.7	80.8	86.5
Avg.	76.0	79.5	71.6	76.9	75.0	82.3

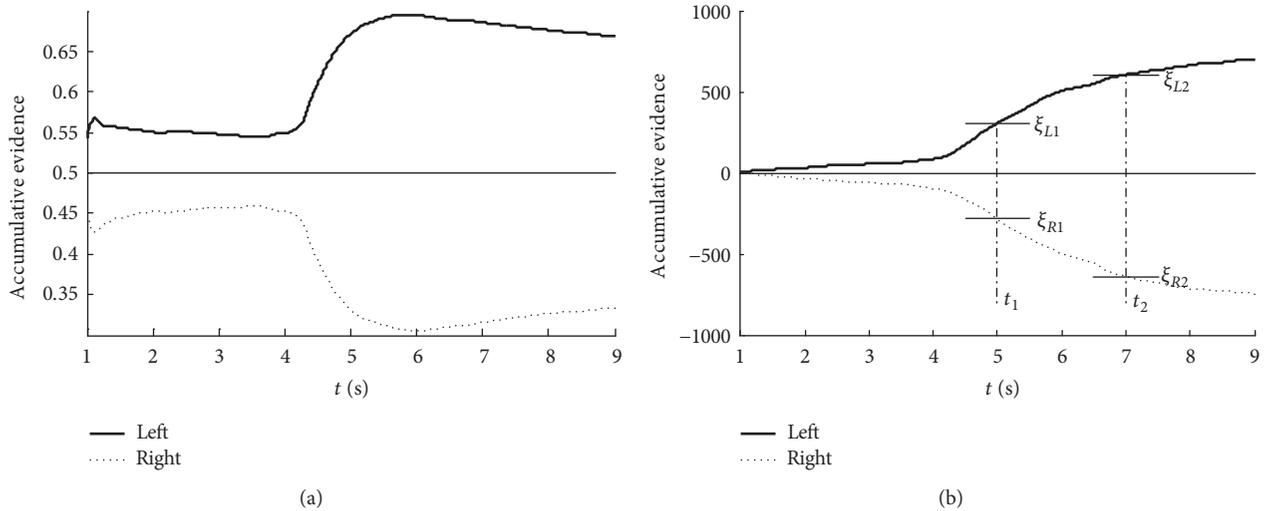


FIGURE 3: The average accumulative process of classification information for subject S1: (a) SB method; (b) SPRT method.

TABLE 3: The classification results of the SPRT method.

Subject	ACC (%)	MI (bit)	$T$ (s)
S1	92.8	0.627	1.23
S2	88.7	0.491	1.95
S3	84.6	0.380	2.80
S4	85.0	0.390	2.35
S5	73.6	0.167	2.33
S6	68.2	0.098	1.23
S7	70.9	0.130	0.87
S8	97.9	0.853	1.33
S9	86.2	0.421	2.48
S10	83.8	0.361	2.54
S11	84.9	0.388	1.76
S12	93.6	0.657	1.69
S13	83.6	0.356	2.23
Avg.	84.1	0.409	1.91

of observations may be required to improve the accuracy, and vice versa. This can be seen from Figure 3(b), given the expected stopping time  $t_1$ , the thresholds are  $\xi_{L1}, \xi_{R1}$ . When the expected stopping time is set to be  $t_2$ , the thresholds change to be  $\xi_{L2}, \xi_{R2}$ . Obviously, it will achieve a higher accuracy with more decision time. This depicts the inherent trade-off between decision time (costs) and accuracy (benefits) of the SPRT method.

The resulting time courses for the classification accuracy, the MI, and SMI (calculated as  $MI(t)/(t - 3s)$  for  $t > 3.5s$ ) for subject S1 are presented in Figure 4. The SMI quantifies the response time. During the first 4 s, the classification performs at a rate no better than chance. Afterwards a steep ascent in the classification accuracy can be observed, meanwhile reflecting a raising MI. The increase of MI indicates an increase in separation ability between left- and right-hand motor imagery. The maximum classification accuracy and the maximum MI are achieved at about 6.4 s and 7.1 s, respectively. The observation of nonmonotonic relationship

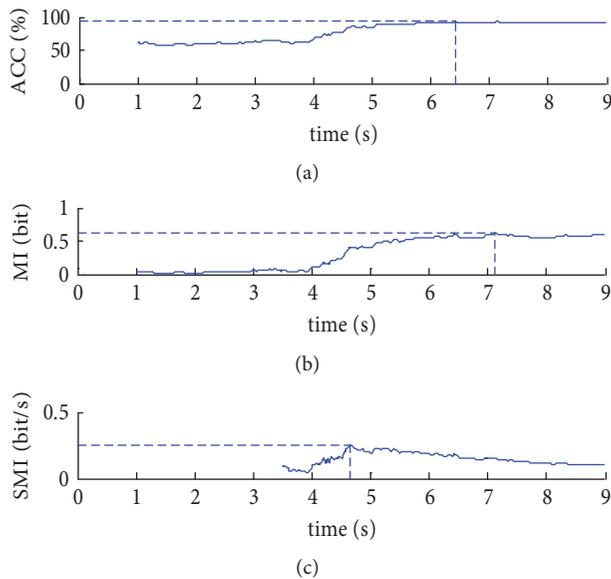


FIGURE 4: The time course of the classification accuracy, the MI, and the steepness of MI for subject S1: (a) ACC; (b) MI; (c) SMI.

between accuracy and the time is due to the limited data available which leads to the final decision. With consideration about time, the maximum steepness of MI is obtained at around 4.6 s.

The general sequential framework of the present approach is customizable to suit different task objectives, such as improving accuracy, MI, or steepness of MI. However, the optimization of one particular objective will come at the expense of the others.

#### 4. Conclusion

In this paper, we present a SPRT method in conjunction with power projective base method to recognize mental states. The power projective method was first developed to determine features by maximizing the average projection energy difference of the two types of signals. With the accumulative evidence curve, the proposed SPRT sets the two-constrained thresholds based on a desired expected stopping time. The SPRT method adds the benefit of a customizable trade-off between accuracy and decision speed. Specifically, the thresholds in this method were determined without predefined error probabilities. Using standardized datasets, improved performances were demonstrated.

Although this study suggests that SPRT would be useful to balance the speed and accuracy for different BCI applications, we realize that further work requires investigation. Future work will attempt to validate the proposed method with a larger dataset and implement it in our BCI-actuated robotic system. Moreover, we will investigate the multiway SPRT theory and its application to the multiclass BCI systems.

#### Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

#### Acknowledgments

This work was supported in part by the grants from the Natural Science Foundation of China (NSFC) (61005088 and 61573079), the Open Research Foundation of State Key Lab. of Digital Manufacturing Equipment & Technology in Huazhong University (DMETKF2015007), the Fundamental Research Funds for the Central Universities to Dr. Liu, and the NSFC-Shenzhen Robotics Research Center Project (U1613219) to Dr. Wu.

#### References

- [1] C. E. Bouton, A. Shaikhouni, N. V. Annetta et al., "Restoring cortical control of functional movement in a human with quadriplegia," *Nature*, vol. 533, pp. 247–250, 2016.
- [2] M. J. Vansteensel, E. G. M. Pels, M. G. Bleichner et al., "Fully implanted brain-computer interface in a locked-in patient with ALS," *The New England Journal of Medicine*, vol. 375, no. 21, pp. 2060–2066, 2016.
- [3] G. R. Muller-Putz and G. Pfurtscheller, "Control of an electrical prosthesis with an SSVEP-based BCI," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 1, pp. 361–364, 2008.
- [4] B. Rebsamen, C. L. Teo, Q. Zeng et al., "Controlling a wheelchair indoors using thought," *IEEE Intelligent Systems*, vol. 22, no. 2, pp. 18–24, 2007.
- [5] V. Lawhern, W. D. Hairston, K. McDowell, M. Westerfield, and K. Robbins, "Detection and classification of subject-generated artifacts in EEG signals using autoregressive models," *Journal of Neuroscience Methods*, vol. 208, no. 2, pp. 181–189, 2012.
- [6] K. Samiee, P. Kovács, and M. Gabbouj, "Epileptic seizure classification of EEG time-series using rational discrete short-time fourier transform," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 2, pp. 541–552, 2015.
- [7] R. Liu, Y.-X. Wang, and L. Zhang, "An FDES-Based shared control method for asynchronous brain-actuated robot," *IEEE Transactions on Cybernetics*, vol. 46, no. 6, pp. 1452–1462, 2016.
- [8] J. I. Gold and M. N. Shadlen, "The neural basis of decision making," *Annual Review of Neuroscience*, vol. 30, pp. 535–574, 2007.
- [9] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Wiley-Interscience, New York, NY, USA, 2nd edition, 2001.
- [10] G. Krausz, R. Scherer, G. Korisek, and G. Pfurtscheller, "Critical decision-speed and information transfer in the "Graz Brain-Computer Interface"," *Applied Psychophysiology and Biofeedback*, vol. 28, no. 3, pp. 233–240, 2003.
- [11] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Müller, and G. Curio, "The non-invasive Berlin brain-computer interface: fast acquisition of effective performance in untrained subjects," *NeuroImage*, vol. 37, no. 2, pp. 539–550, 2007.
- [12] A. Wald, *Sequential Analysis: Design Methods and Applications*, John Wiley & Sons, Inc., New York; Chapman & Hall, Ltd., London, 1947.
- [13] B. Blankertz, K. R. Müller, G. Curio, T. M. Vaughan et al., "The BCI competition 2003 set III: probabilistic modeling of sensorimotor  $\mu$  rhythms for classification of imaginary hand movements," *IEEE Transactions on Biomedical Engineering*, vol. 51, pp. 1044–1051, 2004.
- [14] B. Blankertz, K.-R. Müller, D. J. Krusienski et al., "The BCI competition III: validating alternative approaches to actual BCI

- problems,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 2, pp. 153–159, 2006.
- [15] N. Brodu, F. Lotte, and A. Lécuyer, “Exploring two novel features for EEG-based brain-computer interfaces: Multifractal cumulants and predictive complexity,” *Neurocomputing*, vol. 79, pp. 87–94, 2012.
- [16] J. D. R. Millan, F. Renkens, J. Mouriño, and W. Gerstner, “Non-invasive brain-actuated control of a mobile robot by human EEG,” *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1026–1033, 2004.
- [17] S. Lemm, C. Schäfer, and G. Curio, “BCI Competition 2003—data set III: modeling of sensorimotor  $\mu$  rhythms for classification of imaginary hand movements,” *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1077–1080, 2004.
- [18] S. Lemm, C. Schafer, and G. Curio, “Aggregating classification accuracy across time: application to single trial EEG,” in *Advances in Neural Information Processing Systems*, vol. 19, pp. 825–832, Aggregating classification accuracy across time, application to single trial EEG, 2007.



**Hindawi**  
Submit your manuscripts at  
<https://www.hindawi.com>

