Biomedical Signal Processing and Control 38 (2017) 302–311

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Subject-specific time-frequency selection for multi-class motor imagery-based BCIs using few Laplacian EEG channels☆

Yuan Yang[a,c,d,*], Sylvain Chevallier[b], Joe Wiart[a,c], Isabelle Bloch[a,c]

[a] LTCI, CNRS, Télécom ParisTech, Université Paris-Saclay, Paris, France
[b] Université de Versailles St-Quentin, Vélizy, France
[c] Whist Lab, Paris, France
[d] Department of Biomechanical Engineering, Delft University of Technology, Delft, The Netherlands

ARTICLE INFO

Article history:
Received 6 February 2017
Received in revised form 9 June 2017
Accepted 30 June 2017

Keywords:
FDA-type F-score
Time-frequency selection
Multi-class classification
Brain–computer interfaces
Motor imagery

ABSTRACT

The essential task of a motor imagery brain–computer interface (BCI) is to extract the motor imagery-related features from electroencephalogram (EEG) signals for classifying motor intentions. However, the optimal frequency band and time segment for extracting such features differ from subject to subject. In this work, we aim to improve the multi-class classification and to reduce the required EEG channel in motor imagery-based BCI by subject-specific time-frequency selection. Our method is based on a criterion namely Fisher discriminant analysis-type F-score to simultaneously select the optimal frequency band and time segment for multi-class classification. The proposed method uses only few Laplacian EEG channels (C3, Cz and C4) located around the sensorimotor area for classification. Applied to a standard multi-class BCI dataset (BCI competition III dataset IIIa), our method leads to better classification performance and smaller standard deviation across subjects compared to the state-of-art methods. Moreover, adding artifacts contaminated trials to the training dataset does not necessarily deteriorate our classification results, indicating that our method is tolerant to artifacts.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Brain–computer interfaces measure specific brain activities (e.g. attention level, motor imagery) and then decode them to build a direct interaction between brains and computers [1]. Existing BCIs can be controlled by various types of signals from the brain, such as electroencephalography (EEG) and fMRI. As it is inexpensive, portable and non-invasive, EEG seems to be the most promising brain signals for BCI [2]. Numerous BCI systems has been proposed using EEG signals, such as P300 BCI [3,4], SSVEP BCI [5,6], motor imagery BCI [7] as well as hybrid BCI using more than one type of signal [8,9].

EEG oscillations at μ and β bands measured around sensorimotor cortex are known as sensorimotor rhythms and associated with body movements or motor imagery [10]. Motor imagery (MI)-based BCI is one typical EEG-based BCI, which predicts the subject's motor intentions through classifying sensorimotor rhythms. Several factors indicate that MI-based BCI is quite promising for motor rehabilitation [11] and general public applications. On the one hand, this type of BCI employs the signals from the sensorimotor cortex, which are directly linked to the motor output pathway in the brain. MI-based BCI realizes motor tasks without involving the spinal cord and the periphery, and therefore can be used to help patients with spinal cord injury [12] and lock-in syndrome [13]. On the other hand, MI-based BCI can be driven by voluntary brain activities without any cues from conditioning protocols, so that the control can be independent and self-paced [14]. This advantage makes MI-based BCI also suitable for virtual reality and neuro-games [15–17].

However, individual differences of sensorimotor rhythms (frequency bands, time windows) and poor signal-to-noise ratio (SNR) impair the inter-subject robustness of MI-based BCI. Although data driven spatial filtering algorithms, such as common spatial pattern (CSP) [18,19] and independent component analysis [20], can greatly improve the SNR of sensorimotor rhythms, they usually require a large number of EEG channels. Multi-channel EEG recording reduces the portability of daily use BCI and therefore constitutes a main drawback for end users [21,22].

☆ This paper is based on Yuan Yang's Ph.D. work in Télécom ParisTech. Yuan Yang has finished his PhD in Télécom ParisTech and been with Department of Biomechanical Engineering, Delft University of Technology, Delft, the Netherlands since November 2013.

* Corresponding author.
E-mail address: Y.Yang-2@tudelft.nl (Y. Yang).

http://dx.doi.org/10.1016/j.bspc.2017.06.016
1746-8094/© 2017 Elsevier Ltd. All rights reserved.

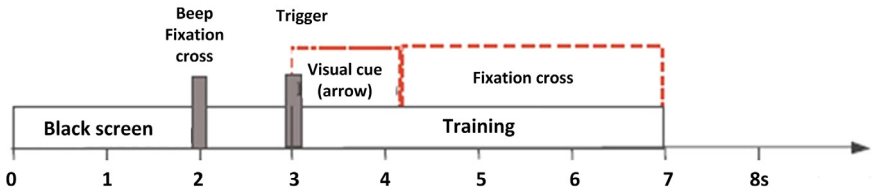

Fig. 1. Timing of the experimental paradigm for BCI competition III dataset IIIa [34].

Thus, many methods have been developed to reduce the number of channels in MI-based BCI by using machine learning techniques to select an optimal channel subset [21–27]. Due to individual differences between subjects, the estimated optimal subsets of channels usually vary with subjects. A high-density EEG recording may still be needed for finding an optimal subset for each individual. Hence, those methods did not really improve the portability and practicability of BCI. A few studies aimed to address this problem by using transfer learning frameworks to provide task-specific methods, so as to seek the channel subset that could be shared among all subjects for a particular MI task [28]. However, their classification performances are usually not as good as those obtained by subject-specific channel selection approaches. Moreover, task-specific methods are difficult to apply to the multi-class BCI, since the optimal channel subset is for a specific MI task (e.g. hand MI).

To address this challenge, we have proposed an alternative solution to increase inter-subject robustness of MI-based BCI using few EEG channels [29,30]. Instead of using machine learning techniques to select the optimal subset of channels, we simply used a few EEG channels located around the sensorimotor cortex. By selecting optimal time-frequency areas to extract subject-specific band power (BP) features, our preliminary study yielded better classification performances using fewer channels than the state-of-art methods in decoding hand MI [29,30]. Similar strategies have also been used in some other recent studies. For example, Yang and colleagues proposed a subject-based Fisher wavelet packet decomposition method using a few EEG channels to improve classification accuracy and inter-subject robustness of MI-based BCI [31]. Luo and colleagues developed a dynamic frequency feature selection method based on three bipolar channels for a similar purpose [32]. Liang and colleagues proposed a Hilbert-Huang transform based algorithm to extract the subject-specific time-frequency area for the hand MI discrimination based on two Laplacian channels [17]. All these studies focused on the two-class BCIs mainly for decoding the hand MI tasks, though multi-class BCIs can be more useful in daily applications. The optimal time-frequency areas for extracting the features could be different for different MI tasks. Increasing the number of classes will increase the inter-subject difference of optimal time-frequency areas as well as the dimensionality of features. All these eventually impair the inter-subject robustness.

In this work we focused on multi-class BCI to address this challenge. We proposed a novel method to select the optimal time-frequency areas for extracting subject-specific features, in order to improve the multi-class BCI performance using few EEG channels. The proposed method is based on a criterion called FDA-type F-score to simultaneously select both frequency band and time segment for extracting optimal BP features. FDA-type F-score is a simplified measure based on Fisher discriminant analysis (FDA) for measuring the discriminative power of a group of features. In our previous studies, we have successfully applied this criterion to seek the optimal channel subset for two-class BCIs [21] and to select the best kinematic feature group for two-class motor recognition [33]. Based on our preliminary work for two-class cases, here we extended the application of FDA-type F-score to multi-class cases using a one-versus-rest strategy. We evaluated our method in a standard multi-class BCI dataset (BCI competition III dataset IIIa [34]) to classify four classes of MI tasks (left hand, right hand, foot and tongue). This dataset has 60 channels EEG. Thus, it is suitable to test the effectiveness of our methods in improving multi-class classification when using a reduced number of EEG channels, and to compare with the state-of-the-art methods using full 60 channels or reduced channels. Meanwhile, part of this data has artifacts, so this dataset is also useful for evaluating the robustness of our methods to artifacts.

Table 1
Numbers of training and testing trials in BCI competition III dataset IIIa [34].

Subject	Training trials			Testing trials
	Clean trials	AC trials	Total	
k3	149	31	180	180
k6	92	28	120	120
l1	84	36	120	120

We used only three Laplacian channels C3, Cz and C4 (according to the standard 10/20 EEG recording system) out of 60 EEG channels. These channels are around the sensorimotor cortex, providing a better signal-to-noise ratio than the other channels, and are thus more important in decoding MI tasks [10,35]. The results are compared with those obtained by using other feature extraction methods based on all 60 channels [34,36–41], as well as using BP features extracted from the broad band (8–30 Hz) and full length (from the start to the end of motor imagery) EEG signals at the same three Laplacian channels. Furthermore, we also test our method with different amounts of artifacts contaminated trials to evaluate its robustness to artifacts.

2. Materials and methods

2.1. Experimental dataset

The BCI competition III dataset IIIa [34] was used in this study. This dataset is provided by Laboratory of Brain–Computer Interfaces, Graz University of Technology. It contains four classes of 4-s MI tasks: left hand, right hand, foot and tongue. The timing of experimental paradigm is shown in Fig. 1. In each trial, the subjects sat quietly for the first two seconds. Then, a "beep" was given, with a cross presented at the center of the screen, to indicate the beginning of the trial. A visual cue in form of an arrow pointing either to left, right, up or down (corresponding to left hand, right hand, both feet and tongue) was given at 3 s and stayed on the screen for one second. The subjects were required to imagine the corresponding movement during four seconds after the cue on-set. The cues indicating different motor imagery tasks were displayed in a randomized order.

The dataset was recorded from three subjects (denoted "k3", "k6", "l1") using 60 channels EEG. There are 180 training vs. 180 testing trials for Subject "k3", and 120 training vs. 120 testing trials for Subjects "k6" and "l1". The dataset providers have marked out the training trials contaminated with artifacts by the data property "HDR.ArtifactSelection". The numbers of training and testing trials for each subject are listed in Table 1, including the number of artifacts contaminated (AC) trials in the training data. Trials with

artifacts were marked out by the dataset provider. According to the data description provided in the BCI competition website (http://www.bbci.de/competition/iii/), the source derivation based on the center and the four nearest neighbor electrodes [59] was calculated for visually identifying the muscle and ocular artefacts. Artifacts in boundary electrodes were not considered. Our method used central electrodes only. Thus, the artifacts marked by the data provider are sufficient for testing our methods.

2.2. FDA-type F-score

FDA-type F-score is a simplified criterion based on Fisher discriminant analysis (FDA) for estimating the discriminative power of a group of features (a feature vector) [22]. We defined this measure first for two-class (e.g. Class "L" vs "R") cases:

$$F = \frac{\left\|\vec{\mu}^{L} - \vec{\mu}^{R}\right\|_2^2}{tr(\Sigma^{L}) + tr(\Sigma^{R})} \tag{1}$$

where $\vec{\mu}$ denotes the mean of the feature vector, $\|\cdot\|_2$ the $L2$-norm (Euclidean norm), Σ the covariance matrix of the feature vector and $tr(\cdot)$ the trace of a matrix. Thus, FDA-type F-score relies on the Euclidean distance between class centers to evaluate the difference between classes, and employs the trace of the covariance matrix to estimate the variance within one class. FDA-type F-score, as a simplified measure, avoids estimating a projection direction in multi-dimensional FDA, and has been successfully used in two-class BCI and motor recognition studies for channel and feature selection [21,22,33,42].

2.3. Subject-specific time-frequency optimization for multi-class EEG classification

In this study, we used three EEG channels, i.e. C3, Cz and C4, nearby the sensorimotor cortex. To improve the SNR of EEG signals, a small-distant Laplacian transformation [43] was applied as a spatial high-pass filter to each channel, in order to reduce the signal correlation and common noise caused by the head volume conduction. The EEG signals at C3, Cz and C4 were decomposed into a series of overlapping time-frequency areas $(\omega_m \times \tau_n)$, $m \in \{1, \ldots, M\}, n \in \{1, \ldots, N\}$, with successive frequency bands $\omega_m = [f_m, f_m + F - 1], f_{m+1} = f_m + F_s$ (F is the frequency bandwidth, F_s is the step in the frequency domain) and overlapping time intervals $\tau_n = [t_n, t_n + T - 1], t_{n+1} = t_n + T_s$ (T is the time interval width, T_s is the step in the time domain). The details of the parameter setting in this preprocessing step are provided in Section 2.4

We used FDA-type F-score to estimate the optimal time-frequency areas for extracting the subject-specific BP features. According to Eq. (1), in the optimal time-frequency area, the populations of features from different classes should have the largest FDA-type F-score distance. This concept can be extended to multi-class cases using the one-versus-rest (OVR) strategy, which is often used in multi-class classification [44]. Let us denote by $\mathcal{C}$ the set of all classes, and $\mathcal{C}\setminus\{O\}$ for all classes except class O. If we can consider $\mathcal{C}\setminus\{O\}$ as a large class, the multi-class problem can be transferred to a two-class problem, so as to compute the OVR-based F-score:

$$\hat{F} = \frac{\left\|\vec{\mu}^{O} - \vec{\mu}^{\mathcal{C}\setminus\{O\}}\right\|_2^2}{tr(\Sigma^{O}) + tr(\Sigma^{\mathcal{C}\setminus\{O\}})} \tag{2}$$

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

The feature vector, $[BP_{C3}, BP_{Cz}, BP_{C4}]$, contains the BP features (BP) from three Laplacian channels C3, Cz and C4, so the OVR-based F-score is calculated by:

$$\hat{F} = \frac{(\bar{BP}^{O}_{C3} - \bar{BP}^{\mathcal{C}\setminus\{O\}}_{C3})^2 + (\bar{BP}^{O}_{Cz} - \bar{BP}^{\mathcal{C}\setminus\{O\}}_{Cz})^2 + (\bar{BP}^{O}_{C4} - \bar{BP}^{\mathcal{C}\setminus\{O\}}_{C4})^2}{(\bar{S}^{O}_{C3} + \bar{S}^{O}_{Cz} + \bar{S}^{O}_{C4}) + (\bar{S}^{\mathcal{C}\setminus\{O\}}_{C3} + \bar{S}^{\mathcal{C}\setminus\{O\}}_{Cz} + \bar{S}^{\mathcal{C}\setminus\{O\}}_{C4})} \tag{3}$$

with:

$$\bar{BP} = \frac{1}{K}\sum_{k=1}^{K} BP(k) \tag{4}$$

$$\bar{S} = \frac{1}{K-1}\sum_{k=1}^{K} (BP(k) - \bar{BP})^2 \tag{5}$$

where K is the number of trials.

Before calculating the F-score, we applied the logarithm on BP features to make their distributions approximately normal [45]. In probability theory, a random vector is considered to be multivariate normally distributed if all linear combinations of its components obey univariate (one-dimensional) normal distributions [46]. Thus, the feature vector, $[BP_{C3}, BP_{Cz}, BP_{C4}]$, should be multivariate normally distributed. In practice, Mardia's test can be used to check whether a given dataset obeys the multivariate normal distribution with a given significance level of 0.05 [47].

For each class, we calculated $\hat{F}(\omega_m, \tau_n)$ at each time-frequency area $(\omega_m \times \tau_n)$ using Eq. (3), in order to measure its discriminative power for separating the class against all the others. The optimal time-frequency area (ω^*, τ^*) for separating each class was estimated by searching the maximum value of $\hat{F}(\omega_m, \tau_n)$ among all $M \times N$ time-frequency areas (M is the total number of frequency bands, N is the total amount of time intervals):

$$\hat{F}(\omega^*, \tau^*) = \max\left\{\hat{F}(\omega_m, \tau_n) \mid m \in \{1, 2\ldots, M\}, n \in \{1, 2\ldots, N\}\right\} \tag{6}$$

Without loss of generality, Fig. 2 presents the scheme of multi-class F-score based time-frequency selection (MFTFS) for a four-class problem. The optimal time-frequency area for separating one class from all the others can be considered as the characteristic time-frequency area for the class, since it contains information that makes the class different from all the others. Let us assume that the characteristic time-frequency areas for I different classes are $(\omega^*, \tau^*)_i, (i = 1, \ldots, I)$. BP features, $BP_{C3}, BP_{Cz}, BP_{C4}$, are extracted from the characteristic time-frequency area $(\omega^*, \tau^*)_i$ of each class i in Laplacian channels C3, Cz and C4, so called the class-relevant feature vector. Considering the selected characteristic time-frequency areas for some classes may be the same in practice, we only use the set of feature vectors from all different (ω^*, τ^*) for classification. This step can be achieved by:

$$\mathcal{A} = UNIQUE(\{(\omega^*, \tau^*)_i \mid i \in 1\ldots, I\}) \tag{7}$$

where $\mathcal{A}$ is the set of all different time-frequency areas (ω^*, τ^*), the operator $UNIQUE$ eliminates possible time-frequency area repetitions in the set of $\{(\omega^*, \tau^*)_i \mid i \in 1\ldots, I\}$. Then, the feature vector $[BP_{C3}, BP_{Cz}, BP_{C4}]$ is extracted from each time-frequency area in $\mathcal{A} = \{(\omega^*, \tau^*)_i \mid i \in 1\ldots, K, 1 \le K \le I\}$ for classification.

2.4. Data processing, method evaluation and comparison

Fig. 3 indicates how the EEG signal is decomposed into a series of overlapping time-frequency areas in each channel. This decomposition procedure is identical for different channels. For each Laplacian channel, 5th order Butterworth filters were employed to obtain 15

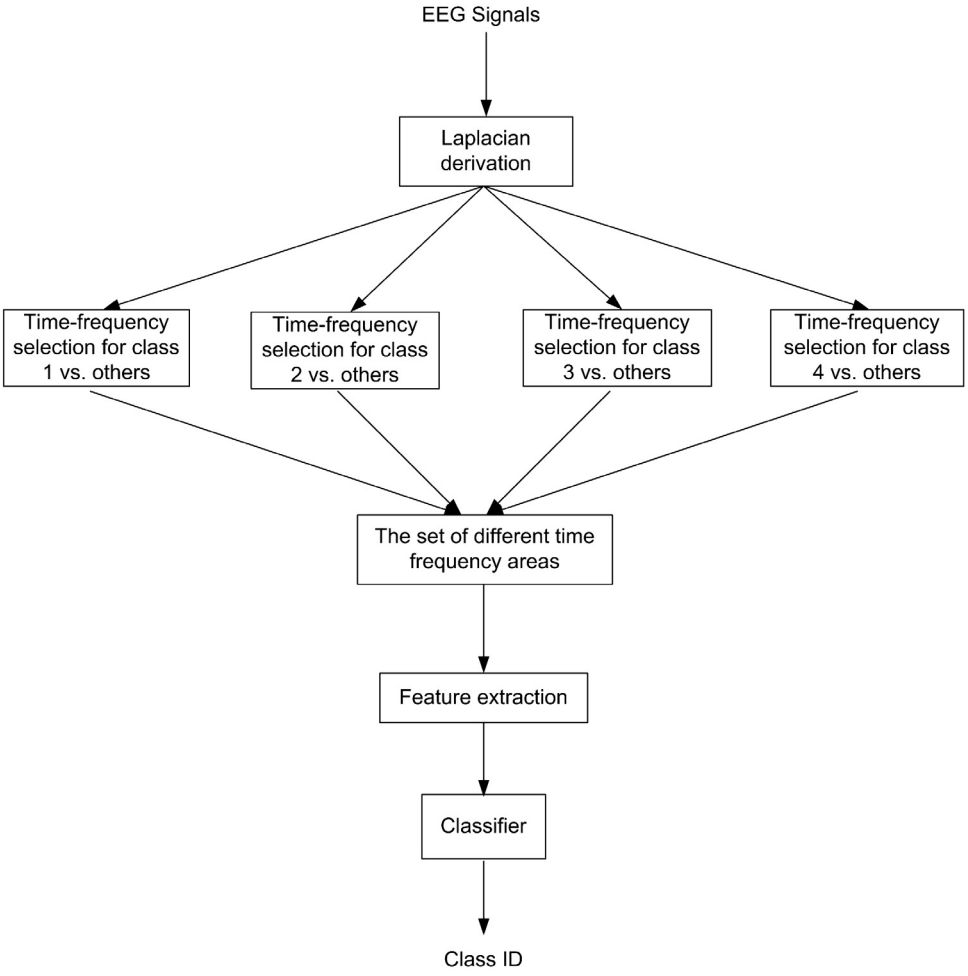

Fig. 2. Scheme of multi-class *F* score based time-frequency selection (MFTFS).

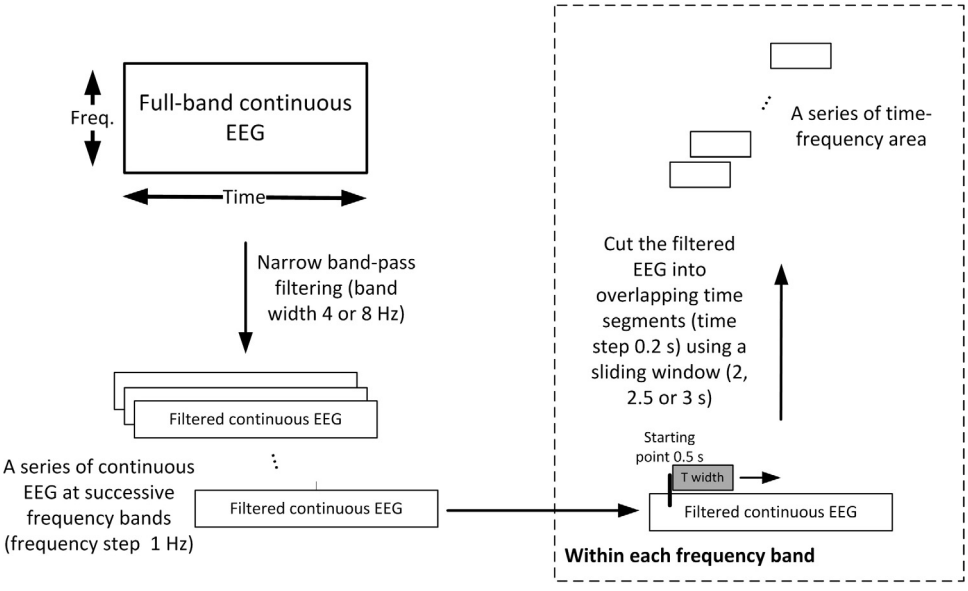

Fig. 3. Scheme of signal decomposition into a series of overlapping time-frequency areas.

successive 8 Hz-wide frequency bands (F = 8 Hz, F_s = 1 Hz): 8–16, 9–17, 10–18, . . ., 22–30 Hz, and 19 successive 4 Hz-wide frequency bands (F = 4 Hz, F_s = 1 Hz): 8–12, 9–13, 10–14 . . ., 26–30 Hz. Butterworth filter has flattest pass-band region among the commonly used filter (e.g. Chebyshev, Elliptic fliters), so it has the least attenuation over the desired frequency range. Thus, Butterworth filter is suitable for narrow band filtering. The order of Butterworth filter here was chosen to have a stable frequency response characteris-

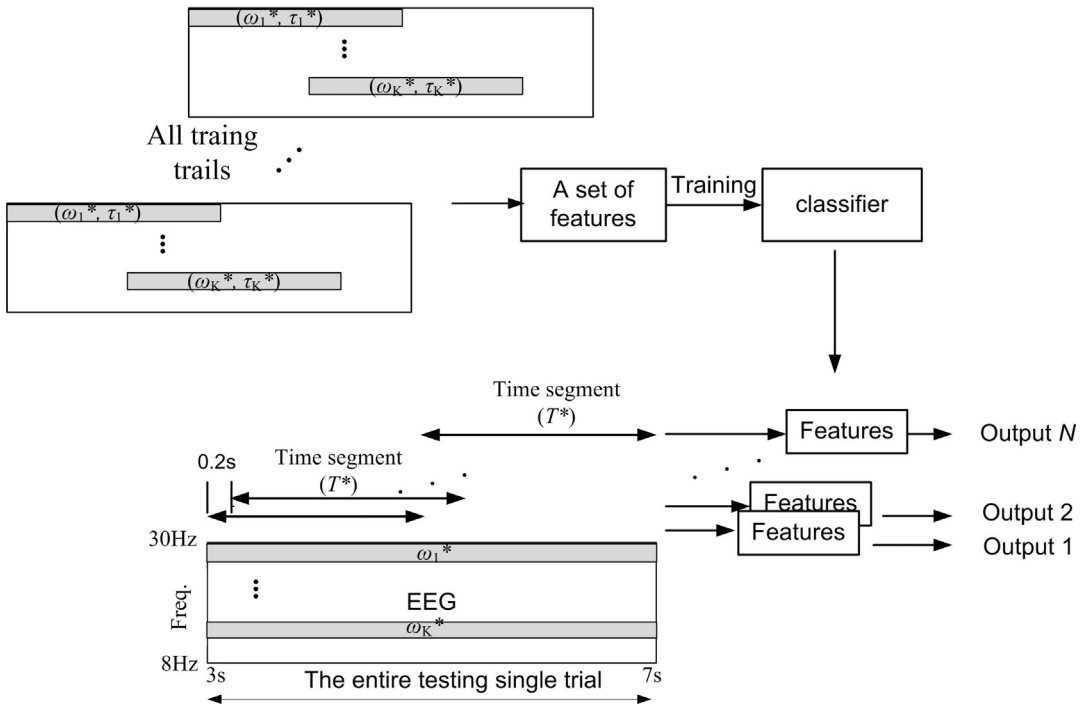

Fig. 4. Scheme of multi-class BCI based on subject-specific time-frequency features.

tic and a steep boundary. Shown in previous studies [29,48–50], the 5th order Butterworth filters work well with different feature extraction methods for various frequency bands including the frequency bands we used in this study.

Previous studies have shown that the event-related (de)synchronization typically occurs around 0.5 s after cue-on. Therefore, we used sliding windows in the period from 0.5 s after cue-on until the end of MI task to segment the band-pass filtered data. Different widths of time windows are used to find the optimal one. Specially, we used 2, 2.5 and 3 s wide (i.e. $T = 2$, 2.5 and 3 s, respectively) sliding windows (12 segments for each sliding window) with 0.2s-step (i.e. $T_s = 0.2$ s) moving from 0.5 s after the cue on-set until the end of MI task to get 36 overlapping time segments in each frequency band.

Thus, there were 34 frequency bands (i.e. fifteen 8 Hz-wide and nineteen 4 Hz-wide frequency bands) × 36 time segments (three different sizes of time windows and twelve segments for each size) = 1224 time-frequency areas in total. The F-score is calculated in each time-frequency area to find the optimal time-frequency area with maximum F-score using Eq. (6) for each class (i.e. one time-frequency area for one class).

When optimal time-frequency areas $\{(\omega^*, \tau^*)_i \mid i \in 1, \ldots, K, 1 \le K \le 4\}$ were selected for each subject, we extracted subject-specific BP features from the training trials to train the classifier. We used the multi-class Fisher's linear discriminant analysis (LDA) as the classifier, which was in line with the FDA-type F-score. The ω_i^*-bandpass ($i \in 1, \ldots, K, 1 \le K \le 4$) filtered EEG segments with the same time length as τ^* were generated from the whole single-trial of testing data via a 0.2-s step sliding window to obtain continuous classification results (see Fig. 4)

Firstly, we evaluated our method in the BCI competition III dataset IIIa [34] by comparing with other methods. To keep consistent with previous studies on this dataset, we used all training trials to seek for the optimal time-frequency areas and train the classifier, and then tested it with the independent testing trials. The results were compared with those obtained by data providers [34] and other recently developed methods applied to the same dataset, as well as using BP features extracted from the broad band (8–30 Hz) and full length (from the start to the end of motor imagery) EEG signals at the same three Laplacian channels. Since several recent studies used 10 fold cross-validations instead of independent testing trails to evaluate their methods, we also provided the results of 10 fold cross-validations. In the 10 fold cross-validations, the whole dataset is randomly split into 10 folds of equal size. At each iteration, 9 folds are used to train the classifier and test it with the rest one fold. We averaged over all folds to estimate the mean classification performance of cross-validation. Kappa coefficient and classification accuracy are two metrics commonly used for evaluating classification results. The relationship between kappa coefficient (κ) and classification accuracy (Acc) is $\kappa = (Acc - ch)/(1 - ch)$, where ch is the chance level for classification ($ch = 0.25$ for four-class classification). Kappa coefficient is generally thought to be a more robust measure than classification accuracy, since Kappa coefficient takes into account the possibility of the agreement occurring by chance. Thus, kappa coefficient is recommended in the BCI competitions [36]. However, many BCI studies provide classification accuracy as well, since classification accuracy directly indicates the agreement between classification results and truth labels. Here we provide results in both kappa coefficient and classification accuracy to compare our results with those obtained by other methods. This part of results and corresponding discussions are provided in Section 3.1

Secondly, we tested our method with different amounts of artifacts contaminated (AC) trials to evaluate its robustness to artifacts. We started with using only "clean" trials for time-frequency selection and classifier training, and test the classifier with all testing trials. Then, we ran several tests by randomly picking up certain amount of AC trials to add to training data to repeat the evaluation. In each test, we repeated 20 times for the same amount of AC trials to estimate the mean classification performance and its standard deviation. For each new test, we added more AC trials (the amount equal to 5% of total training data, so $180 \times 5\% = 9$ trials for k3 and $120 \times 5\% = 6$ trails for k6 and l1) than previous one until all AC trials were added, so as to get the curve of classification performance

Table 2
Comparisons between the MFTFS and other methods in BCI competition III using independent testing datasets. 'NC' is the number of channels used for feature extraction and classification. The best performances are highlighted in bold.

	Kappa coefficient					Classification accuracy					*NC*
	k3	k6	l1	Mean	Std.	k3	k6	l1	Mean	Std.	
MFTFS (our method)	0.64	**0.71**	**0.72**	**0.69**	**0.04**	0.73	**0.78**	**0.79**	**0.77**	**0.03**	**3**
Broad band BP	0.14	0.01	0.05	0.07	0.07	0.36	0.26	0.29	0.30	0.04	**3**
AAR [34]	0.70	0.37	0.39	0.49	0.19	0.78	0.53	0.54	0.62	0.11	60
ICA + PCA [36]	**0.95**	0.41	0.52	0.63	0.29	**0.96**	0.58	0.64	0.72	0.17	60
CSP [36]	0.90	0.43	0.71	0.68	0.23	0.93	0.58	0.78	0.76	0.14	60
JAD-CSP [37]	0.76	0.41	0.53	0.57	0.18	0.82	0.56	0.65	0.68	0.11	60

Table 3
Comparisons between the MFTFS and the state-of-the-art methods (published after 2012) using 10-fold cross-validation. 'NC' is the (average) number of channels used for feature extraction and classification. The best performances are highlighted in bold.

	Kappa coefficient					Classification accuracy					*NC*
	k3	k6	l1	Mean	Std.	k3	k6	l1	Mean	Std.	
MFTFS (our method)	0.71	**0.77**	**0.85**	**0.78**	**0.04**	0.79	**0.84**	**0.89**	**0.84**	**0.03**	**3**
STFSCSP [39]	0.84	0.66	0.82	0.77	0.08	0.88	0.75	0.86	0.83	0.06	16
ARR [40]	0.69	0.36	0.39	0.48	0.15	0.77	0.52	0.54	0.61	0.11	20
DCT [40]	0.17	0.29	0.25	0.24	0.06	0.38	0.47	0.44	0.43	0.04	20
KL-CSP [38]	0.78	0.49	0.56	0.61	0.15	0.84	0.62	0.67	0.71	0.09	60
KL-LTCSP [38]	0.79	0.50	0.54	0.61	0.16	0.84	0.63	0.66	0.71	0.10	60
LP-SVD + logistic model tree [51]	**0.87**	0.68	0.70	0.75	0.08	**0.90**	0.76	0.78	0.81	0.06	20
LP-SVD + AR + error variance [40,41]	0.69	0.56	0.45	0.57	0.12	0.77	0.67	0.59	0.68	0.07	20

over different amounts of AC trials. This part of results and corresponding discussions are provided in Section 3.2. The results are provided in kappa coefficient only.

3. Results and discussion

3.1. Improving classification performance based on few EEG channels

The classification performances obtained by using our method MFTFS and other methods are provided in Tables 2 and 3. The performance is evaluated with kappa coefficient (κ) and classification accuracy (Acc). Higher kappa coefficient and classification accuracy indicate better classification performance. Table 2 shows the results using independent testing dataset as BCI competition required. Table 3 provides the results of 10-fold cross-validation using all the available data.

Using BP features directly extracted from the broad band (8–30 Hz) and full length (from the start to the end of motor imagery) EEG signals at the three Laplacian channels (C3, Cz and C4) yields the poorest classification performance among all methods, which is just above the chance level (zero kappa). We also provide receiver operating characteristic (ROC) curves of four classes for our method MFTFS in comparison to using the broad band, full length data (Fig. 5). From the ROC curves, we can see that using our method MFTFS to extract BP features from the optimal time-frequency areas greatly improves classification performances for all classes. Thus, it is necessary to seek for the optimal time-frequency areas for BP feature extraction when using only three Laplacian channels for multi-class BCI.

We also compared our method with the methods applied to the same dataset in BCI competition including adaptive autoregressive (AAR) used by data providers [34],[1] independent common analysis (ICA), principal component analysis (PCA) and common spatial pattern (CSP) [36] and a previous study using joint approximate diagonalization based CSP (JAD-CSP) [37]. The results are summarized in Table 2. All these results were obtained on the dependent testing dataset.

Furthermore, we compared our method with the methods that were proposed in the last five years from 2012 to 2016, including sparse time-frequency segment CSP (STFSCSP) [39], Kullback-Leibler (KL) divergence based CSP (KL-CSP) [38], KL divergence based local temporal common spatial patterns (KL-LTCSP) [38], discrete cosine transform (DCT) [40] and linear prediction singular value decomposition (LP-SVD) [40,41,51]. Since all these recent works used 10-fold cross-validation to test their methods, we provide our results in 10-fold cross-validation as well to compare with them in Table 3.

Our method MFTFS yielded the best mean classification accuracy (0.77 for independent testing data, 0.84 for 10-fold cross validation) and kappa coefficient (0.69 for independent testing data and 0.78 for 10-fold cross validation) with smallest standard deviation among all methods in both comparisons (see Tables 2 and 3). Moreover, our method used only three Laplacian channels, which are far less than other methods in comparison. These results indicate our method is very effective in improving classification performance based on few EEG channels.

Motor imagery BCI classifies subject's motor intentions based on the features extracted from EEG signals. The underlying neurophysiological mechanism is that motor imagery of a specific body part induces EEG power changes (known as event-related desynchronization and synchronization) in the specific frequency band of EEG oscillation (i.e. sensorimotor rhythm) over the sensorimotor cortex [35]. Thus, we used band power features to identify the task-relevant band power change in the sensorimotor rhythm for classifying the motor imagery tasks. We noted that several recent studies used other features for classification in the same dataset. In comparison with methods using other features, i.e. AAR, ICA, DCT features (see Tables 2 and 3), we cannot conclude that using other features could be superior to using band power features; while more computational cost may be required for calculating other features, such as ICA features.

For using band power features, it is essential to extract the task-relevant band power change in the sensorimotor rhythm. As

[1] Using AAR as the feature extraction method, Schlögl and colleagues provided classification results obtained by four different classifiers, namely minimum distance analysis, linear discriminant analysis, k-nearest-neighbor and support vector machine [34]. Here we list their best results for comparison.

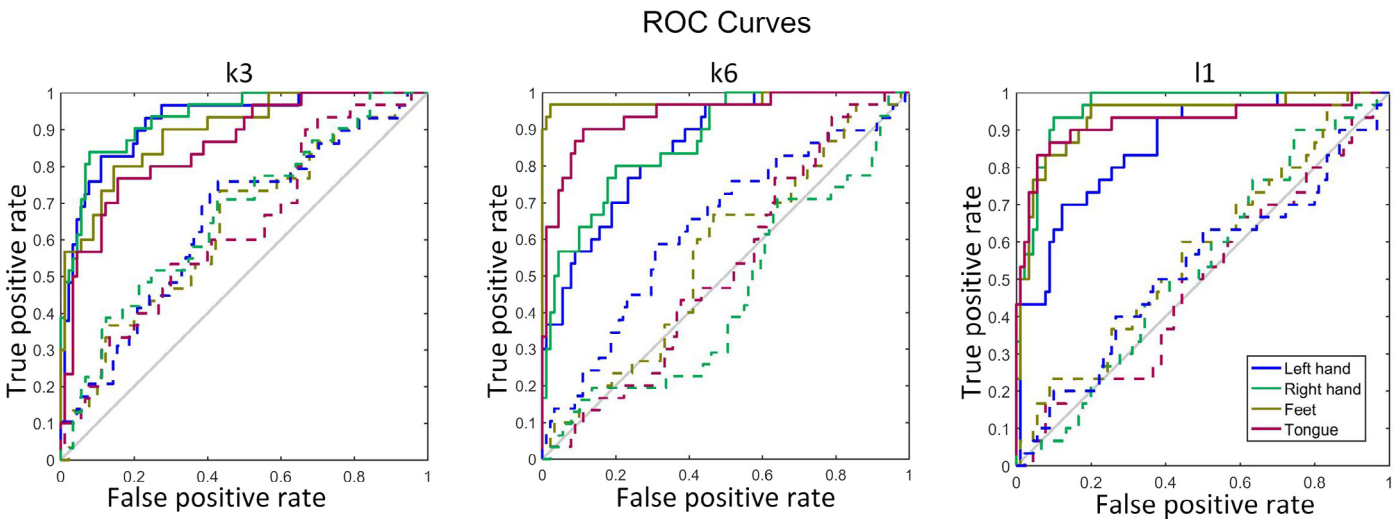

Fig. 5. ROC curves for our method (MFTFS, solid lines) in comparison to using the broad band, full length data (dashed lines).

indicated in Table 5, the exact frequency band of sensorimotor rhythm can varies from subject to subject. Individual differences in the frequency band of sensorimotor rhythm will deteriorate BCI performances when using a general parameter setting for all subjects (e.g. features from the broad frequency band 8–30 Hz), as shown in Fig. 5 and Table 2. Additionally, the delay between the cue onset and the starting of motor imagery varies with individuals [21]. Previous studies have demonstrated that estimating optimal time-frequency areas for each individual can greatly improve the classification performance [52,29]. Due to the volume conduction effect, spatial filtering is also needed to reduce the correlation of EEG signals and eliminate the artefacts [18]. Data-driven spatial filtering methods, such as common spatial pattern (CSP), usually need a large number of EEG channels. Alternatively, the bipolar or Laplacian derivation can be used when there are only a few electrodes placed around the sensorimotor cortex. This option has been successfully used in the previous studies for two-class cases, generating even better results than CSP [53,54,29]. The central channels C3, Cz and C4 are close to functional regions of the sensorimotor cortex for motor imageries of left vs. right hand, feet and tongue [10]. Moreover, the central electrodes are relatively free of muscle artefact compared to the electrodes in the boundary of EEG cap [55]. Our method selects the optimal time-frequency areas to extract band power features based on the Laplacian channels C3, Cz and C4, which combines the knowledge discussed above to improve classification results based on few channels. As a result, our method can yield better performances in comparison to other methods.

Noteworthy, our time-frequency selection method is not expensive in computational time (2.5 min in average for finding the characteristic time-frequency area for each class, so around 10 min in total for the four-class problem, using Matlab 7.10.0 on Windows 7 Professional 64bit, RAM 2.0G, CPU 2.66G Hz). For on-line BCI, this step could be performed off-line before on-line classification.

Table 4
Evaluation of the MFTFS sensitivity (evaluated with *kappa coefficient*) to artifacts on BCI competition III dataset IIIa.

	Subjects			
	k3	k6	l1	Mean
All	0.64	**0.71**	**0.72**	**0.69**
Clean	**0.69**	0.70	0.60	0.66

3.2. Robustness to artifacts

Table 4 lists the classification results (evaluated by *kappa coefficient* [56], κ) of using all training trials (All) v.s. using only the "clean" trials. We examined the results for each subject individually. For Subject *k*3, using only the clean trials ($\kappa = 0.69$) yields a better classification result than using all training trials ($\kappa = 0.63$). However, for Subjects *k*6 and *l*1, the classification results obtained by using only the clean trials (Subject *k*6: $\kappa = 0.70$, Subject *l*1: $\kappa = 0.60$) are worse than those obtained by using all training trials (Subject *k*6: $\kappa = 0.71$, Subject *l*1: $\kappa = 0.72$).

Generally speaking, artifacts should theoretically deteriorate the classification results, since artifacts reduce the signal-to-noise ratio. However, removing AC trials will definitely decrease the number of trials for training, where a limited amount of data might not be sufficient to train the classifier. Therefore, a trade-off should be made between rejecting AC trials and maintaining the amount

Table 5
Characteristic time-frequency areas for different classes on BCI competition III dataset IIIa, obtained by using all training trials (All) and only the "clean" trials, respectively.

Subject			Left hand	Right hand	Feet	Tongue
k3	All	Freq.(Hz)	12–20	13–21	13–21	12–20
		Time(s)	0.9–3.9	1.9–4.9	2.1–5.1	0.7–3.7
	Clean	Freq.(Hz)	12–20	12–20	13–21	12–20
		Time(s)	2.5–5.0	2.1–5.1	2.1–5.1	0.7–3.7
k6	All	Freq.(Hz)	22–30	26–30	22–30	19–27
		Time(s)	0.5–3.5	0.9–3.9	2.7–5.7	2.7–5.7
	Clean	Freq.(Hz)	22–30	26–30	22–30	18–26
		Time(s)	0.5–3.5	0.9–3.9	2.3–5.3	2.7–5.7
l1	All	Freq.(Hz)	20–28	12–20	22–30	12–20
		Time(s)	2.7–5.7	1.1–4.1	2.5–5.5	1.1–4.1
	Clean	Freq.(Hz)	22–30	13–17	22–30	12–20
		Time(s)	0.7–3.7	1.8–4.8	2.7–5.7	2.7–5.7

of training data. This trade-off, in some extent, also depends on the sensitivity of the method to noise. If the method is very sensitive to noise, then adding AC trials to the training dataset can only deteriorate the classification result. As a result, using only the clean trials will yield better results than adding AC trials. On the contrary, if a method is robust to artifacts, then adding AC trials may not decrease the classification results any more.

In Table 4, we can see that using all training trials ($\bar{\kappa} = 0.69$) generates a better mean classification result than using only clean trials ($\bar{\kappa} = 0.64$), indicating that MFTFS is robust to artifacts. In this dataset, subject *l*1 has the least clean training trials (84 trials, see Table 1). If only using clean trials, the classification result for *l*1 ($\kappa = 0.60$) is the worst among all the subjects.

Table 5 lists the characteristic time-frequency areas for different classes, obtained by using all training trials (All) and only the clean trials, respectively. From Table 5, we can see that the characteristic time-frequency areas shift for some classes for each subject between using all training trials and only the clean trials. Thus, MFTFS can adapt to the amount of training data and noise to find the optimal time-frequency areas for extracting the most discriminative features.

To further study the effect of adding AC trials on the performance of MFTFS, we add different amounts of AC data to training data, so as to get a curve of classification performance with respect to different numbers of training data with different amounts of noise. In this test, we randomly picked out certain amount of AC trials adding to training data and repeated the evaluation 20 times. Note that the repeated evaluation is not for the case using only "clean" trials (no AC trials at all) and the case using all trials (including all AC trials). Therefore, there are no error bars for these two cases. From Fig. 6, we can see that adding AC trials to the training dataset does not necessarily deteriorate classification performances, indicating that our method is tolerant to artifacts. For Subjects "k6" and "l1", who have the limited numbers of "clean" trials (92 trials for "k6" and 84 trials for "l1"), adding AC trials can generally improve the classification performance until the training data are more than hundred trials. The best performance ($\kappa = 0.77$) is achieved with 116 training trials (AC trials/training trials = 24/92) for Subject "k6" and 114 training trials (AC trials/training trials = 30/84) for Subject "l1". However, for Subject "k3", adding AC trials cannot improve the classification results. The best performance is achieved when using only clean trials for this subject. The reason is that this subject already has a large number of training trials (149 "clean" trials, see Table 1). Usually, a training dataset should have five to ten times as many trials as the dimensionality of features to obtain a good performance of a classifier [57,58]. Increasing the number of trials in this range usually will improve the classification performance. For a four-class problem, the dimensionality of MFTFS selected BP features are around $3 \times 4 = 12$, so that 120 trials is the upper boundary in this range, which is closed to the reported optimal number of trials for Subjects "k6" and "l1". For Subject "k3", the amount of "clean" trials (149 trials) are already above this boundary. In this case, adding AC trials deteriorates the classification performance, since it mainly introduces artifacts rather than providing more useful information for training.

To sum up, using MFTFS, when the number of clean training trials is less than ten times of the feature dimension, it is possible to improve classification results by adding AC trials. On the contrary, if the number of clean training trials is large enough, adding AC trials is not helpful any more.

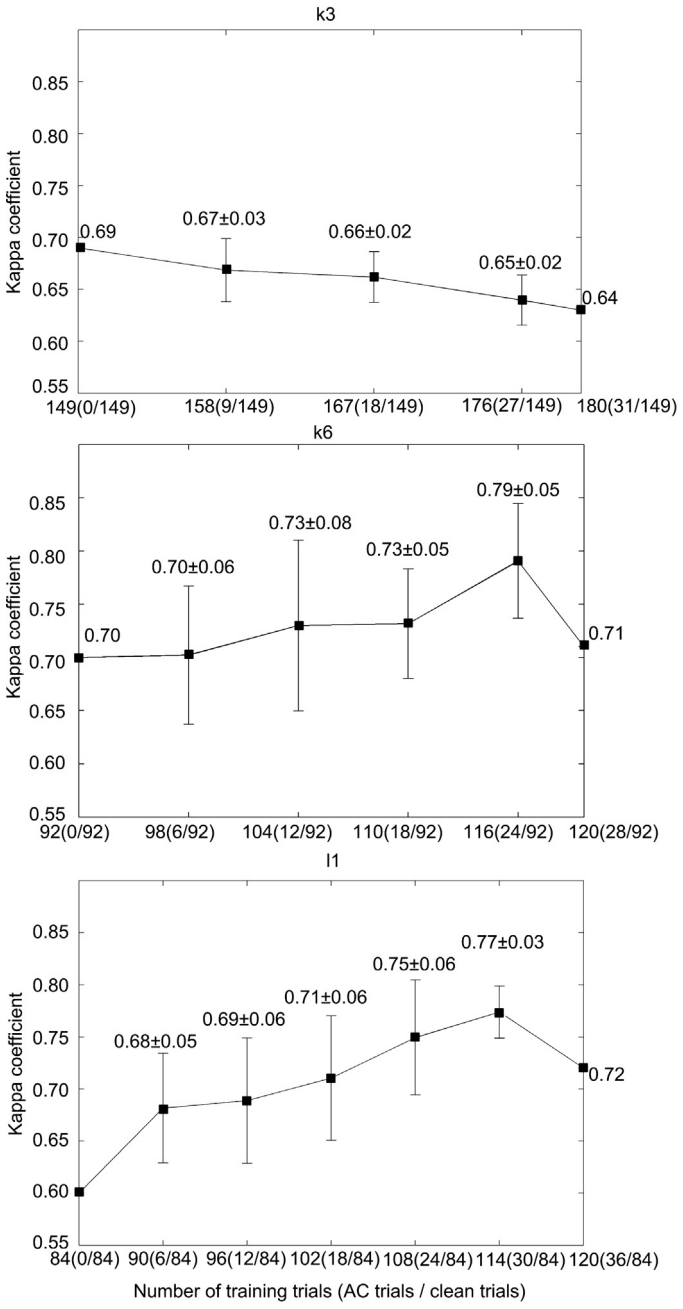

Fig. 6. Curve of classification performance (measured by κ values) with respect to different numbers of training data with different amounts of AC trials for different subjects. Results are averaged over 20 repetitions. We randomly pick out certain amount of AC trials and added to training data for each repetition. Error bars show standard deviations of κ values. We also provided the numerical values (mean ± std.).

4. Conclusion

Previous studies have demonstrated the importance of feature extraction and using few EEG channels to the widespread use of BCI. The optimal frequency band and time segment to extract motor-imagery features differ from subject to subject. Extracting the feature from a uniform time-frequency area cannot guarantee the best BCI performance for all subjects and therefore reduces the inter-subject robustness. To address this challenge, we proposed an effective method to seek optimal time-frequency areas for extracting subject-specific features, in order to improve the classification performance using few channels EEG. Different from our previous work, this study focused on the multi-class motor imagery BCI, which aims to provide the user with more freedom in the control. The method was evaluated on a standard multi-class BCI dataset (BCI competition III dataset IIIa). The experimental results show

that our method yields better classification performance than other methods, using fewer EEG channels. Moreover, the performance of our method is stable over subjects and robust to artifacts. This work could contribute to daily use multi-class BCIs based on few EEG channels. This BCI based on few channels can be used for patients with lock-in syndrome or spinal injury, whose cortical structure and function are similar to healthy subjects. Nevertheless, it could be a challenge for patients with brain injury such as a stroke, when brain functions have been affected by the lesion and the post-injury brain reorganization. The brain activity recorded from C3, Cz and C4 (around the sensorimotor cortex) may be too weak to perform any MI-based control. A few groups are aiming to solve this problem, e.g. using invasive electrocorticogram recording [11]. In the future, we will combine our proposed method with a subject-specific channel selection algorithm [21] to provide a non-invasive solution for non-severe chronic stroke patients.

Ethical statements

We validated our method by an open-access BCI dataset, i.e. the dataset IIIa [34] from BCI competition III (http://www.bbci.de/competition/iii/). Thus, our study did not involve any self-recorded datasets from human participants or animals. The authors declare that they have no conflict of interest.

Acknowledgements

Authors thank Dr. Teodoro Solis-Escalante and Dr. Olexiy Kyrgyzov for useful discussions about the method, and Dr. Alois Schlögl for providing the BCI dataset. Authors also thank Prof. Gert Kwakkel and Prof. Julius Dewald for useful discussions about BCI in stroke. This work was partially supported by funding from Orange Labs, France Télécom. Since November 2013, Dr. Yuan Yang has been funded by the European Research Council under the European Union's Seventh Framework Programme (FP/2007–2013)/ERC Grant Agreement No. 291339 (4D-EEG project) obtained by Prof. Dr. Frans C.T. van der Helm and Prof. Dr. Gert Kwakkel in 2012.

References

[1] A. Vallabhaneni, T. Wang, B. He, Brain–computer interface, Neural Eng. (2005) 85–121.
[2] E.A. Curran, M.J. Stokes, Learning to control brain activity: a review of the production and control of EEG components for driving brain–computer interface (BCI) systems, Brain Cogn. 51 (2003) 326–336.
[3] L.A. Farwell, E. Donchin, Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials, Electroencephalogr. Clin. Neurophysiol. 70 (1988) 510–523.
[4] J. Jin, B.Z. Allison, E.W. Sellers, C. Brunner, P. Horki, X. Wang, C. Neuper, An adaptive p300-based control system, J. Neural Eng. 8 (2011) 036006.
[5] J.J. Vidal, Toward direct brain–computer communication, Annu. Rev. Biophys. Bioeng. 2 (1973) 157–180.
[6] X. Chen, Y. Wang, M. Nakanishi, X. Gao, T.-P. Jung, S. Gao, High-speed spelling with a noninvasive brain–computer interface, Proc. Natl. Acad. Sci. U. S. A. 112 (2015) E6058–E6067.
[7] Z. Qiu, B.Z. Allison, J. Jin, Y. Zhang, X. Wang, W. Li, A. Cichocki, Optimized motor imagery paradigm based on imagining Chinese characters writing movement, IEEE Trans. Neural Syst. Rehabil. Eng. (2017).
[8] G. Pfurtscheller, B.Z. Allison, C. Brunner, G. Bauernfeind, T. Solis-Escalante, R. Scherer, T.O. Zander, G. Mueller-Putz, C. Neuper, N. Birbaumer, The hybrid BCI, Front. Neurosci. 2 (2010) 1–12.
[9] M. Wang, I. Daly, B.Z. Allison, J. Jin, Y. Zhang, L. Chen, X. Wang, A new hybrid BCI paradigm based on p300 and SSVEP, J. Neurosci. Methods 244 (2015) 16–25.
[10] G. Pfurtscheller, C. Brunner, A. Schlögl, F.H. Lopes da Silva, Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks, Neuroimage 31 (2006) 153–159.
[11] S. Silvoni, A. Ramos-Murguialday, M. Cavinato, C. Volpato, G. Cisotto, A. Turolla, F. Piccione, N. Birbaumer, Brain–computer interface in stroke: a review of progress, Clin. EEG Neurosci. 42 (2011) 245–252.
[12] R. Swaminathan, S. Prasad, Brain computer interface used in health care technologies, in: Next Generation DNA Led Technologies, Springer, 2016, pp. 49–58.
[13] A. Kübler, F. Nijboer, J. Mellinger, T.M. Vaughan, H. Pawelzik, G. Schalk, D.J. McFarland, N. Birbaumer, J.R. Wolpaw, Patients with ALS can use sensorimotor rhythms to operate a brain–computer interface, Neurology 64 (2005) 1775–1777.
[14] R. Scherer, A. Schlögl, F. Lee, H. Bischof, J. Janša, G. Pfurtscheller, The self-paced Graz brain–computer interface: methods and applications, Comput. Intell. Neurosci. 2007 (2007) 79826.
[15] Y. Yang, J. Wiart, I. Bloch, Towards next generation human-computer interaction-brain–computer interfaces: applications and challenges, 1st International Symposium of Chinese CHI (Chinese CHI 2013) (2013) 1–2.
[16] D virtual reality environments.
[17] S. Liang, K.-S. Choi, J. Qin, W.-M. Pang, Q. Wang, P.-A. Heng, Improving the discrimination of hand motor imagery via virtual reality based visual guidance, Comput. Methods Programs Biomed. 132 (2016) 63–74.
[18] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, K.R. Müller, Optimizing spatial filters for robust EEG single-trial analysis, IEEE Signal Process. Mag. 25 (2008) 41–56.
[19] Y. Yang, S. Chevallier, J. Wiart, I. Bloch, Automatic selection of the number of spatial filters for motor-imagery BCI, 20th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2012) (2012) 109–114.
[20] S. Wang, C.J. James, Extracting rhythmic brain activity for brain–computer interfacing through constrained independent component analysis, Comput. Intell. Neurosci. 2007 (2007) 41468.
[21] Y. Yang, I. Bloch, S. Chevallier, J. Wiart, Subject-specific channel selection using time information for motor imagery brain–computer interfaces, Cogn. Comput. 8 (2016) 505–518.
[22] Y. Yang, O. Kyrgyzov, J. Wiart, I. Bloch, Subject-specific channel selection for classification of motor imagery electroencephalographic data, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2013) (2013) 1277–1280.
[23] L. He, Y. Hu, Y. Li, D. Li, Channel selection by Rayleigh coefficient maximization based genetic algorithm for classifying single-trial motor imagery EEG, Neurocomputing 121 (2013) 423–433.
[24] M. Arvaneh, C. Guan, K.K. Ang, C. Quek, Optimizing the channel selection and classification accuracy in EEG-based BCI, IEEE Trans. Biomed. Eng. 58 (2011) 1865–1873.
[25] O. Kyrgyzov, I. Bloch, Y. Yang, J. Wiart, A. Souloumiac, Data ranking and clustering via normalized graph cut based on asymmetric affinity, in: Image Analysis and Processing-ICIAP 2013, Springer, 2013, pp. 562–571.
[26] H. Shan, H. Xu, S. Zhu, B. He, A novel channel selection method for optimal classification in different motor imagery BCI paradigms, Biomed. Eng. Online 14 (2015) 93.
[27] J. Wang, F. Xue, H. Li, Simultaneous channel and feature selection of fused EEG features based on Sparse Group Lasso, BioMed Res. Int. 2015 (2015) 703768.
[28] A. Barachant, S. Bonnet, Channel selection procedure using Riemannian distance for BCI applications, 5th International IEEE/EMBS Conference on Neural Engineering (NER 2011) (2011) 348–351.
[29] Y. Yang, S. Chevallier, J. Wiart, I. Bloch, 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) 38.
[30] Y. Yang, S. Chevallier, J. Wiart, I. Bloch, Time-frequency selection in two bipolar channels for improving the classification of motor imagery EEG, 34th IEEE Annual International Conference of Engineering in Medicine and Biology Society (EMBC 2012) (2012) 2744–2747.
[31] B. Yang, H. Li, Q. Wang, Y. Zhang, Subject-based feature extraction by using fisher WPD-CSP in brain–computer interfaces, Comput. Methods Programs Biomed. 129 (2016) 21–28.
[32] J. Luo, Z. Feng, J. Zhang, N. Lu, Dynamic frequency feature selection based approach for classification of motor imageries, Comput. Biol. Med. 75 (2016) 45–53.
[33] C. Ansuini, A. Cavallo, A. Koul, M. Jacono, Y. Yang, C. Becchio, Predicting object size from hand kinematics: a temporal perspective, PLoS ONE 10 (2015) e0120432.
[34] A. Schlögl, F. Lee, H. Bischof, G. Pfurtscheller, Characterization of four-class motor imagery EEG data for the BCI-competition 2005, J. Neural Eng. 2 (2005) L14.
[35] G. Pfurtscheller, F.H. Lopes da Silva, Event-related EEG/MEG synchronization and desynchronization: basic principles, Clin. Neurophysiol. 110 (1999) 1842–1857.
[36] B. Blankertz, K.R. Müller, D.J. Krusienski, G. Schalk, J.R. Wolpaw, A. Schlögl, G. Pfurtscheller, J.R. Millán, M. Schroder, N. Birbaumer, The BCI competition III: validating alternative approaches to actual BCI problems, IEEE Trans. Neural Syst. Rehabil. Eng. 14 (2006) 153–159.
[37] M. Grosse-Wentrup, M. Buss, Multiclass common spatial patterns and information theoretic feature extraction, IEEE Trans. Biomed. Eng. 55 (2008) 1991–2000.
[38] H. Wang, Harmonic mean of Kullback-Leibler divergences for optimizing multi-class EEG spatio-temporal filters, Neural Process. Lett. 36 (2012) 161–171.
[39] M. Miao, H. Zeng, A. Wang, C. Zhao, F. Liu, Discriminative spatial-frequency-temporal feature extraction and classification of motor imagery EEG: an sparse regression and weighted Naïve Bayesian classifier-based approach, J. Neurosci. Methods 278 (2017) 13–24.

[40] M. Mesbah, A. Khorshidtalab, H. Baali, A. Al-Ani, Motor imagery task classification using a signal-dependent orthogonal transform based feature extraction Neural Information Processing, vol. 9490, Springer, 2015, pp. 1–9.
[41] H. Baali, A. Khorshidtalab, M. Mesbah, M.J. Salami, A transform-based feature extraction approach for motor imagery tasks classification, IEEE J. Transl. Eng. Health Med. 3 (2015) 1–8.
[42] C. Ansuini, A. Cavallo, A. Koul, et al., Grasping others' movements: rapid discrimination of object size from observed hand movements, J. Exp. Psychol. Hum. Percept. Perform. 42 (7) (2016) 918.
[43] J. Deng, J. Yao, J.P. Dewald, Classification of the intention to generate a shoulder versus elbow torque by means of a time-frequency synthesized spatial patterns BCI algorithm, J. Neural Eng. 2 (2005) 131.
[44] D.M.J. Tax, R.P.W. Duin, Using two-class classifiers for multiclass classification International Conference on Pattern Recognition, vol. 2, IEEE, 2002, pp. 124–127.
[45] C. Vidaurre, N. Kramer, B. Blankertz, A. Schlögl, Time domain parameters as a feature for EEG-based brain–computer interfaces, Neural Netw. 22 (2009) 1313–1319.
[46] R.A. Johnson, D.W. Wichern, Applied Multivariate Statistical Analysis, Prentice Hall, Englewood Cliffs, NJ, 1992.
[47] K.V. Mardia, Measures of multivariate skewness and kurtosis with applications, Biometrika 57 (1970) 519–530.
[48] H.-I. Suk, S.-W. Lee, A novel Bayesian framework for discriminative feature extraction in brain–computer interfaces, IEEE Trans. Pattern Anal. Mach. Intell. 35 (2013) 286–299.
[49] F. Lotte, C. Guan, Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms, IEEE Trans. Biomed. Eng. 58 (2011) 355–362.
[50] C. Park, D. Looney, N. ur Rehman, A. Ahrabian, D.P. Mandic, Classification of motor imagery BCI using multivariate empirical mode decomposition, IEEE Trans. Neural Syst. Rehabil. Eng. 21 (2013) 10–22.
[51] A. Khorshidtalab, M.J. Salami, A. Rini, Motor imagery task classification using transformation based features, Biomed. Signal Process. Control 33 (2017) 213–219.
[52] K.K. Ang, Z.Y. Chin, C. Wang, C. Guan, H. Zhang, Filter bank common spatial pattern algorithm on BCI Competition IV datasets 2a and 2b, Front. Neurosci. 6 (2012) 39.
[53] B. Lou, B. Hong, X. Gao, S. Gao, Bipolar electrode selection for a motor imagery based brain–computer interface, J. Neural Eng. 5 (2008) 342–349.
[54] T. Solis-Escalante, G. Müller-Putz, G. Pfurtscheller, Overt foot movement detection in one single Laplacian EEG derivation, J. Neurosci. Methods 175 (2008) 148–153.
[55] S. Fitzgibbon, D. DeLosAngeles, T. Lewis, D. Powers, E. Whitham, J. Willoughby, K. Pope, Surface Laplacian of scalp electrical signals and independent component analysis resolve EMG contamination of electroencephalogram, Int. J. Psychophysiol. 97 (2015) 277–284.
[56] A. Schlögl, J. Kronegg, J.E. Huggins, S.G. Mason, Toward Brain-Computer Interfacing, MIT Press, 2007, pp. 327–360.
[57] A. Jain, D. Zongker, Feature selection: evaluation, application, and small sample performance, IEEE Trans. Pattern Anal. Mach. Intell. 19 (1997) 153–158.
[58] S.J. Raudys, A.K. Jain, Small sample size effects in statistical pattern recognition: recommendations for practitioners, IEEE Trans. Pattern Anal. Mach. Intell. 13 (1991) 252–264.
[59] B. Hjorth, An on-line transformation of EEG scalp potentials into orthogonal source derivations, Electroencephalogr. Clin. Neurophysiol. 39 (5) (1975) 526–530.