

# *Subject-Specific Channel Selection Using Time Information for Motor Imagery Brain–Computer Interfaces*

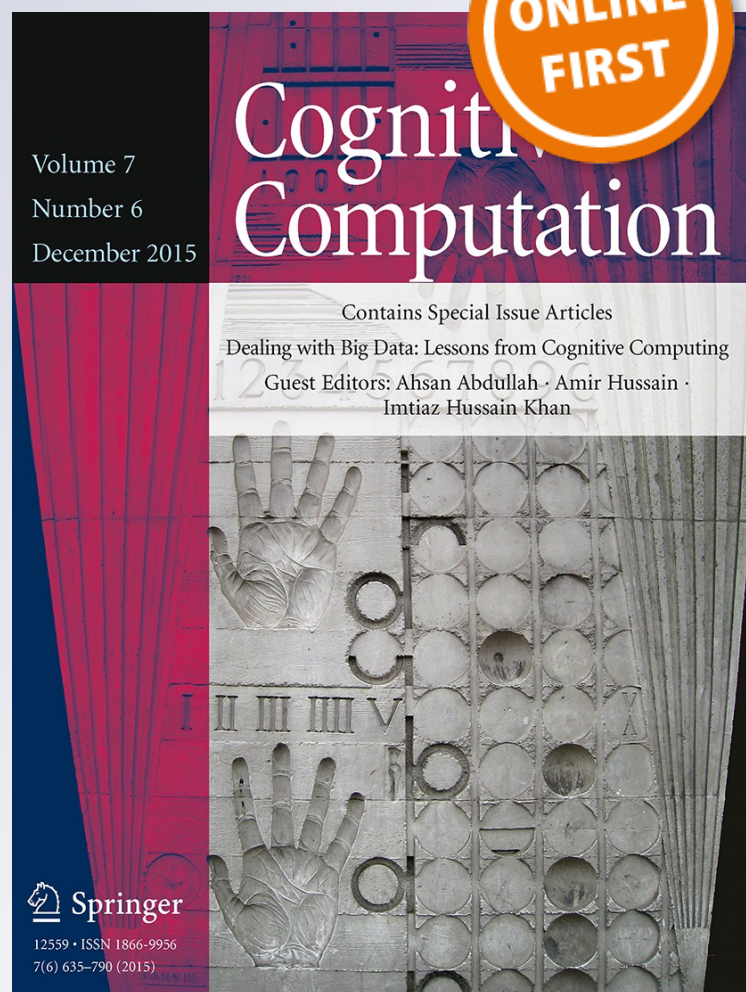
**Yuan Yang, Isabelle Bloch, Sylvain  
Chevallier & Joe Wiart**

**Cognitive Computation**

ISSN 1866-9956

Cogn Comput

DOI 10.1007/s12559-015-9379-z



**Your article is protected by copyright and all rights are held exclusively by Springer Science +Business Media New York. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at [link.springer.com](http://link.springer.com)".**

# Subject-Specific Channel Selection Using Time Information for Motor Imagery Brain–Computer Interfaces

Yuan Yang<sup>1,2,3</sup> · Isabelle Bloch<sup>1,2</sup> · Sylvain Chevallier<sup>4</sup> · Joe Wiart<sup>1,2,5</sup>

Received: 3 July 2015 / Accepted: 28 December 2015  
© Springer Science+Business Media New York 2016

**Abstract** Keeping a minimal number of channels is essential for designing a portable brain–computer interface system for daily usage. Most existing methods choose key channels based on spatial information without optimization of time segment for classification. This paper proposes a novel subject-specific channel selection method based on a criterion called  $F$  score to realize the parameterization of both time segment and channel positions. The  $F$  score is a novel simplified measure derived from Fisher’s discriminant analysis for evaluating the discriminative power of a group of features. The experimental results on a standard dataset (BCI competition III dataset IVa) show that our method can efficiently reduce the number of channels (from 118 channels to 9 in average) without a decrease in mean classification accuracy. Compared to two state-of-the-art methods in channel selection, our method leads to comparable or even better classification results with less selected channels.

**Keywords** Brain–computer interfaces · Channel reduction · Time information · EEG · Fisher’s discriminant analysis

## Introduction

Brain–computer interfaces (BCIs) are systems that support a direct communication between brain and computer without any use of peripheral nerves and muscle movements [1, 31]. The basic structure of a BCI typically includes four essential parts: brain signal acquisition, feature extraction, feature-to-command translation and command output pathway. Some systems may contain a feedback. The brain signal can be recorded by various techniques, either invasive or noninvasive [19].

The BCIs based on electroencephalography (EEG) are noninvasive BCIs, which record EEG signal with electrodes placed on the surface of the scalp [1, 8, 23]. EEG studies show that imaginary movements of different body parts can cause a power decrease in sensorimotor rhythms of EEG, i.e.,  $\mu$  (8–13 Hz) and  $\beta$  rhythms (14–35 Hz), called event-related desynchronization (ERD), at corresponding “active” cortex areas [25]; meanwhile, a power increase in sensorimotor rhythms called event-related synchronization (ERS) might be observed at other “idling” areas during the motor imagery [24]. Thus, motor imagery of different body parts can be identified by classifying ERD/ERS patterns, which gives birth to a type of EEG-based BCI called motor imagery BCI [31].

The advantage of this type of BCI is that it is inexpensive, of low risk and portable. However, due to volume conduction through the scalp, skull and other layers of the brain, the EEG recorded by a scalp sensor is a “blurred” copy of multi-source activities (e.g., visual-related

✉ Yuan Yang  
Y.Yang-2@tudelft.nl

Isabelle Bloch  
isabelle.bloch@telecom-paristech.fr

<sup>1</sup> LTCI, CNRS, Télécom ParisTech, Université Paris-Saclay, Paris, France

<sup>2</sup> Whist Lab, 46 rue Barrault, Paris 75013, France

<sup>3</sup> Present Address: Department of BioMechanical Engineering, Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands

<sup>4</sup> LISV - IUT de Vélizy, Université de Versailles St-Quentin, 10-12 Avenue de l’Europe, Vélizy 78140, France

<sup>5</sup> Orange Labs R&D, 38 40 rue du Général Leclerc, Issy les Moulineaux 92794, France

activities and motor imagery) [11, 22], which reduces the signal-to-noise ratio (SNR) and therefore increases the difficulty of signal decoding. The classical solution is to use a multi-channel recording and spatial filtering algorithms, such as common spatial patterns (CSP), to improve the SNR and extract discriminative features from overlapping signals [7]. However, this setting may reduce the portability and practicability of BCI, because it typically requires a large number of EEG electrodes (e.g., 64 or 128), which represents a main drawback for final users in daily usage (e.g., neuro-games) [36].

To develop a daily use system, several advanced algorithms were proposed to reduce the number of electrodes in BCI by selecting some key EEG channels [4, 12, 15, 16, 27, 30, 37]. A thorough review of channel selection algorithms for EEG signal processing can be found in [2]. Most of existing studies addressed the issue of channel selection using only spatial information, disregarding the potential impact of time and frequency information. In this case, the optimal combination of time, frequency and channel (electrode) position may not be achieved in a BCI design. Although a recent study showed that a broad frequency band (8–30 Hz) that covers both  $\mu$  (8–12 Hz) and  $\beta$  (18–25 Hz) bands can generally be used when employing features, called time domain parameters (TDPs) [28], the existing channel selection methods mainly work with the popular band power (BP) feature, which is sensitive to frequency band and time segment [10, 17, 34].

As motor imagery BCIs typically rely on decoding sensorimotor rhythm, in practice, many researchers simply placed electrodes at three key positions (C3, Cz and C4 of 10–20 recording system [14]) in the sensorimotor areas to reduce the number of electrodes, which we call 3C setup. The advantages of the 3C setup are that it does not need a full EEG cap, training data or machine learning methods to find the optimal positions for recording. It can be used when only a few electrodes are available. However, due to the limited information and low SNR of signal, it may not achieve good classification results in most cases. Our previous studies indicated that some preprocessing steps, such as the time–frequency optimization, were often needed to improve the performance of 3C setup (see [33, 34], for details). Moreover, general users may not be skillful enough to place the electrodes at the precise locations of C3, Cz and C4 each time, if a standard EEG cap is not used.

Here, we present a novel channel selection method using TDP features. As TDP features are less sensitive to frequency band, we used broadband (8–30 Hz) EEG signals in this work. Different from the existing methods [4, 12, 16, 27, 30, 37] and our previous work on channel selection [15], this novel approach considers the effect of time window on channel selection, so as to find the optimal combination of time segment and subset of channels for BCI design. A new

criterion based on Fisher's discriminant analysis, namely  $F$  score, was used in our method to measure the discrimination power of TDP features extracted from different channels and different time segments. The application of this new criterion has first been demonstrated in our previous study by Yang et al. [34] for time–frequency optimization in BCIs, showing better results than the state-of-art methods. Later, this new criterion has also been successfully applied to a motor cognition study by Ansuini et al. [3] for classifying kinematic features.

We evaluated our method in a standard dataset (BCI competition III dataset IVa [5]). We performed the comparisons between the channel selection using time information (CSTI), the channel selection based on the long time segment from the cue on-set to the ending of the cue, the 3C setup, the full-cap-based CSP and two state-of-the-art methods in channel selection (the  $l_1$ -norm-based sparse CSP [37] and the Riemannian distance-based channel selection [4]) to validate the contribution of our method (CSTI). Additionally, the effects of electrode misplacing and data evolution were also examined to study their potential influence on classification.

## Materials and Methods

### Time Domain Parameters

The EEG signals are band-pass-filtered between 8 and 30 Hz using a 5th-order Butterworth filter. For one channel (electrode) and one trial, we denote by  $x(t)$  the filtered EEG signal in a time segment  $[t_0, t_0 + T - 1]$ . Time domain parameters (TDPs) are a set of broadband (i.e., 8–30 Hz) EEG features defined in the time domain [28]:

$$\text{TDP}^{(p)} = \log \left( \text{var}_{t \in [t_0, t_0 + T - 1]} \left( \frac{d^p x(t)}{dt^p} \right) \right), \quad p = 0, 1, 2, \dots \quad (1)$$

The logarithm is applied here to make the distribution of TDPs approximately Gaussian (for details, see [28]), since the linear classifier we use here typically assumes that the input features follow Gaussian distributions [21]. Note that the TDP of order 0,  $A = \text{TDP}^{(0)}$ , is the BP feature. It characterizes the EEG pattern in terms of amplitude. Although TDP features are defined in the time domain, they can as well be interpreted as frequency domain filters. Therefore, the frequency domain information has already been integrated in the TDP features. The TDP of order 1,  $M = \text{TDP}^{(1)}$ , can be considered as a feature that reflects the EEG pattern in terms of high frequency (mainly the beta band), and the TDP of order 2,  $C = \text{TDP}^{(2)}$ , reflects the change in frequency [28]. We use these three TDPs,



[ $A, M, C$ ], in this work, since they carry more information than the only BP feature, and have clearer physical meanings than TDPs of higher orders in BCI research.

### A Criterion Based on Fisher's Discriminant

Fisher's linear discriminant analysis (Fisher's LDA) is a very popular classification algorithm in BCI research [21], because it has a very low computational cost and usually yields good results for motor imagery BCIs [18]. It projects high-dimensional data onto a direction and performs a linear classification in this one-dimensional space. The optimal projection is found by maximizing the separation between two classes. Let us assume that we have two classes of observations,  $h$  and  $f$ . In a one-dimensional feature space, the separation between two classes is defined using the Fisher criterion [21]:

$$FC = \frac{(\mu^h - \mu^f)^2}{(\sigma^h)^2 + (\sigma^f)^2} \quad (2)$$

where  $\mu^h$  and  $\mu^f$  are the mean values of the feature over all trials for classes  $h$  and  $f$ , respectively, and  $(\sigma^h)^2$  and  $(\sigma^f)^2$  are the variances of the feature.

In feature selection,  $FC$  can be used to evaluate the discrimination power of each single feature [21]. However, it is not directly suitable to evaluate the discrimination power of a group of features. Thus, we proposed a novel and simplified criterion based on Fisher's discriminant, called  $F$  score [34],  $\hat{F}$ , and used it to estimate the discrimination power of a group of features (here TDP feature vector [ $A, M, C$ ]):

$$\hat{F} = \frac{\left\| \vec{\mu}^h - \vec{\mu}^f \right\|_2^2}{tr(\Sigma^h) + tr(\Sigma^f)} \quad (3)$$

where  $\Sigma$  denotes the covariance matrix of the feature vector,  $\vec{\mu}$  denotes the mean of the feature vector,  $\|\cdot\|_2$  denotes the  $L_2$ -norm (Euclidean norm), and  $tr(\cdot)$  the trace of a matrix.

Compared to  $FC$ ,  $\hat{F}$  is a derived version relying on the Euclidean distance between class centers,  $\left\| \vec{\mu}^h - \vec{\mu}^f \right\|_2$ , to estimate the difference between classes, and employing the trace of the covariance matrix to evaluate the variance within a class. Note that this simplified expression avoids estimating a projection direction as required by the general multi-dimensional expression of Fisher's LDA.

### F Score-Based Channel Selection

A spatial filtering is performed in each channel based on the small-distance Laplacian derivation [20] to reduce the

signal correlation and common noise among neighboring channels. The TDPs, [ $A_e^\chi(i), M_e^\chi(i), C_e^\chi(i)$ ], are computed for a time segment [ $t_n, t_n + T - 1$ ] for each single trial  $i$  at channel  $e$  for class  $\chi$  ( $\chi \in \{h, f\}$ ). Then, the discrimination power of channel  $e$  is estimated by the  $F$  score:

$$\hat{F}_e = \frac{(\bar{A}_e^h - \bar{A}_e^f)^2 + (\bar{M}_e^h - \bar{M}_e^f)^2 + (\bar{C}_e^h - \bar{C}_e^f)^2}{\tilde{A}_e^h + \tilde{A}_e^f + \tilde{M}_e^h + \tilde{M}_e^f + \tilde{C}_e^h + \tilde{C}_e^f} \quad (4)$$

with

$$\overline{TDP}_e^\chi = \frac{1}{K^\chi} \sum_{i=1}^{K^\chi} TDP_e^\chi(i) \quad (5)$$

$$\widetilde{TDP}_e^\chi = \frac{1}{K^\chi - 1} \sum_{i=1}^{K^\chi} (TDP_e^\chi(i) - \overline{TDP}_e^\chi)^2 \quad (6)$$

where  $\overline{TDP}_e^\chi(i)$  and  $\widetilde{TDP}_e^\chi(i)$  are the estimated means and variances of time domain parameters (TDP, i.e., [ $A_e^\chi(i), M_e^\chi(i), C_e^\chi(i)$ ]);  $K^\chi$  is the number of training trials for class  $\chi$ .

Existing methods typically determine the number of selected channels based on user's experience [30] or exhaustive searching strategy [4, 16], which is either arbitrary or time-consuming. Here, we propose an automatic approach, by considering the properties of both features and classifier to determine the size of the subset of selected channels.

Let  $\hat{F}_m$  be the largest  $F$  score among all channels. The relative discrimination power of each channel  $e$  is defined as:

$$\rho_F(e) = \frac{\hat{F}_e}{\hat{F}_{\max}} \quad (7)$$

The value of  $\rho_F(e)$  is between 0 and 1. A larger  $\rho_F(e)$  indicates a larger relative discrimination power. Thus, a threshold  $\hat{\rho}$  can be set to extract the channels with  $\rho_F(e) > \hat{\rho}$  to be used for classification. A lower value of  $\hat{\rho}$  tends to pick out more channels. In practice, the training trials should have several times as many as the dimensionality of features to guarantee a good performance of the classifier [13]. Based on this knowledge, the range of  $\hat{\rho}$  can be shrunk to [ $P, 1.0$ ] to feed the classifier, where  $P$  is obtained by:

$$\begin{aligned} &\min_P \text{Num}(P) \text{ s.t.} \\ &P \in [0, 1.0], \text{Num}(P) \geq K/3R \end{aligned} \quad (8)$$

where  $\text{Num}(P)$  is the number of selected channels with  $\rho_F(e) > P$ ,  $K$  is the number of trials for training, and  $R$  is the ratio of the number of trials to the number of features for a specific classifier. Note that each channel yields three TDPs, so here we have  $\text{Num}(P) \geq K/3R$ . As a linear classifier, such as Fisher's LDA, typically needs 5–10

times training trials as many as the dimensionality of features [18], we set  $R = 5$  to have a loose range of  $\hat{\rho}$  for further optimization. Different subsets of channels according to different  $\hat{\rho} \in [P, 1.0]$  are used to train the classifier. The optimal  $\hat{\rho}$  is obtained by seeking the subset with the lowest training error (ERR) in the classifier training. The training error is defined as the observed overall disagreement between classification outputs and true classes. If there are more than one optimal value obtained, we use the largest one.

### Channel Selection Using Time Information (CSTI)

This method aims to find the optimal combination of time segment and subset of channels for classification. The general scheme of the method, called CSTI, is shown in Fig. 1. First, we compute the TDP features and the  $F$  score for each channel in a series of overlapping  $T$ -width time segments  $[t_n, t_n + T - 1]$  ( $n = 1, \dots, N$ ),  $t_{n+1} = t_n + T_s$  ( $T_s$  is the step), during the motor imagery duration  $[T_0, T_e]$ , where  $T_0$  is the beginning time of motor imagery and  $T_e$  is the ending time. Then, the optimal subsets of channels  $S(t_n)$  and their corresponding training error  $ERR(\hat{\rho}^*(t_n))$  are obtained by the  $F$  score-based channel selection proposed above for different time segments  $[t_n, t_n + T - 1]$  ( $n = 1, \dots, N$ ), where  $\hat{\rho}^*(t_n)$  is the optimal  $\hat{\rho}$  in the time segment  $[t_n, t_n + T - 1]$ . The optimal time segment  $[t^*, t^* + T - 1]$  is found by seeking the lowest training error  $ERR(\hat{\rho}^*(t_n))$  among all time segments, in order to obtain the optimal subset of channels  $S(t^*)$  in the optimal time segment  $[t^*, t^* + T - 1]$ .

### Experimental Data and Goals

The dataset IVa [5] from BCI competition III is used in this study. As it consists of EEG signals recorded using 118 electrodes, this dataset is very suitable for a fine selection of EEG channels. Five subjects, denoted “aa,” “al,” “av,” “aw” and “ay,” have performed 280 trials of cue-driven motor imagery (right hand: 140 trials, right foot: 140 trial) during the recording. The acquisition process was driven by visual cues, presented during 3.5 s, and separated by randomly chosen intervals, ranging from 1.75 to 2.25 s. Subjects were required to perform the corresponding motor imagery task during the presentation of a cue and to relax in the intermission. Thus,  $T_0 = 0$  is the time point of the cue-onset, and  $T_e = 3.5$  s is the ending of the cue. Ground truth is available for all subjects in this dataset.

The aim of the experiment is to perform classification of the signal, for each subject, into two classes (right hand and right foot), with as few electrodes as possible. The  $F$  score-based channel selection was performed in five ( $N = 5$ ) overlapping time segments of 0–2.0, 0.5–2.5, 1.0–3.0, 1.5–

3.5 and 2.0–4.0 s after the cue on-set ( $t_n = 0, 0.5, 1.0, 1.5, 2.0$  s,  $T = 2$  s,  $T_s = 0.5$ ) to find the optimal combination of time segment and subset of channels by CSTI. To verify the importance of time segment selection, we also performed  $F$  score-based channel selection in a long time segment from the cue on-set to the ending of the cue for comparison. Moreover, we also compared our method with full EEG cap-based CSP and 3C setup. The optimal CSP patterns are selected by using an automatic algorithm proposed in our previous work [32]. *Fisher's* LDA was used as the classifier in this study, since  $F$  score is based on *Fisher's* discriminant, and it works well with TDP and BP features [18, 28]. The paired-sample  $t$  test was employed to reveal the statistical significance of the difference between the results of different methods.

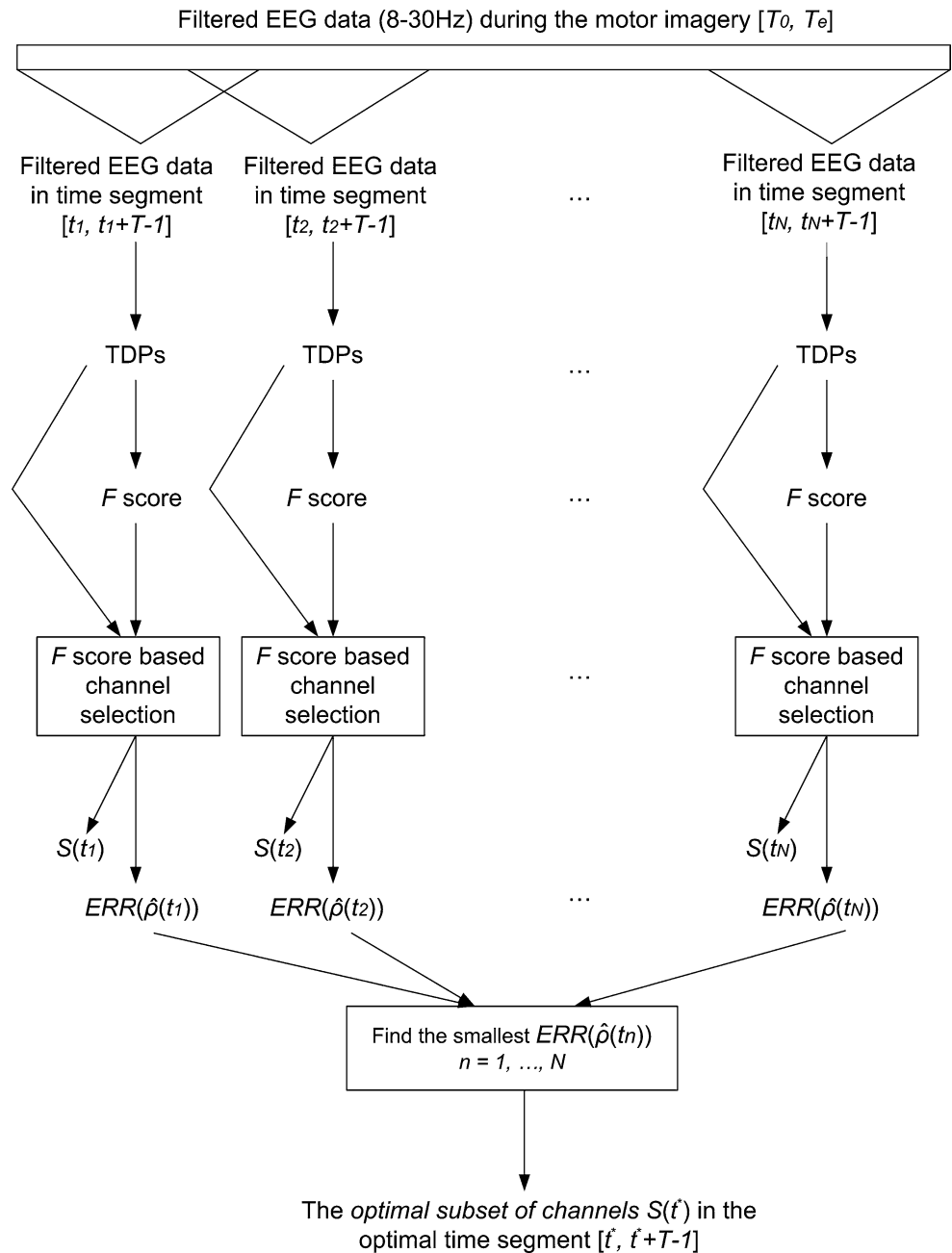
First, we used the first 70 trials for each class for training, and the remaining ones for the independent testing, to evaluate the contributions of our methods. The results are provided in “Effect of Time Segment on Channel Selection and Classification” to “Comparisons with Other Methods” sections. This choice of training/testing data corresponds to a usual situation in real applications, where the training data are recorded before the testing data. Using 50 % trials for training makes the information for training comparable to that for testing. Secondly, considering the data evolution, we also tested our method with randomly selected training and testing data (70 training trials vs. 70 testing trials for each class) to evaluate the robustness of our method. The results are provided in “Effect of Data Evolution” section.

## Results and Discussion

### Effect of Time Segment on Channel Selection and Classification

The spatial distribution of the  $F$  score and the selected electrodes in different time segments are shown in Fig. 2, where the selected time segments are marked out by squares. The testing results obtained when using the selected electrodes in different time segments of 2 s are provided in Table 1, and the results from the selected time segments are in *Italic*. The results are evaluated by classification accuracy (ACC), which is defined as the observed overall agreement between classification outputs and true classes. From Fig. 2, we can see that the subsets of selected electrodes vary with time segments for each subject, indicating that time segment is an important factor that should be considered in electrode selection. Among all possible combinations of time segment and subset of electrodes, the selected combination yields the highest classification

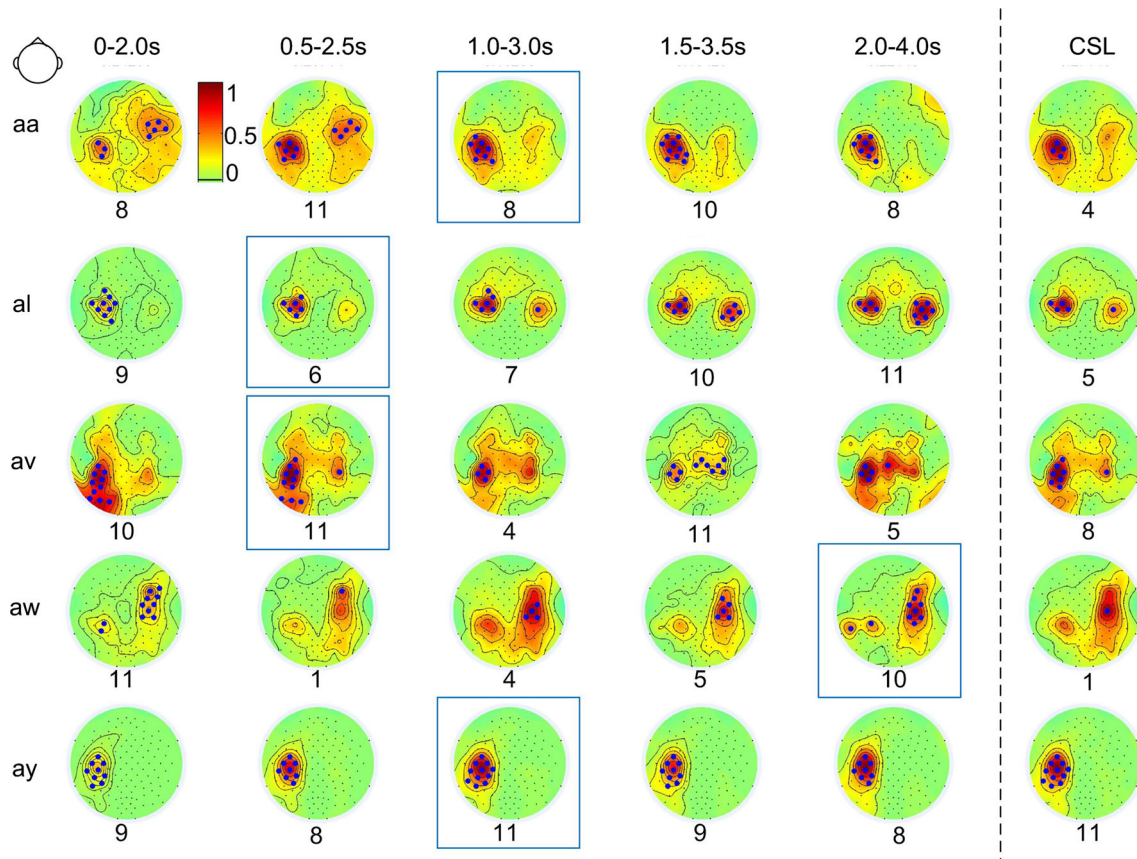
**Fig. 1** General scheme of CSTI



accuracy (ACC) on the testing data. This result shows that CSTI is effective in finding the optimal combination of time segment and subset of electrodes. However, CSTI has a computational cost, which is at least  $N$  times ( $N$  is the number of different time segments, here  $N = 5$ ) the one of the methods only performing channel selection in only one single time segment. In our experiments, the computational time for CSTI was 11 s MATLAB 7.10.0, Window 7 Professional 64 bits, CPU 2.66 GHz, RAM 2.0 GB). It was 5 times the computational time for channel selection in a single time segment (around 2 s). Nevertheless, this

additional calibration time remains acceptable for several applications, such as neuro-games [36].

In this study, we also performed  $F$  score-based channel selection in a long time segment (CSL) from the cue on-set to the cue ending (that covers the whole period of motor imagery) [35] to see: (1) whether a long time segment will improve the results of channel selection (i.e., selecting less electrodes and/or improving classification accuracy), (2) whether the effect of time segment can be ignored by using a long time segment that covers the full period of motor



**Fig. 2** Topographic maps of the  $F$  score (color scale) and selected electrodes (marked by bold points) in different time segments. The number of selected electrodes is given below each map. The selected

time segments are marked out by rectangles. Results in the long time segment obtained by CSL are provided in the last column

**Table 1** ACC results when using the selected subset of electrodes in different time segments

Subject	0–2.0 s	0.5–2.5 s	1.0–3.0 s	1.5–3.5 s	2.0–4.0 s
aa	0.53	0.67	<i>0.67</i>	0.67	0.64
al	0.78	<i>0.89</i>	0.87	0.87	0.88
av	0.49	<i>0.61</i>	0.61	0.61	0.58
aw	0.66	0.51	0.71	0.75	<i>0.81</i>
ay	0.73	0.90	<i>0.92</i>	0.84	0.80

The ACC obtained from the selected time segment is in italic

imagery, so as to save computational time. Comparisons between CSTI and CSL are provided in Table 2. Compared to CSTI, CSL selected less electrodes (except for “ay”) and used less computational time (see Fig. 2 and Table 2). However, CSL only improves ACC for one subject (“av”). For the other subjects, CSL yields significantly worse ACC

**Table 2** The effect of time information on channel selection and classification

Subject	NS		CT (s)		ACC			
	CSL	CSTI	CSL	CSTI	CSL	CSTI	FL	FTI
aa	4	8	2	11	0.61	0.67	0.47	0.59
al	5	6	2	11	0.82	0.89	0.94	0.87
av	8	11	2	11	0.63	0.61	0.69	0.62
aw	1	10	2	11	0.74	0.81	0.94	0.56
ay	11	11	2	11	0.84	0.92	0.84	0.69
Mean	6	9	2	11	0.73	0.78	0.78	0.67

NS number of selected channels, CT computational time for channel selection under the environment: MATLAB 7.10.0, Window 7 Professional 64 bits, CPU 2.66 GHz, RAM 2.0 GB, ACC classification accuracy, CSL channel selection using a long time segment, CSTI channel selection using time information, FL full-cap classification using a long time segment, LTI full-cap classification using time information



than CSTI does ( $t$  test:  $p < 0.05$ ). Considering the deteriorated ACC, using a long time segment in channel selection cannot fully improve the results for most subjects. Thus, the effect of time segment cannot simply be ignored by using the long time segment.

Although CSTI tends to select more electrodes than CSL does, the number of CSTI selected electrodes is no more than 11 (see Fig. 2 and Table 2), which is comparable to commercial BCI system Emotiv EPOC,<sup>1</sup> which has 14 electrodes. Thus, the number of electrodes selected by CSTI is still reasonable and acceptable for general applications (e.g., in a game environment).

Additionally, we also investigated the effect of time segment selection on the classification accuracy with the full-cap data. Experimental results show that time segment selection alone did not improve the classification accuracy (see Table 2). Thus, time segment selection may only be necessary with channel selection.

### Comparisons with Other Methods

Table 3 lists the testing results (evaluated by ACC) of the full-cap CSP and the 3C setup using BP and TDP features, as well as two state-of-the-art methods in channel selection using BP features, i.e., the  $l1$ -norm-based sparse CSP (SCSP) [37] and the Riemannian distance-based channel selection method (Rd) [4]. To make the comparison easy, the testing results of CSL and CSTI are also reminded in Table 3.

For the full-cap CSP as well as for the 3C setup, using TDPs yields better mean ACC ( $\overline{ACC} = 0.78$  for full-cap CSP,  $\overline{ACC} = 0.72$  for 3C setup) than using BP ( $\overline{ACC} = 0.76$  for full-cap CSP,  $\overline{ACC} = 0.71$  for 3C setup). The difference is not significant ( $p > 0.05$ ) due to the limited number of subjects in this dataset. For most subjects, using TDPs did improve ACC, which is in agreement with the results in [28]. With the BioSig toolbox [26], TDPs are easy and fast to calculate (2 ms using MATLAB 7.10.0, Window 7 Professional 64 bits, CPU 2.66 GHz, RAM 2.0 GB). Unlike BP which often requires the selection of frequency bands to improve classification results [34], TDPs save computation time during the frequency band selection. All of these indicate the interest of using TDPs in motor imagery BCI.

The results obtained using CSTI ( $\overline{ACC} = 0.78$ ) are significantly better ( $p < 0.05$ ) than simply using 3C setup ( $\overline{ACC} = 0.71$  when using BPs,  $\overline{ACC} = 0.72$  when using TDPs). The mean classification accuracy when using CSTI is better than using full-cap CSP with BP features ( $\overline{ACC} = 0.76$ , not significant with  $p > 0.05$ ) and equal to using full-cap CSP with TDP features ( $\overline{ACC} = 0.78$  when

using TDPs). For some subjects (“aa” and “ay”), CSTI even yields higher ACC than full-cap CSP. Thus, CSTI meets the goal of largely reducing the number of electrodes (from 118 channels to 9 in average), without a drop of the mean classification performance. This result is better than the  $l1$ -norm-based sparse CSP (SCSP) [37] with higher mean ACC (0.78 vs. 0.73) and less selected channels (9 vs. 13). Although there is no difference between CSTI and the Riemannian distance-based method (Rd) [4] in the mean ACC over subjects (both  $\overline{ACC} = 0.78$ ), CSTI selects slightly less channels than the Riemannian distance-based method (9 vs. 10) and leads to better individual results in three out of five subjects (subject “av,” “aw” and “ay”). Moreover, CSTI uses a relatively shorter time segment (2 s length) than the methods in comparison (3.5 s length). For most subjects (except “aw”), the classification outputs are obtained before the ending of cue, which indicates that less time (here, less than 3.5 s) is required for recording the training data for these subjects.

Recently, Wang et al. [29] have introduced a sophisticated method for a similar purpose as our method (CSTI). Their experimental results showed that their method can simultaneously achieve channel and feature selection with a lower error rate (22.22 %). Thus, their classification performance could be similar to our method (CSTI). However, their method selected a larger number of channels (i.e., 17–23 channels) than our method for motor imagery BCI.

Additionally, we found that CSL generates slightly better mean ACC ( $\overline{ACC} = 0.73$ ) than simply using 3C setup. However, this improvement is not significant ( $p > 0.05$ ) and does not occur for all subjects. Moreover, CSL tends to select more than three channels and needs a full EEG cap to acquire training data for seeking the optimal subset of electrodes. Thus, CSL is not cost-efficient in real applications.

Among all methods, the mean ACC of 3C setup is the worst, but it uses the least number of electrodes (only three channels) and can yield better ACC than the full-cap CSP for one subject in the dataset (“aa”). Moreover, 3C setup has no additional computational cost and does not need full-cap training data for calculating CSP filters or seeking the optimal subset of electrodes. Thus, for electrode reduction, the choice between CSTI and 3C setup may depend on a preference between the number of electrodes, the computation cost, the amount of training data and the classification performance. This choice can be left to the user.

### Effect of Electrode Misplacement

The electrode positions might have undergone slight changes compared to the standard 10–20 recording system

<sup>1</sup> <https://www.emotiv.com/>.

**Table 3** Classification accuracy (ACC) and mean number of electrodes for different methods. The best ACC with selected channels for each subject is marked in bold

Subject	ACC for full-cap (118)		ACC for selected channels (average number of channels)					
	CSP		3C setup (3)		SCSP (13)	Rd (10)	CSL (6)	CSTI (9)
	BP	TDPs	BP	TDPs	BP	BP	TDPs	TDPs
aa	0.46	0.47	0.64	0.59	0.57	<b>0.74</b>	0.61	0.67
al	0.94	0.94	0.79	0.81	0.87	<b>0.98</b>	0.82	0.88
av	0.68	0.69	0.58	0.59	0.54	0.59	<b>0.63</b>	0.61
aw	0.94	0.94	0.73	0.79	<b>0.84</b>	0.80	0.74	0.81
ay	0.75	0.84	0.81	0.82	0.84	0.81	0.84	<b>0.92</b>
Mean	0.76	0.78	0.71	0.72	0.73	<b>0.78</b>	0.73	<b>0.78</b>

[14] in real applications, in particular for general users who may not be proficient in EEG recording. For example, an inexperienced user may put the EEG cap a little bit left; as a result, all electrodes are placed at the left side of the standard positions during the recording.

In practice, the training and testing data may be recorded in two different ways. In the first way, they are recorded in one session without re-placing the electrodes. In this case, if misplacement happens, both the training and testing data are recorded at the same non-standard positions. For machine learning-based methods, e.g., CSTI, the effect of electrode misplacement can be neglected, because the optimal subset of electrodes is estimated based on the actual positions, where the data are recorded, instead of standard positions, while for 3C setup, this effect should be examined, because the selected channels (C3-Cz-C4) are defined according to the standard positions. When the cap is put incorrectly, nominal channels (C3, Cz and C4) of 3C setup will not be in their standard positions.

In the second way, the training and testing data are recorded in two sessions (maybe in two different days) with re-placing the electrodes. As a result, the training and testing data may be recorded at different non-standard positions. Usually, not only the shift of electrodes should be considered in this case, but also the change of the mental state of the user [6]. It is a very complicated problem, so-called the challenge of “session-to-session transfer” [6]. In fact, all methods face this challenge. As both the change of mental state and the shift of electrodes may exist but are unpredictable, even if a method has achieved a good performance in one “session-to-session transfer” test, it may fail in the next one if the changes are too large. In real applications, commercial BCI systems (Emotiv and Neurosky) require the user to wait a few seconds (or minutes) for calibration after putting the cap (to check the electrode impedance) and to perform a training session with feedback before the real play, to overcome this challenge. As a result, this calibration costs users some additional time for collecting the training dataset.

To examine the effect of electrode misplacement on 3C setup, we compared the classification results obtained using

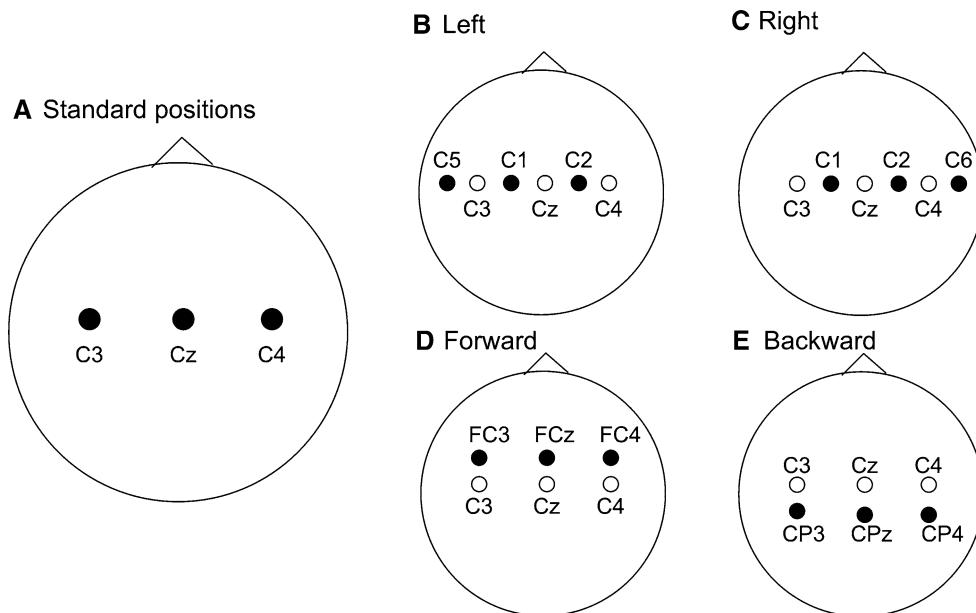
the standard 3C setup (C3-Cz-C4) and using the non-standard 3C setup with the electrodes placed a little left (C5-C1-C2), right (C1-C2-C6), forward (FC3-FCz-FC4), backward (CP3-CPz-CP4) with respect to the standard positions (see Fig. 3). Table 4 shows that using the electrodes placed a little backward, the classification results are improved for subjects “aa,” “al” and “av,” but deteriorated for subjects “aw” and “ay.” However, for all subjects, the results using the electrodes placed a little forward are significantly worse than using the electrodes placed at the standard positions ( $p < 0.01$ ) and a little backwards ( $p < 0.01$ ). Using electrodes placed a little left or right, the results are deteriorated compared to those obtained with the electrodes placed at the standard positions. Compared to those obtained with the electrodes placed a little right, the results obtained when using the electrodes placed a little left are better for subjects “aa,” “av” and “ay,” but worse for subjects “al” and “aw.”

Figure 4 shows that the large values of  $F$  score are mainly distributed in the post-central areas of the brain for all subjects,<sup>2</sup> which explains why using the electrodes placed a little backward always generates better results than using the electrodes placed a little forward. Meanwhile, for subjects “aa,” “av” and “ay,” the distributions of large values of  $F$  score show a left-brain dominance. Thus, the results obtained with the electrodes placed a little left are better than those obtained with the electrodes placed a little right for those subjects.

To sum up, the effect of changes of electrode position on classification results depends on the subject and the direction of error placement. As an inexperienced user may unconsciously misplace the electrodes, the effect will be unpredictable when simply using 3C setup and may lead to a deteriorated result. Concerning this effect, CSTI can be recommended to users who are not very professional in EEG recording. However, training data and computation

<sup>2</sup> We provide the locations of electrodes selected by CSL instead of CSTI for comparison, since this part of analysis is based on the whole time period for 3C setup.

**Fig. 3** Standard 3C setup (C3-Cz-C4) and the non-standard 3C setups with the electrodes placed a little left (C5-C1-C2), right (C1-C2-C6), forward (FC3-FCz-FC4), backward (CP3-CPz-CP4)



**Table 4** ACC results for standard 3C setup (C3-Cz-C4) and the non-standard 3C setups with the electrodes placed a little left (C5-C1-C2), right (C1-C2-C6), forward (FC3-FCz-FC4), backward (CP3-CPz-CP4)

Subject		Standard	Left	Right	Forward	Backward
aa	BP	0.64	0.57	0.48	0.52	<i>0.69</i>
	TDP	0.59	0.56	0.51	0.49	<i>0.63</i>
al	BP	0.79	0.65	0.74	0.61	<i>0.80</i>
	TDP	0.81	0.70	0.77	0.68	<i>0.86</i>
av	BP	0.58	0.54	0.53	0.59	<i>0.61</i>
	TDP	0.59	<i>0.64</i>	0.59	0.62	0.62
aw	BP	<i>0.73</i>	0.58	0.66	0.65	0.66
	TDP	<i>0.79</i>	0.54	0.66	0.64	0.70
ay	BP	<i>0.81</i>	0.77	0.50	0.69	0.79
	TDP	<i>0.82</i>	0.74	0.59	0.66	0.77

The best performance for each subject is in *Italic*

time are needed for finding the optimal subset of electrodes.

**Effect of Data Evolution**

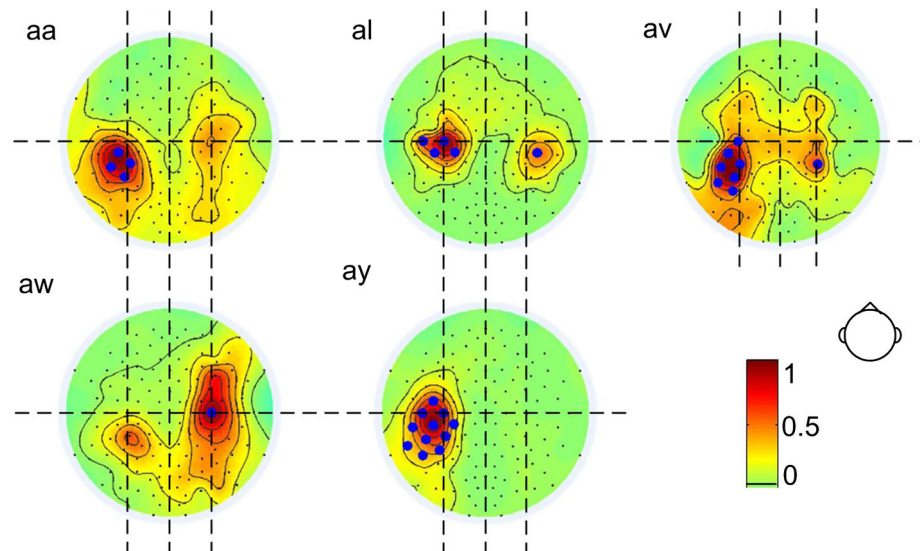
The non-stationarity of EEG is a common problem in BCI [9]. As mentioned above, it is common to discuss this issue for session-to-session transfer. However, the data evolution problem may also exist in one session data when the recording period is relatively long, since the non-stationarity of EEG can result from several causes. For example, changes in electrode impedance may occur when the electrically conductive gel between skin and electrode dries

out or an electrode gets loose. Additionally, the task involvement and attention level of a subject may change over the course of a BCI experiment. All these factors will lead to some unpredictable modulations in EEG signals even when both training and testing data are recorded in the same session, resulting in a poor SNR in a time segment or at a channel, which may impact the selection of time segment and channel.

To examine this effect, we randomly selected 140 trials (70 trials for right hand and 70 trials for right foot) as the training dataset to find the optimal combination of time segment and subset of electrodes by CSTI for each subject, the remaining data forming the testing dataset. We repeated this procedure 100 times. For comparison, we also calculated the subset of electrodes based on the long time segment by CSL.

The experimental results generated by CSTI show that the optimal time segments are not always the same for different training datasets even for the same subject. A possible reason for this result is that the subject may not have the same response time to the cue in different trials due to different mental states and possible fatigue during the BCI experiment [10]. The distribution of optimal time segments for each subject is given in Fig. 5. It shows that the optimal time segments mainly appear in the range of 0.5–3.0 s (i.e., the second and third time segments) for subjects “aa,” “av” and “ay,” while a little bit later (i.e., the fourth time segment 1.0–3.5 s) for subjects “al” and “aw,” indicating that some subjects may need relatively longer time for recording the useful data in each trial compared to other subjects. The subsets of selected electrodes also vary with different training datasets for the

**Fig. 4** Distribution of  $F$  score for different subjects in the long time segment from the cue on-set to the ending of the cue. Electrodes selected by CSL are marked by *bold* points



same subject. These results indicate that the effect of data evolution exists not only for “session-to-session transfer” but also when the training and testing data are recorded in the same session.

The probabilities of channels being selected are shown in Fig. 6. The red areas indicate the important brain areas where the channels are often selected. We also marked out the key channels with the selection probabilities above 80 %. The similarity is shown between CSTI and CSL, although there are more key channels when using CSTI. For most subjects (except subject “av”), the key channels are distributed over the hand representative area of the sensorimotor cortex. Motor imagery of the right hand typically elicits strong ERD in the hand representative area of the sensorimotor cortex of the left brain (see Fig. 7). Nevertheless, for some subjects (e.g., subjects “al” and “aw”), the key electrodes are also found over the right hemisphere (see Fig. 6). The reason is that motor imagery can also cause an ERS in a “non-active” area [24]. For example, performing a foot motor imagery can generate an ERS in the hand representation area (see Fig. 7). The ERS can also contribute to classification [25]. Channels in central, frontal and occipital cortices are with very low selection probabilities, indicating that those areas are less important for distinguishing motor imagery of foot and hand. This result implies the possibility of using a part instead of all of the electrodes in an EEG cap to find the optimal subset of channel.

Among all subjects, subject “av” does not have any key channels. Thus, compared to other subjects, subject “av” needs a relatively larger number of electrodes and computation time for finding the optimal subset of electrodes.

## Conclusions and Future Work

Although earlier studies have presented the need for selecting and reducing the electrodes required in a BCI system [4, 16, 17, 30], they addressed this issue based only on spatial information, disregarding the potential impact of temporal information. The contribution in this paper, with the proposition of a novel method, CSTI, emphasizes the potential effects of the chosen time segment on channel selection. A criterion derived from *Fisher’s* criterion is proposed to evaluate the discrimination power of a group of features and applied on time domain parameters (TDP), which overcomes the disadvantage of classical *Fisher’s* criterion [21] on TDP feature selection.

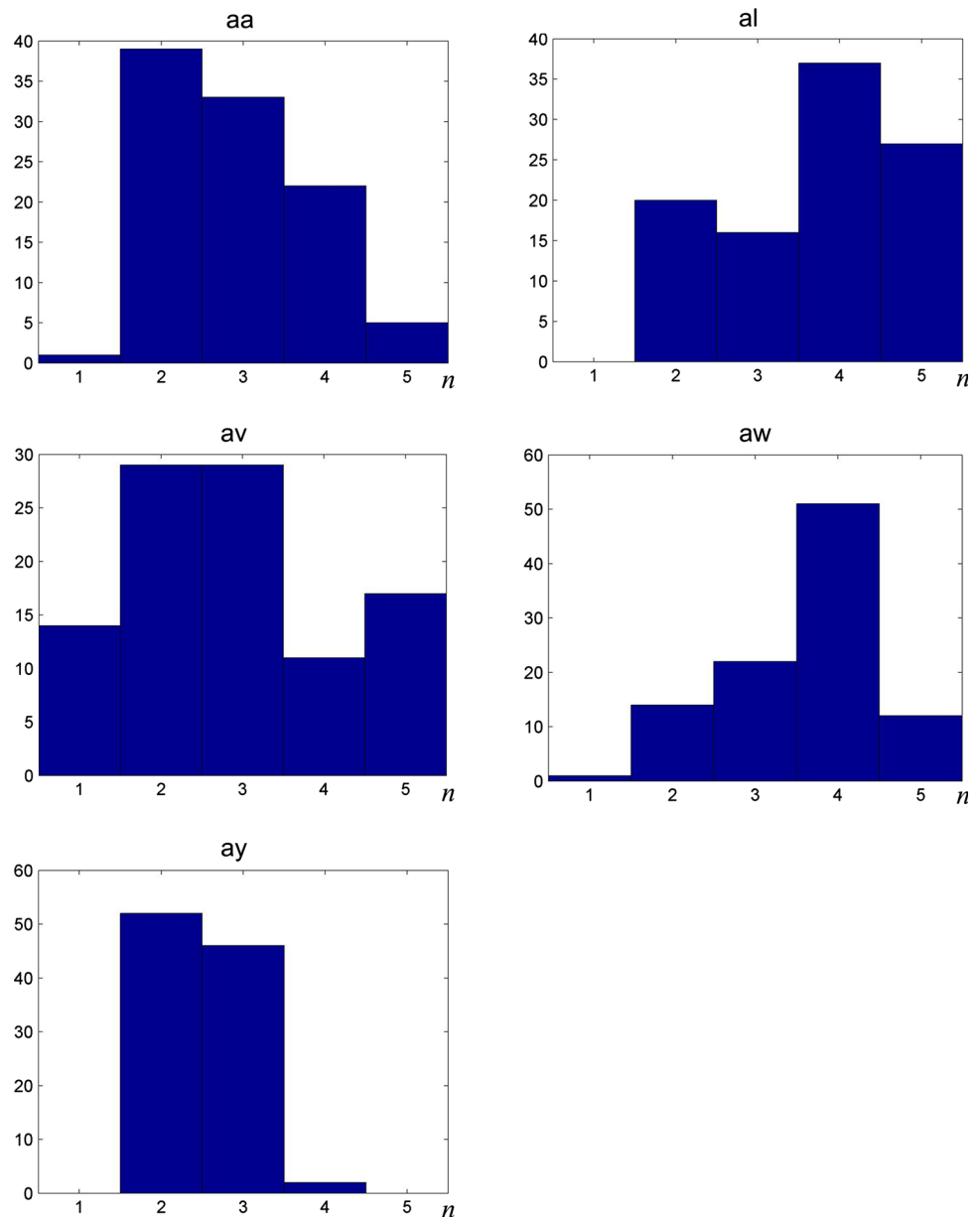
Comparisons between CSTI, CSL, 3C setup and full-cap CSP were performed. The comparisons of their average performances on classification accuracy and reducing the number of channels, their computational costs and training data required for finding the optimal subset of electrodes can be summarized as follows:

- Mean classification accuracy: 3C setup < CSL < CSTI = full-cap CSP;
- Mean number of channels used: 3C setup < CSL < CSTI < full-cap CSP;
- Computational cost for finding the optimal subset of electrodes: 3C setup = full-cap CSP (no computational cost) < CSL (2 s in the experiment) < CSTI (11 s);
- Training data required for finding the optimal subset of electrodes: 3C setup = full-cap CSP (not needed) < CSL = CSTI (needed).

A full-cap setup with the CSP algorithm employs the largest number of electrodes among all methods. The



**Fig. 5** Distribution of optimal time segment for each subject. The horizontal axis  $n$  indicates the time segments  $[t_n, t_n + T - 1]$  ( $n = 1, \dots, 5$ ). The vertical axis shows the number of times each time segment is selected



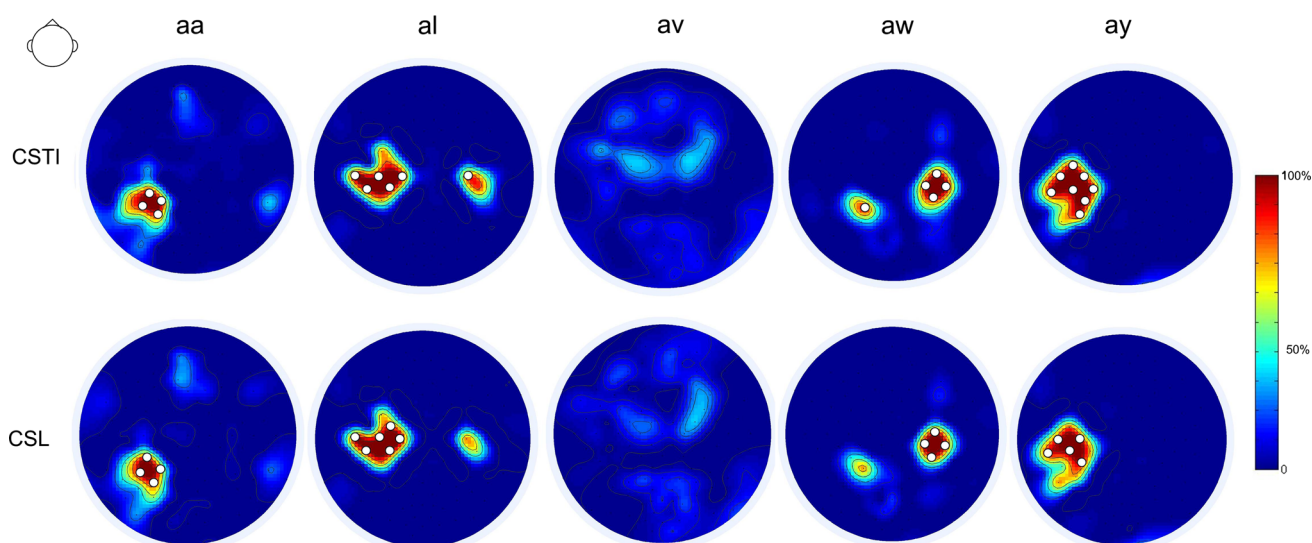
tedious placement of EEG electrodes unavoidably reduces its practicability in non-clinical applications, such as for a home use of BCI systems. Moreover, the classification performance obtained by full-cap CSP is not always the best and may be even worse than 3C setup in some cases. Thus, the classification performance is not proportional to the number of electrodes, and it is possible to reduce the number of electrodes without deteriorating the classification results.

The 3C setup uses only three channels (C3, Cz and C4) that cover the sensorimotor areas of the brain. This setting has the lowest number of electrodes and does not need a standard EEG cap, training data and computation time to find the optimal subset of electrodes. It is an ideal choice when only very few electrodes (i.e., less than 10) are

available. However, in most cases, its classification accuracies are not as good as for other methods due to the limited information it exploits. Moreover, the 3C setup relies on a precise placement of electrodes, so it may not be easy to use for users who are not professional in EEG data recording.

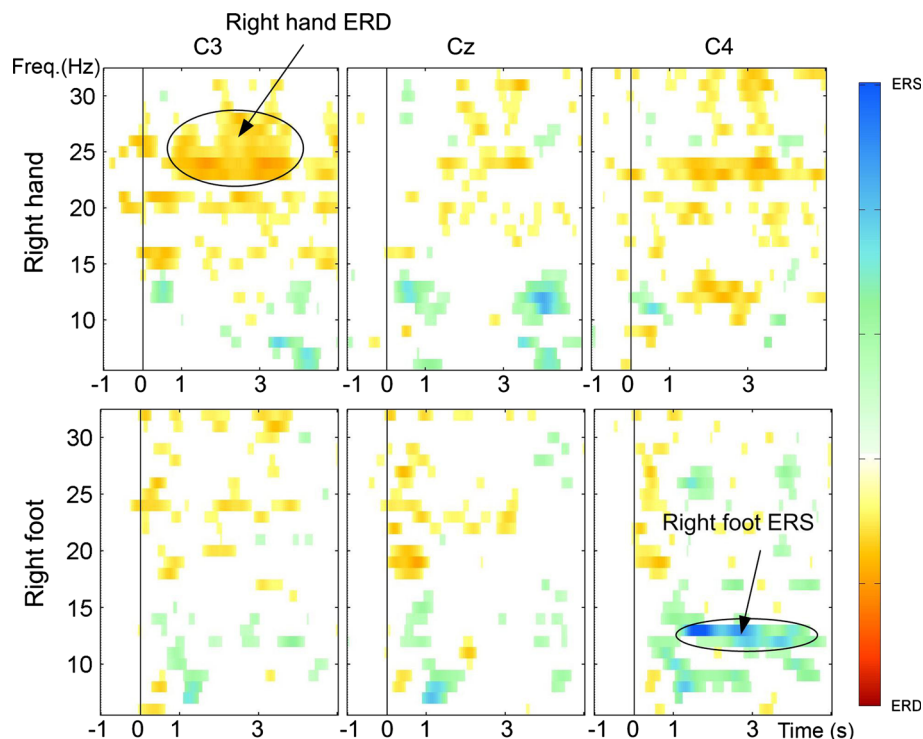
CSL often chooses more than 3 channels for classification; however, it can only slightly improve classification accuracy compared to 3C setup. Thus, it may not be a good choice in most cases.

CSTI can largely reduce the number of channels (from 118 channels to 9 in average), shorten the time window length and achieve the mean classification accuracy comparable to the full-cap CSP. Compared to two existing



**Fig. 6** Topography of selection probabilities of channels for each subject. The key channels with probabilities more than 80 % are indicated by “o.” CSTI: channel selection using time information, CSL: channel selection using a long time segment

**Fig. 7** Time–frequency visualization of ERD/ERS for subject “aw.” It was generated by the BioSig toolbox [26], using overlapping 2 Hz bands (step = 1 Hz) in the frequency range between 6 and 32 Hz, from 1 s before cue on-set to 4 s after cue on-set (for details, see [25])



channel selection methods, the experimental results on a publicly accessible BCI dataset show that our method is better than them with less selected electrode and higher classification accuracy for most subjects. The number of electrodes selected by CSTI is less than that of the commercial BCI system Emotiv EPOC. Thus, our method can be used in designing BCI systems using few channels (electrodes) for subject-specific applications. This work

can also help the BCI system designer to decide on the best compromise between accuracy, easy use and portability, according to the user’s needs.

In this study, we performed a subject-specific channel selection. Although a non-subject-specific channel selection seems more promising, the individual differences between subjects are still hard to overcome. A non-subject-specific channel selection based on the training datasets

recorded from a few subjects may not capture the whole inter-individual variability. A robust non-subject-specific selection requires a very large database, and estimating its minimum size is still an open question. In the future, we will try to solve this problem and extend the study to multi-class BCIs.

**Acknowledgments** Authors would like to thank Dr. Olexiy Kyrzyov and Dr. Teodoro Solis-Escalante for their useful discussions and Prof. Benjamin Blankertz for providing the BCI dataset: open access BCI dataset, i.e. the dataset IVa [5] from BCI competition III (<http://www.bbci.de/competition/iii/>). This work was partially supported by grants from China Scholarship Council and Orange Labs.

#### Compliance with Ethical Standards

**Conflict of Interest** Yuan Yang, Isabelle Bloch, Sylvain Chevallier, and Joe Wiart declare that they have no conflict of interest.

**Informed Consent** All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2008 (5).

**Human and Animal Rights** This article does not contain any studies with human or animal subjects performed by the any of the authors.

#### References

- Allison BZ, Wolpaw EW, Wolpaw JR. Brain-computer interface systems: progress and prospects. *Expert Rev Med Dev.* 2007;4(4):463–74.
- Alotaiby T, El-Samie FEA, Alshebeili SA, Ahmad I. A review of channel selection algorithms for EEG signal processing. *EURASIP J Adv Signal Process.* 2015;2015(1):1–21.
- Ansuini C, Cavallo A, Koul A, Jacono M, Yang Y, Becchio C. Predicting object size from hand kinematics: a temporal perspective. *PLoS One.* 2015;10(3):e0120432.
- Barachant A, Bonnet S. Channel selection procedure using Riemannian distance for BCI applications. In: 5th international IEEE/EMBS conference on neural engineering (NER 2011). 2011; p. 348–351.
- Blankertz B, Dornhege G, Krauledat M, Müller KR, Kunzmann V, Losch F, Curio G. The Berlin brain-computer interface: EEG-based communication without subject training. *IEEE Trans Neural Syst Rehabil Eng.* 2006;14(2):147–52.
- Blankertz B, Müller KR, Krusienski DJ, Schalk G, Wolpaw JR, Schlögl A, Pfurtscheller G, Millán JR, Schroder M, Birbaumer N. The BCI competition III: validating alternative approaches to actual BCI problems. *IEEE Trans Neural Syst Rehabil Eng.* 2006;14(2):153–9.
- Blankertz B, Tomioka R, Lemm S, Kawanabe M, Müller KR. Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal Process Mag.* 2008;25(1):41–56.
- Duan L, Zhong H, Miao J, Yang Z, Ma W, Zhang X. A voting optimized strategy based on ELM for improving classification of motor imagery BCI data. *Cogn Comput.* 2014;6(3):477–83.
- Gouy-Pailler C, Congedo M, Brunner C, Jutten C, Pfurtscheller G. Nonstationary brain source separation for multiclass motor imagery. *IEEE Trans Biomed Eng.* 2010;57(2):469–78.
- Gouy-Pailler C, Sebag M, Larue A, Souloumiac A. Single trial variability in brain-computer interfaces based on motor imagery: learning in the presence of labeling noise. *Int J Imaging Syst Technol.* 2011;21(2):148–57.
- Graimann B, Pfurtscheller G. Quantification and visualization of event-related changes in oscillatory brain activity in the time-frequency domain. *Progr Brain Res.* 2006;159:79–97.
- He L, Hu Y, Li Y, Li D. Channel selection by Rayleigh coefficient maximization based genetic algorithm for classifying single-trial motor imagery EEG. *Neurocomputing.* 2013;121:423–33.
- Jain A, Zongker D. Feature selection: evaluation, application, and small sample performance. *IEEE Trans Pattern Anal Mach Intell.* 1997;19(2):153–8.
- Jasper HH. The ten-twenty electrode system of the International Federation. *Electroencephalogr Clin Neurophysiol.* 1958;10(2):371–5.
- Kyrzyov O, Bloch I, Yang Y, Wiart J, Souloumiac A. Data ranking and clustering via normalized graph cut based on asymmetric affinity. In: *Image Analysis and Processing-ICIAP 2013.* Springer; 2013. p. 562–571.
- Lal TN, Schröder M, Hinterberger T, Weston J, Bogdan M, Birbaumer N, Schölkopf B. Support vector channel selection in BCI. *IEEE Trans Biomed Eng.* 2004;51(6):1003–10.
- Li Y, Koike Y. A real-time BCI with a small number of channels based on CSP. *Neural Comput Appl.* 2011;20(8):1187–92.
- Lotte F, Congedo M, Lécuyer A, Lamarche F, Arnaldi B. A review of classification algorithms for EEG-based brain-computer interfaces. *J Neural Eng.* 2007;4:R1–14.
- McFarland DJ, Wolpaw JR. Brain-computer interface operation of robotic and prosthetic devices. *Computer.* 2008;41(10):52–6.
- McFarland Dennis J, McCane Lynn M, David Stephen V, Wolpaw Jonathan R. Spatial filter selection for EEG-based communication. *Electroencephalogr Clin Neurophysiol.* 1997;103(3):386–94.
- Müller KR, Krauledat M, Dornhege G, Curio G, Blankertz B. Machine learning techniques for brain-computer interfaces. *Biomed Eng Biomed Tech.* 2004;49(1):11–22.
- Naeem M, Brunner C, Pfurtscheller G. Dimensionality reduction and channel selection of motor imagery electroencephalographic data. *Comput Intell Neurosci.* 2009;1–8:2009.
- Naji M, Firoozabadi M, Azadfallah P. Classification of music-induced emotions based on information fusion of forehead biosignals and electrocardiogram. *Cogn Comput.* 2014;6(2):241–52.
- Neuper C, Wörtz M, Pfurtscheller G. ERD/ERS patterns reflecting sensorimotor activation and deactivation. *Progr Brain Res.* 2006;159:211–22.
- Pfurtscheller G, Brunner C, Schlögl A, Lopes da Silva FH. Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks. *NeuroImage.* 2006;31(1):153–9.
- Schlögl A, Brunner C. BioSig: a free and open source software library for BCI research. *Computer.* 2008;41(10):44–50.
- Shan H, Xu H, Zhu S, He B. A novel channel selection method for optimal classification in different motor imagery BCI paradigms. *Biomed Eng Online.* 2015;14(1):93.
- Vidaurre C, Kramer N, Blankertz B, Schlögl A. Time domain parameters as a feature for EEG-based brain-computer interfaces. *Neural Netw.* 2009;22(9):1313–9.
- Wang J, Xue F, Li H. Simultaneous channel and feature selection of fused EEG features based on sparse group lasso. *BioMed Res Int.* 2015;2015:703768.
- Wang Y, Gao S, Gao X. Common spatial pattern method for channel selection in motor imagery based brain-computer interface. In: *27th IEEE annual international conference of the engineering in medicine and biology society (EMBS 2005).* 2006; p. 5392–5395.
- Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM. Brain-computer interfaces for communication and control. *Clin Neurophysiol.* 2002;113(6):767–91.

32. Yang Y, Chevallier S, Wiart J, Bloch I. Automatic selection of the number of spatial filters for motor-imagery BCI. In: 20th European symposium on artificial neural networks, computational intelligence and machine learning (ESANN 2012). 2012; p. 109–114.
33. Yang Y, Chevallier S, Wiart J, Bloch I. Time-frequency selection in two bipolar channels for improving the classification of motor imagery EEG. In 34th IEEE annual international conference of engineering in medicine and biology society (EMBC 2012). 2012; p. 2744–2747.
34. Yang Y, Chevallier S, Wiart J, Bloch I. Time-frequency optimization for discrimination between imagination of right and left hand movements based on two bipolar electroencephalography channels. *EURASIP J Adv Signal Process.* 2014;2014(1):38.
35. Yang Y, Kyrgyzov O, Wiart J, Bloch I. Subject-specific channel selection for classification of motor imagery electroencephalographic data. In: IEEE international conference on acoustics, speech and signal processing (ICASSP 2013). 2013; p. 1277–1280.
36. Yang Y, Wiart J, Bloch I. Towards next generation human–computer interaction–brain–computer interfaces: applications and challenges. In: 1st international symposium of Chinese CHI (Chinese CHI 2013). 2013; p. 1–2.
37. Yong X, Ward RK, Birch GE. Sparse spatial filter optimization for EEG channel reduction in brain-computer interface. In: IEEE international conference on acoustics, speech and signal processing (ICASSP 2008). 2008; p. 417–420.