

RESEARCH

Open Access

Time-frequency optimization for discrimination between imagination of right and left hand movements based on two bipolar electroencephalography channels

Yuan Yang^{1,4}, Sylvain Chevallier², Joe Wiart^{3,4} and Isabelle Bloch^{1,4*}

Abstract

To enforce a widespread use of efficient and easy to use brain-computer interfaces (BCIs), the inter-subject robustness should be increased and the number of electrodes should be reduced. These two key issues are addressed in this contribution, proposing a novel method to identify subject-specific time-frequency characteristics with a minimal number of electrodes. In this method, two alternative criteria, time-frequency discrimination factor (TFDF) and F score, are proposed to evaluate the discriminative power of time-frequency regions. Distinct from classical measures (e.g., Fisher criterion, r^2 coefficient), the TFDF is based on the neurophysiologic phenomena, on which the motor imagery BCI paradigm relies, rather than only from statistics. F score is based on the popular Fisher's discriminant and purely data driven; however, it differs from traditional measures since it provides a simple and effective measure for quantifying the discriminative power of a multi-dimensional feature vector. The proposed method is tested on BCI competition IV datasets IIa and IIb for discriminating right and left hand motor imagery. Compared to state-of-the-art methods, our method based on both criteria led to comparable or even better classification results, while using fewer electrodes (i.e., only two bipolar channels, C3 and C4). This work indicates that time-frequency optimization can not only improve the classification performance but also contribute to reducing the number of electrodes required in motor imagery BCIs.

Keywords: Brain-computer interface; Electroencephalography; Time-frequency analysis; Electrode reduction; Feature extraction

1 Introduction

After several decades of development, current brain-computer interface (BCI) systems can now be driven based on various types of brain signals obtained by techniques such as electroencephalography (EEG) [1], functional magnetic resonance imaging (fMRI) [2], near-infrared spectroscopy (NIRS) [3], etc. Thanks to its low-cost, non-invasivity, and high temporal resolution, the scalp EEG is a popular technique for BCIs [1]. One typical paradigm of EEG-based BCI is motor imagery BCI, which classifies subject's motor intention based on the spatial difference of EEG patterns. The underlying physiological

phenomenon is that motor imagery of a specific body part (e.g., left hand) induces an event-related desynchronization (ERD) and/or synchronization (ERS) in the μ and β bands over the corresponding functional area in the sensorimotor cortex [4]. Thus, the essential task of a motor imagery BCI is to extract the task-relevant ERD/ERS patterns from EEG signals for classifying subject's motor intentions.

However, poor signal-to-noise ratio (SNR) of raw EEG signal and mixture of different EEG rhythms (e.g., α and μ rhythms) make it difficult to extract ERD/ERS features for BCI classification [5]. One popular solution is to apply a data-driven spatial filtering technique, such as common spatial pattern (CSP) [6], on multi-channel (e.g. 64 or 128 channels) monopolar recording EEG data, which can improve the SNR of signal and extract discriminative features from the mixture of signals, especially for

*Correspondence: Isabelle.Bloch@telecom-paristech.fr

¹ Institut Mines-Telecom, Telecom ParisTech/CNRS LTCI, 46 rue Barrault, Paris 75013, France

⁴ Whist Lab, Paris 75020, France

Full list of author information is available at the end of the article

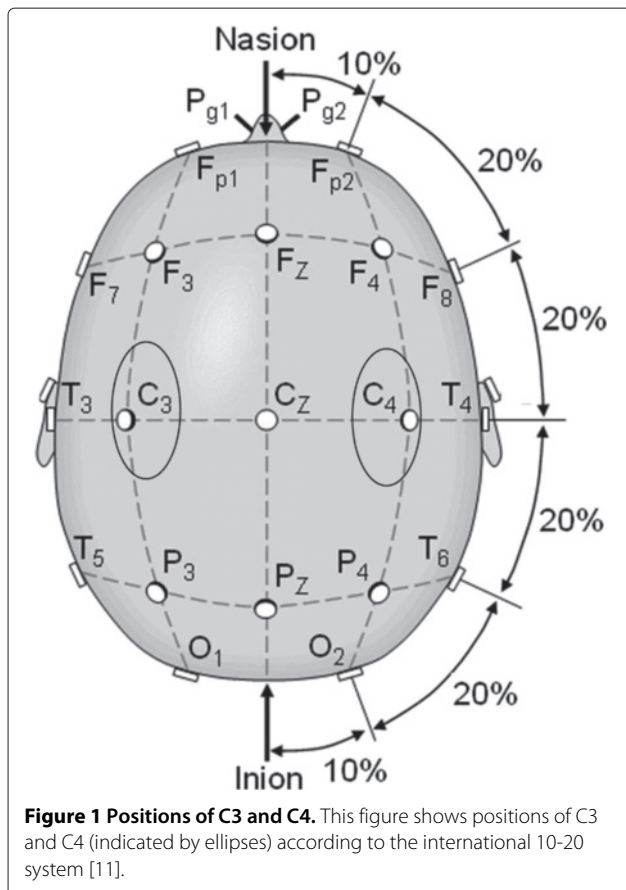
two-class discriminations [7]. But such a multi-channel setting inevitably reduces the portability and practicability of BCIs, which represents a main drawback for end users.

Thus, bipolar recordings are recommended in portable BCI systems to reduce the number of electrodes [8,9]. A bipolar channel of EEG is obtained by subtracting two monopolar EEG signals [10]. This acquisition improves the SNR by eliminating shared artifacts between two monopolar channels (for details, see [9]). Therefore, they may achieve as good performances as usual multi-channel monopolar settings, using only a few electrodes (i.e., two or three pairs of active electrodes) placed around task-relevant sensorimotor areas. The positions of electrodes in bipolar recording can be optimized algorithmically or using prior knowledge on the spatial location of brain activity during motor imagery [9]. Typically, the bipolar electrodes are placed on locations C3 and C4 of the international 10-20 system [11] (see Figure 1) for hand-related motor imagery tasks since these places correspond to the hand representation areas in the cerebral cortex [12].

However, ERD/ERS patterns are typically short-lasting (half to few seconds) and their frequency range may vary with subjects [13]. Thus, only optimizing the position of electrodes may not be sufficient to achieve a good

classification, and a BCI system using bipolar recording also requires more precise user-specific time-frequency parameterization in the feature extraction step. To address this problem, a number of approaches were proposed to estimate time-frequency characteristics of motor imagery EEG [13-16], but only a few were successfully applied to bipolar recording data. Among those methods, the filter bank CSP (FBCSP) method seems to be the most effective one, yielding the best BCI performances on BCI competition datasets [17]. FBCSP was initially proposed only for frequency band optimization and then extended to include an optimal temporal selection process [14]. However, FBCSP-based methods involve feature selection procedures based on mutual information, which require tedious iterative steps that greatly increase their complexity. Moreover, the latest version of FBCSP selects the optimal time segment from only a few different options, which did not yield better results on bipolar recording data (BCI competition IV dataset IIB) compared to previous versions [14].

In this paper, we address the issue of time-frequency optimization with only two bipolar channels (C3 and C4) for the discrimination between right and left hand motor imagery tasks. On the contrary to the coarse selection of the time segment in FBCSP, we propose to take into account fine subject-specific time-frequency characteristics for feature extraction. Moreover, our approach is neither based on CSP algorithm nor combined with complex algorithms, such as mutual information-based algorithms, and employs less electrodes than CSP-based methods. The strategy of subject-specific time-frequency optimization builds on our preliminary work in [18] and includes three steps: (1) the time-frequency domain of the input bipolar channels is divided into a set of overlapping regions with different time segments and frequency bands, (2) the discriminative power of each time-frequency region is measured, and (3) the optimal time-frequency region is selected by finding the region with the largest discriminative power. Once the optimal time-frequency region for each subject is found, the classification is performed using a simple linear classifier, i.e., Fisher's linear discriminant analysis (LDA). This classifier has a very low computational cost and usually yields good results for motor imagery BCIs [19]. Two novel criteria are proposed for the evaluation step (2). One is based on domain-specific knowledge of neurophysiology and is called time-frequency discrimination factor (TFDF), while the other is purely data driven and is named F score. Different from classical criteria (e.g., Fisher criterion, r^2 coefficient) used for ranking one-dimensional observations [20], the proposed criteria are more suitable to quantify how informative a multi-dimensional feature vector is for distinguishing two classes. The comparison between these two criteria, as well as with state-of-the-art



methods, is performed on a standard bipolar dataset (the BCI competition IV Iib), and their contribution to electrode reduction is evaluated on BCI Competition IV Iia dataset.

2 Time-frequency optimization for classification

The EEG signals at C3 and C4 (see Figure 1) are decomposed into signal components first, in a series of overlapping time-frequency regions $(\omega_m \times \tau_n)$, $m \in \{1, \dots, M\}$, $n \in \{1, \dots, N\}$ with different frequency bands $\omega_m = [f_m, f_m + F - 1]$, $f_{m+1} = f_m + F_s$ (F is the bandwidth, F_s is the frequency step), and time intervals $\tau_n = [t_n, t_n + T - 1]$, $t_{m+1} = t_m + T_s$ (T is the interval width, T_s is the time step). The aim of time-frequency optimization is to find a time-frequency region that contains the most discriminative information, so-called the region of interest (ROI), for a given subject. The selected ROI is then used to extract key features for classification.

A measure evaluating the discriminative power of a region should be defined as an increasing function of the discriminative power. This criterion is denoted S . The best ROI $(\omega^* \times \tau^*)$ is estimated by exhaustively searching the largest value of $S(\omega_m, \tau_n)$ among all regions:

$$S(\omega^*, \tau^*) = \max\{S(\omega_m, \tau_n) \mid m \in \{1, 2, \dots, M\}, n \in \{1, 2, \dots, N\}\} \quad (1)$$

The exhaustive search is reasonable with the chosen values of M and N (1,224 regions in our experiments, as detailed in the next section).

The authors in [20] have reviewed the popular measures, such as Fisher criterion, t -scaled difference, r^2 coefficient, which are often used in feature selection for BCI systems. However, those measures are typically used for quantifying the discriminative power of one-dimensional feature and are not appropriate for multi-dimensional feature vector evaluation, in particular for our time-frequency optimization. Thus, two novel criteria, TFDF and F score, are designed to address this problem. Different from those statistical measures often used in BCIs, TFDF is based on neurophysiological background rather than only statistical distribution of features. F score is based on the popular Fisher's discriminant, which can be considered as an extended version of Fisher criterion, but more suitable for estimating the discriminative power of a multi-dimensional feature vector.

2.1 Time-frequency discrimination factor

The proposed criterion for finding the ROI is based on two neurophysiological principles:

1. Motor imagery of one hand typically generates ERD in the contralateral side of brain, so it is possible to discriminate between the imaginations of right and left hand movements by using bipolar electrodes

placed over corresponding hand representation areas, i.e., C3 and C4 [12]. To achieve good classification performances, (1) the pattern difference between imaginations of left and right hand movements should exist in the selected time-frequency region (ROI) at each channel, and (2) the difference between C3 and C4 should also exist in the ROI for both motor imageries.

2. Electrophysiological studies have emphasized the role of volume conduction, so that neural activities in one area are distributed on multiple electrode positions [21]. Due to this effect, the signals of some undesirable EEG rhythms (i.e., common components) are also recorded and mixed with the specific signals of different hand movements, which may deteriorate the classification results [7]. Although bipolar recording can eliminate this effect to some extent, it cannot completely remove all of those common components. Thus, we should consider the influence of those common components in selecting the ROI.

In BCI signal classification, ERD patterns are often estimated by the logarithm of the variance of band-pass filtered EEG in a specific time interval, the so-called logarithmic band power (BP) estimator [22]. The variance of EEG segment in the time domain for each trial i and each channel e is computed as:

$$v_e(i) = \frac{1}{T-1} \sum_{j=t_n}^{t_n+T-1} (x_{ij} - \bar{x}_i)^2 \quad (2)$$

where x_{ij} is the j th sample in the time interval $\tau_n = [t_n, t_n + T - 1]$ of the i th trial of the ω_m -bandpass filtered EEG data, and \bar{x}_i is the mean value over all samples of filtered EEG in the time interval τ_n of the i th trial. Then, the band power feature in each channel is defined as:

$$\text{BP}_e(i) = \log(v_e(i)) \quad (3)$$

The logarithm is applied to make the distribution of BP features approximately normal, so as to feed the linear classifier, Fisher's LDA.

According to this definition, the overall BP, $\widetilde{\text{BP}}_e^\chi$, for each class ($\chi = L, R$) and each channel ($e = C3, C4$) is defined by taking the logarithm of the median or the mean of data variances over trials [18]. Here, we use the median value because it is more robust to outliers. The overall BP then writes:

$$\widetilde{\text{BP}}_e^\chi = \log(\tilde{v}_e^\chi) \quad (4)$$

where \tilde{v}_e^χ denotes the median of data variances $v_e(i)$ over all trials for class χ .

Thus, the pattern difference (PD_e) between two conditions (left vs. right hand) in a time-frequency region ($\omega_m \times \tau_n$) in each channel is expressed as:

$$PD_{C3}(\omega_m, \tau_n) = \widetilde{BP}_{C3}^L(\omega_m, \tau_n) - \widetilde{BP}_{C3}^R(\omega_m, \tau_n) \quad (5)$$

$$PD_{C4}(\omega_m, \tau_n) = \widetilde{BP}_{C4}^L(\omega_m, \tau_n) - \widetilde{BP}_{C4}^R(\omega_m, \tau_n) \quad (6)$$

The sign of PD_e reflects the tendency (increase or decrease) of the BP modulation from condition L (imagination of left hand movement) to condition R (imagination of right hand movement) in channel e .

Imagination of left and right hand movements usually elicit contrary contralateral dominance of ERD at channels C3 and C4 [12,23]. These task-related spatial discriminative modulations can be measured by $|PD_{C3}(\omega_m, \tau_n) - PD_{C4}(\omega_m, \tau_n)|$, called *discriminative force* $Fd(\omega_m, \tau_n)$, to estimate this positive contribution in a time-frequency region ($\omega_m \times \tau_n$). A large $Fd(\omega_m, \tau_n)$ indicates that large discriminative modulations occur in the time-frequency region ($\omega_m \times \tau_n$).

On the other hand, it has been proven that other sources (non-target motor imagery sources) will generate signals (e.g., α -rhythm from the visual cortex) in the same frequency as ERD during the motor imagery (for details, see [7,9]). For example, subjects are looking at the screen during both motor imagery tasks, which can generate visually related common modulations at C3 and C4. Although these sources are not near C3 and C4, they will conduct through scalp and be mixed with discriminative components because of the volume conduction [24]. Meanwhile, neural activities at C3 and C4 will also affect the contralateral channels due to volume conduction. These are what we call common components. They overlap with the discriminative modulations, which present a negative effect on the classification. Thus, we define the *blurring force* $Fb(\omega_m, \tau_n) = |PD_{C3}(\omega_m, \tau_n) + PD_{C4}(\omega_m, \tau_n)|$ to estimate those common modulations in the time-frequency region ($\omega_m \times \tau_n$). A small $Fb(\omega_m, \tau_n)$ indicates that small common modulations happen in the time-frequency region ($\omega_m \times \tau_n$).

Finally, a $TFDF(\omega_m, \tau_n)$ is defined as the difference between $Fd(\omega_m, \tau_n)$ and $Fb(\omega_m, \tau_n)$ to evaluate the overall contribution of the data in the time-frequency area (ω_m, τ_n) from electrodes C3 and C4 for two-class discrimination:

$$\begin{aligned} TFDF(\omega_m, \tau_n) &= Fd(\omega_m, \tau_n) - Fb(\omega_m, \tau_n) \\ &= |PD_{C3}(\omega_m, \tau_n) - PD_{C4}(\omega_m, \tau_n)| \\ &\quad - |PD_{C3}(\omega_m, \tau_n) + PD_{C4}(\omega_m, \tau_n)| \end{aligned} \quad (7)$$

An ideal time-frequency region for classification should have large discriminative modulations (large $Fd(\omega_m, \tau_n)$) and small common modulations (small $Fb(\omega_m, \tau_n)$), so the ROI ($\omega^* \times \tau^*$) is estimated by seeking the maximum value

of $TFDF(\omega_m, \tau_n)$ among the given $M \times N$ time-frequency regions:

$$TFDF(\omega^*, \tau^*) = \max \left\{ TFDF(\omega_m, \tau_n) \mid m \in \{1, 2, \dots, M\}, n \in \{1, 2, \dots, N\} \right\} \quad (8)$$

To examine the behavior of TFDF, we provide its possible values in Table 1 for different cases and present in Figure 2 the ERD/ERS maps and the corresponding TFDF values of 4 Hz, 2-s wide time-frequency regions for an example from a standard dataset. From Table 1, we can see that (1) the values of TFDF are larger for $PD_{C3} \cdot PD_{C4} < 0$ than for $PD_{C3} \cdot PD_{C4} \geq 0$, and (2) the values of TFDF are determined by $\min\{|PD_{C3}|, |PD_{C4}|\}$. Thus, this method tends to seek the time-frequency region where PD_{C3} and PD_{C4} have different signs and large absolute values.

In the ROI, the right hand motor imagery elicits more significant ERD at C3 compared to the left hand motor imagery, which leads to $\widetilde{BP}_{C3}^L(\omega^*, \tau^*) > \widetilde{BP}_{C3}^R(\omega^*, \tau^*)$, while left hand motor imagery generates more significant ERD at C4 compared to the right hand motor imagery, so we have $\widetilde{BP}_{C4}^L(\omega^*, \tau^*) < \widetilde{BP}_{C4}^R(\omega^*, \tau^*)$. Thus,

$$PD_{C3}(\omega^*, \tau^*) = \widetilde{BP}_{C3}^L(\omega^*, \tau^*) - \widetilde{BP}_{C3}^R(\omega^*, \tau^*) > 0 \quad (9)$$

$$PD_{C4}(\omega^*, \tau^*) = \widetilde{BP}_{C4}^L(\omega^*, \tau^*) - \widetilde{BP}_{C4}^R(\omega^*, \tau^*) < 0 \quad (10)$$

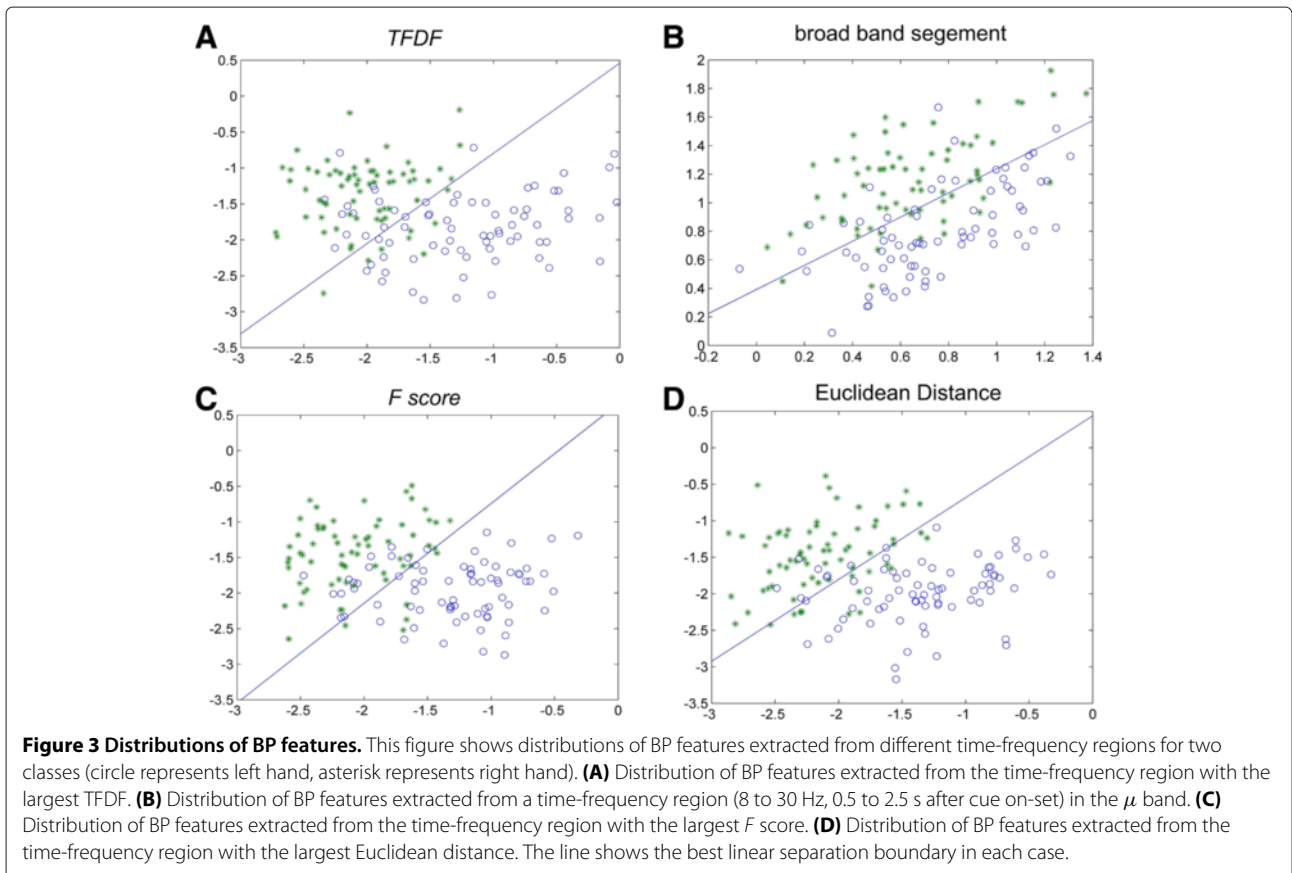
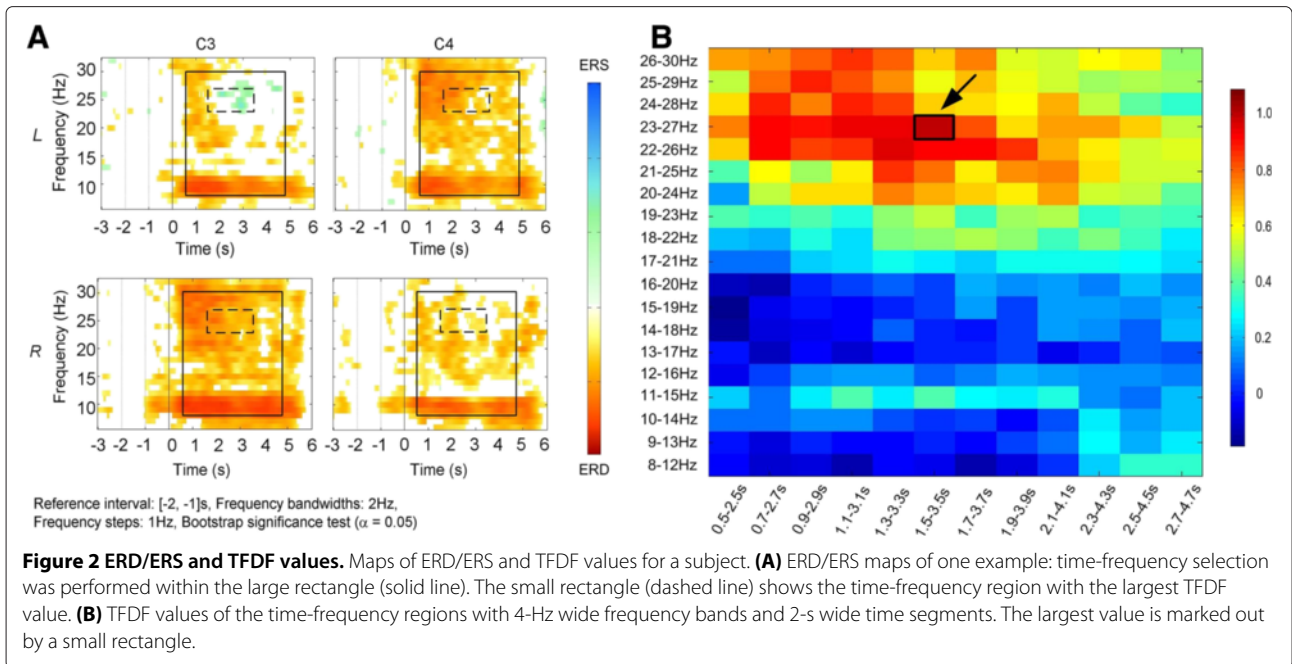
These different signs of PD_{C3} and PD_{C4} reflect the spatial difference of significant ERD between right and left hand motor imageries. On the other hand, large absolute values of PD_e represent the large magnitudes of task-related (i.e., right vs. left hand) difference at channel e (i.e., C3, C4), which also contributes to the discrimination between two tasks.

In the literature, a broad frequency band (i.e., 8 to 30 Hz) EEG segments (0.5 to 2.5 s after cue on-set) was typically chosen for feature extraction because it covers the μ and β bands and usually generates good classification results [6]. For this data example, the frequency band (23 to 27 Hz) of the ROI selected by TFDF is in the range of β band (18 to 25 Hz) but does not completely cover it, and the time segment (1.5 to 3.5 s) is different from the typically used one.

Figure 3A,B shows the distributions of the BP features extracted from the time-frequency region with the

Table 1 Values of TFDF for different pairs of PD_{C3} and PD_{C4}

	$ PD_{C3} > PD_{C4} $	$ PD_{C3} \leq PD_{C4} $
$PD_{C3} \cdot PD_{C4} \geq 0$	$TFDF = -2 PD_{C4} $	$TFDF = -2 PD_{C3} $
$PD_{C3} \cdot PD_{C4} < 0$	$TFDF = 2 PD_{C4} $	$TFDF = 2 PD_{C3} $



largest TDFD and of the ones extracted from the recommended broad frequency band EEG segment in a real data example. The linear separation boundary is obtained by Fisher's LDA in the figure. From the figure, we can see that the BP features extracted from the time-frequency region with the largest TDFD seem more linearly separable than the ones extracted from the recommended broad frequency band EEG segment. The comparison on classification results will be made in the result section to assess the contribution of TDFD to the discrimination between left and right hand motor imageries.

2.2 A criterion based on Fisher's discriminant

Fisher's discriminant analysis (Fisher's LDA) is a very popular classification algorithm in BCI research [19]. It projects high-dimensional data onto a direction and then performs a linear classification in this one-dimensional space. The optimal projection is found by maximizing the separation between two classes. In a one-dimensional feature space, the separation between two classes L and R is defined using the Fisher criterion [20]:

$$FC = \frac{(\mu^L - \mu^R)^2}{(\sigma^L)^2 + (\sigma^R)^2} \quad (11)$$

where μ^L and μ^R are the mean values of the feature over all trials for classes L and R , respectively, and $(\sigma^L)^2$ and $(\sigma^R)^2$ are the trial-wise variances of the feature.

In feature selection, FC can be used to evaluate the discriminative power of each single feature [20]. However, it is not suitable to evaluate the discriminative power of a group of features. Thus, we propose a novel and simplified criterion based on Fisher's discriminant, called F score, \hat{F} , and we use it to estimate the discriminative power of a group of features:

$$\hat{F} = \frac{\|\bar{\mu}^L - \bar{\mu}^R\|_2^2}{tr(\Sigma^L) + tr(\Sigma^R)} \quad (12)$$

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

Without loss of generality, let us discuss this novel criterion in a two-dimensional space with the feature vector $\vec{f}(i) = [f_1(i), f_2(i)]$, $i = 1, \dots, K$, where K is the number of samples (trials) for one class. Thus, the mean of the feature vector for the class is $\bar{\mu} = [\mu_1, \mu_2]$, where μ_1 and μ_2 are the mean values of $f_1(i)$ and $f_2(i)$, respectively. We denote by σ_1^2 and σ_2^2 , the variances of $f_1(i)$ and $f_2(i)$ for the class,

respectively. The trace of the covariance matrix for each class is computed as:

$$\begin{aligned} tr(\Sigma) &= \sigma_1^2 + \sigma_2^2 \\ &= \frac{1}{K-1} \sum_{i=1}^K (f_1(i) - \mu_1)^2 + \frac{1}{K-1} \sum_{i=1}^K (f_2(i) - \mu_2)^2 \\ &= \frac{1}{K-1} \sum_{i=1}^K [(f_1(i) - \mu_1)^2 + (f_2(i) - \mu_2)^2] \end{aligned} \quad (13)$$

Thus, the trace of the covariance matrix for each class is the mean Euclidean distance between samples to the class center, which reflects intra-class spread.

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

In this paper, the BP features $[BP_{C3}(i), BP_{C4}(i)]$ (defined in Equation 3) extracted from the time-frequency ROI are used for classification, so it is also a two-dimensional feature space. We use \hat{F} to estimate the separation between left hand vs. right hand motor imagery in this feature space:

$$\hat{F} = \frac{(\overline{BP}_{C3}^L - \overline{BP}_{C3}^R)^2 + (\overline{BP}_{C4}^L - \overline{BP}_{C4}^R)^2}{\bar{S}_{C3}^L + \bar{S}_{C4}^L + \bar{S}_{C3}^R + \bar{S}_{C4}^R} \quad (14)$$

with:

$$\overline{BP}_e^\chi = \frac{1}{K^\chi} \sum_{i=1}^{K^\chi} BP_e^\chi(i) \quad (15)$$

$$\bar{S}_e^\chi = \frac{1}{K^\chi - 1} \sum_{i=1}^{K^\chi} (BP_e^\chi(i) - \overline{BP}_e^\chi)^2 \quad (16)$$

where K^χ the number of trials for class χ ($\chi \in \{L, R\}$).

We calculate the value of this criterion, $\hat{F}(\omega_m, \tau_n)$, for each time-frequency region ($\omega_m \times \tau_n$), so as to measure whether ($\omega_m \times \tau_n$) contains the most discriminative information. The time-frequency ROI (ω^*, τ^*) is estimated by seeking the maximum value of $\hat{F}(\omega_m, \tau_n)$ among all $M \times N$ time-frequency regions:

$$\hat{F}(\omega^*, \tau^*) = \max \left\{ \hat{F}(\omega_m, \tau_n) \mid m \in \{1, 2, \dots, M\}, n \in \{1, 2, \dots, N\} \right\} \quad (17)$$

Note that outliers are taken into account in the TDFD calculation but not in the F score. The reason is that some outliers (which may be caused by muscle movement of one side of face/body during the experiments) may increase

the difference between the two sides of brain [7]. In this case, the TFDF value will abnormally increase when using the mean value. As a result, the time-frequency area contaminated by noise may be selected by error, which may deteriorate classification results when using Fisher's LDA. On the contrary, for the F score, outliers will increase the intra-class variance, so they will lower the F score. Thus, the time-frequency area contaminated by noise will not be selected (due to a low F score value), though we do not account for the outliers in the calculation of the F score.

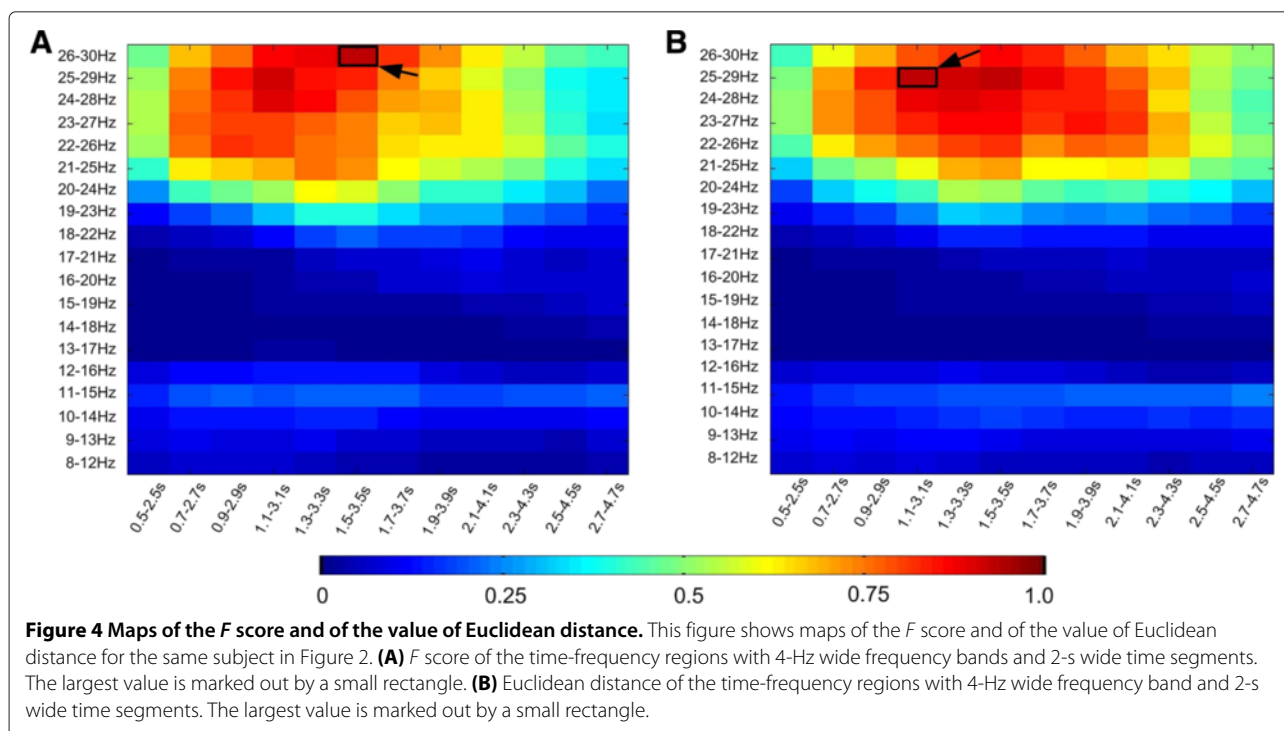
As opposed to the TFDF, the F score is purely based on statistical characteristics of the features, regardless of neurophysiological phenomena linked to a specific task. So, it can be applied in the absence of prior knowledge about task-related neural response.

Figure 4A,B shows the F score and the Euclidean distance between two classes $\|\bar{\mu}^L - \bar{\mu}^R\|_2$ (which reflects only inter-class difference between two classes) in the time-frequency regions with 4-Hz wide frequency bands and 2-s wide time segments for the data example. The large values of F score mainly appear in the frequency band over 20 Hz for this example, which is quite similar to the distribution of Euclidean distance. However, the maximum value appears in different regions. Figure 3C,D shows the distributions of the BP features extracted from the time-frequency regions with the largest F score, and with the largest Euclidean distance, respectively. The linear separation boundary is also obtained by Fisher's LDA in the figure. From Figure 3, we can see that the features

extracted from the time-frequency region with the largest F score and the largest Euclidean distance are more linearly separable than the ones extracted from the broad frequency band EEG segment when using Fisher's LDA as the classifier. Compared to the features from the time-frequency region with the largest Euclidean distance, the intra-class difference is smaller for the features from the time-frequency region with the largest F score. As it considers the intra-class difference, the F score is more adequate than the Euclidean distance as a two-class separation measurement. The overlap area between two classes is smaller for the F score than for the TFDF, so that using the features from the time-frequency region selected by F score has less misclassified data (error rate = 11.25%) than by TFDF (error rate = 13.13%) on the example. This phenomenon indicates that the F score might be more effective than the TFDF. However, the BCI problem is more complicated than a general classification problem [19,20], in particular when training data and testing data are recorded in different sessions [8,25]. Thus, further analysis and comparisons of classification performances with respect to the two criteria on more real data is provided in the 'Experimental results' section.

3 Data description and preprocessing

In this study, we used data of the BCI competition IV dataset Iia [26] and Iib [8]. Dataset Iia was recorded in a multi-channel monopolar setting (22 monopolar channels). The parameters of bipolar channels can be adapted



to the experiments. Dataset IIb was recorded over three bipolar channels, C3, Cz, and C4. Details of these two datasets are provided in the following paragraphs.

3.1 Dataset IIa

BCI competition IV dataset IIa [26] contains one training session and one evaluation session of EEG data from nine subjects who performed four classes of cue-driven motor imagery tasks (left hand, right hand, both feet and tongue). Each trial began with a fixation cross and an additional short acoustic warning tone. After 2 s, a cue in the form of an arrow pointing either to the left, right, down, or up (corresponding to left hand, right hand, foot, or tongue) appeared and stayed on the screen for 1.25 s. The subjects were asked to carry out the motor imagery task until 4 s after cue on-set. No feedback was provided.

The EEG signals were recorded by 22 Ag/AgCl electrodes (with inter-electrode distances of 3.5 cm) using the left mastoid as reference and the right mastoid as ground (sampling rate 250 Hz). The electrode montage is shown in Figure 5A. For extracting a bipolar channel at the position of C3 or C4, three different pairs of electrodes can be used, marked as *a*, *p*, and *l* in Figure 5A. Thus, nine possible channel combinations for C3 and C4 are generated for time-frequency optimization.

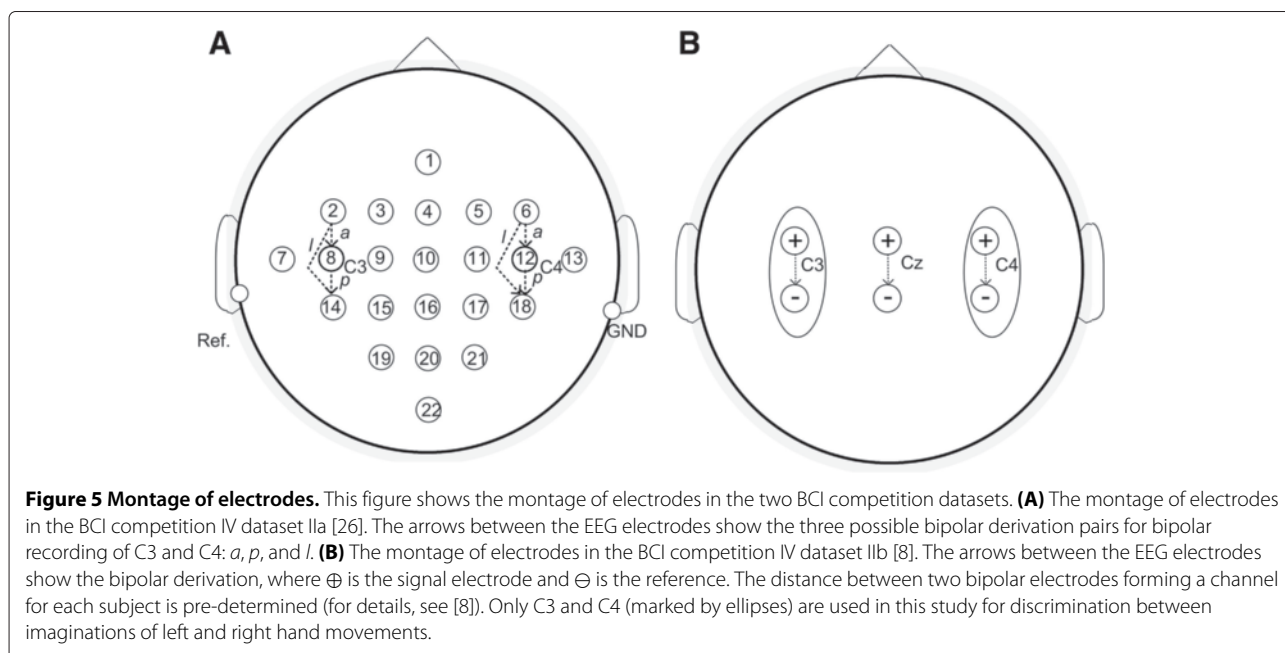
3.2 Dataset IIb

This dataset consists of two classes (left vs. right hand) cue-driven motor imagery BCI data from nine subjects. The EEG data are recorded in three bipolar channels, i.e., at positions C3, Cz, and C4 (see Figure 5B). The distances between the two bipolar electrodes forming a channel are

pre-determined in this dataset (for details, see [8]). For each subject, five sessions are provided, including three training sessions and two evaluation sessions. The first two training sessions consist of 240 single trials (120 single trials per session) totally without visual feedback. Each trial started with a fixation cross and an additional short acoustic warning tone. Later, a visual cue was given to guide the subject to execute the corresponding imagination of hand movement over a period of 4 s. The last training session (160 single trials) and both evaluation sessions (160 trials per session) were recorded with visual feedback from 0.5 to 4.5 s after the cue on-set (for details, see [8]).

3.3 Experimental goals

The proposed time-frequency optimization methods based on different criteria will first be applied on dataset IIb using only two bipolar channels (C3, C4). The goal of this experiment is to evaluate the effectiveness of the methods in improving the performances of BCI based on few channels only. We first train the methods for each subject on the training data and then evaluate them on the testing data for this subject. Note that training and testing sessions are recorded on different days [8,26]. This is called session-to-session transfer [25]. The results of our methods in session-to-session transfers from training sessions to testing sessions will be compared with the winners on this dataset in BCI competition IV. Then, we will apply our methods on two bipolar channels (C3, C4) selected from the 22 monopolar channels of dataset IIa. The classification results obtained on dataset IIa will be compared with those obtained by CSP algorithms using 22



monopolar channels to evaluate the interest of the method for electrode reduction.

3.4 Visualization of ERD/ERS maps

The ERD/ERS patterns are usually expressed as percentage power decrease (ERD) or power increase (ERS) referring to the 1-s interval before the warning tone (for details, see [27]). The time-frequency maps of ERD/ERS for both left (L) and right (R) hands in the bipolar channels C3 and C4 were generated by the Biosig Toolbox using overlapping 2 Hz bands (step = 1 Hz) in the frequency range between 6 and 32 Hz [28]. The statistical significance of the power increases (ERD) and decreases (ERS) was verified by a t -percentile bootstrap test with the confidence interval of $\alpha = 0.05$ [27]. Only the significant ERD/ERS are shown in the maps.

3.5 Data preprocessing for time-frequency optimization

Electrooculogram (EOG), originating from ocular activities (e.g., eye movement, blinks), is the most important source of artifacts in BCI. To prevent the influence of EOG artifacts on the classification results, an automated EOG removal method (for details, see [29]) is first applied on the data as required in the BCI competition [30]. Then, for each bipolar channel, 5th order Butterworth filters are applied to compute 19 successive 4-Hz wide frequency bands of signals ($F = 4$ Hz, $F_s = 1$ Hz): 8 to 12, 9 to 13, 10 to 14 Hz, . . . , 26 to 30 Hz, and 15 successive 8-Hz wide frequency bands of signals ($F = 8$ Hz, $F_s = 1$ Hz): 8 to 16, 9 to 17, 10 to 18 Hz, . . . , 22 to 30 Hz. Then, 36 overlapping time segments in each frequency band were obtained through 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 0.2-s step (i.e. $T_s = 0.2$) moving from 0.5 s after the cue on-set. Those parameters are set based on the experience from competitors reported in BCI competition IV [31] and led to good performances in our previous work [18]. Therefore, there are $(19 + 15) \times 36 = 1,224$ time-frequency areas for subject-specific selection. It has to be mentioned that the selection procedure is not very time-consuming (2 min and 21 s in average, using Matlab 7.10.0 on Window 7 Professional 64bits, CPU 2.66G Hz, RAM 2.0G) and is done offline, so the computational cost is acceptable.

4 Experimental results

In the experiments, the paired sample t test was employed to reveal the statistical significance of the difference between the performances of different methods. The test rejects the null hypothesis at the 0.05 significance level.

4.1 Improving classification performance for dataset IIb

The time-frequency ROI ($\omega_{ROI} \times \tau_{ROI}$) of each training dataset is obtained by maximizing the TDFD and the F

score criteria, respectively. Results are reported in Table 2. These results show that (1) the estimated time-frequency ROIs vary among different subjects, (2) even for the same subject, the estimated ROIs vary among different training sessions, and (3) the two criteria picked out different ROIs for the same training session. The fact that the estimation results depend on the subjects is also reflected in the individual differences of timing and frequency of ERD/ERS patterns. Even for the same subject, the timing and frequency of ERD/ERS may shift across sessions [12], which leads to the intra-subject difference in the estimation of ROIs between sessions. A typical example of time-frequency maps displaying significant ERD (red) and ERS (blue) in a training session (session 3) for a subject (subject 6) in the dataset is shown in Figure 6. The ROIs estimated by the TDFD are marked out by solid rectangles (10 to 14 Hz, 0.7 to 2.7 s), while the ROIs selected by the F score are displayed as dashed rectangles (11 to 15 Hz, 1.1 to 4.1 s). Although the ROIs estimated by the two criteria are different, both ROIs contain discriminative ERD patterns between the two classes, indicating that these two criteria could successfully capture the discriminative part of the signal.

To evaluate the contribution of the proposed time-frequency optimization to classification, ten repetitions of cross-validation are performed on each training session for each subject, using the BP features extracted from the estimated time-frequency ROIs by TDFD and F score, respectively. In each run, we randomly separated the data into calibration (90%) and test (10%) sets and classified the test data using the Fisher's LDA obtained from the calibration set. The classification accuracy (Acc) is defined as the observed agreement between classification outputs and true labels [32]. The cross-validation results are obtained by averaging Acc over 10 runs.

The results are compared to those obtained by using the BP features from the broad band (8 to 30 Hz) EEG segments (0.5 to 2.5 s) with or without CSP filtering. The number of spatial filters used in CSP-based classification for this dataset is two (one pair) because only three bipolar channels (C3, Cz, and C4) are provided in this dataset. The comparisons between different methods are shown in Figure 7 using scatter plots. We can see that using the time-frequency ROI estimated by each criterion can greatly improve the accuracy in most sessions compared to using a broad band EEG segment with or without CSP filtering. Using the F score generates higher accuracy than using the TDFD for most sessions (63.0%), indicating that the F score may be more effective than the TDFD in selecting optimal time-frequency regions for discrimination. In addition, we also observe that using classic CSP filtering generates the worst results in the cross-validations, indicating that classic CSP filtering is not very useful for the data recorded by a very few number of electrodes.

Table 2 Time-frequency ROIs selected by the TDF and F score criteria on dataset IIb

Subject	Session	ω^* (Hz)		τ^* (s)	
		TDF	F score	TDF	F score
1	1	10 to 14	11 to 15	2.9 to 4.9	0.9 to 3.4
	2	10 to 14	10 to 14	2.3 to 4.3	0.9 to 3.9
	3	10 to 14	17 to 25	1.5 to 3.5	0.9 to 3.9
2	1	18 to 22	11 to 19	1.5 to 3.5	2.1 to 4.1
	2	11 to 15	18 to 22	2.7 to 5.2	1.3 to 4.3
	3	11 to 15	9 to 13	2.7 to 5.2	2.9 to 5.4
3	1	8 to 12	12 to 16	0.5 to 3.0	2.3 to 4.3
	2	9 to 13	21 to 25	0.5 to 2.5	2.9 to 5.4
	3	8 to 12	11 to 15	3.5 to 6.5	2.1 to 5.1
4	1	8 to 12	18 to 26	0.5 to 2.5	0.7 to 2.7
	2	9 to 13	10 to 18	0.9 to 2.9	0.7 to 2.7
	3	8 to 12	8 to 12	0.9 to 2.9	0.9 to 2.9
5	1	26 to 30	22 to 30	0.9 to 2.9	0.7 to 2.7
	2	26 to 30	22 to 30	0.7 to 3.2	0.5 to 3.0
	3	23 to 27	22 to 30	1.5 to 3.5	0.5 to 3.5
6	1	11 to 15	11 to 15	1.3 to 3.8	1.9 to 4.4
	2	10 to 14	11 to 15	1.7 to 3.7	0.7 to 3.7
	3	10 to 14	11 to 15	0.7 to 2.7	1.1 to 4.1
7	1	12 to 16	12 to 16	2.7 to 4.7	2.1 to 5.1
	2	12 to 16	11 to 15	3.1 to 5.1	2.9 to 5.4
	3	11 to 15	12 to 16	1.1 to 3.1	0.9 to 2.9
8	1	11 to 15	9 to 13	1.3 to 3.3	1.5 to 4.5
	2	9 to 13	13 to 17	1.1 to 3.1	1.1 to 3.6
	3	10 to 14	8 to 16	0.9 to 2.9	1.3 to 3.3
9	1	22 to 26	21 to 25	1.9 to 4.4	1.1 to 3.1
	2	9 to 13	8 to 12	3.1 to 5.1	3.1 to 6.1
	3	22 to 26	23 to 27	0.9 to 2.9	0.9 to 3.4

To further examine the contributions of these two criteria, session-to-session transfers are performed using the training session which has the best classification result in the cross-validation for each subject. As the independent evaluation data are recorded on a different day than the training sessions, EEG signals of the subjects may change significantly from the training data to the evaluation data. This test aims at evaluating the robustness of the methods to non-stationary signals.

In this test, the classifier is parameterized from the selected training session using the BP features from the corresponding ($\omega^* \times \tau^*$). The ω^* bandpass-filtered EEG segments with the same time length as τ^* (i.e., T) are obtained from each entire single trial of testing data via a 0.2 s step sliding window to generate continuous classification outputs (see Figure 8). According to the BCI competition requirement, the classification performances

in the session-to-session transfers are measured by *kappa coefficient* [32]:

$$\kappa = (\text{Acc} - P_e) / (1 - P_e) \quad (18)$$

where P_e is the chance level for agreement (i.e. $P_e = 0.5$ for two-class problems, so here $\kappa = 2\text{Acc} - 1$). Thus, a larger κ value indicates a better classification performance. The mean kappa value over all subjects of the dataset is denoted by $\bar{\kappa}$.

For this dataset, six BCI groups have reported their results of session-to-session transfers in the BCI competition [30]. We provide information on the methods of the first three winners in Table 3, since their classification results are better than the other three. The first place winner used FBCSP [17], which we have briefly introduced in the introduction section. Except our method based on TDF [18], no one has generated better results

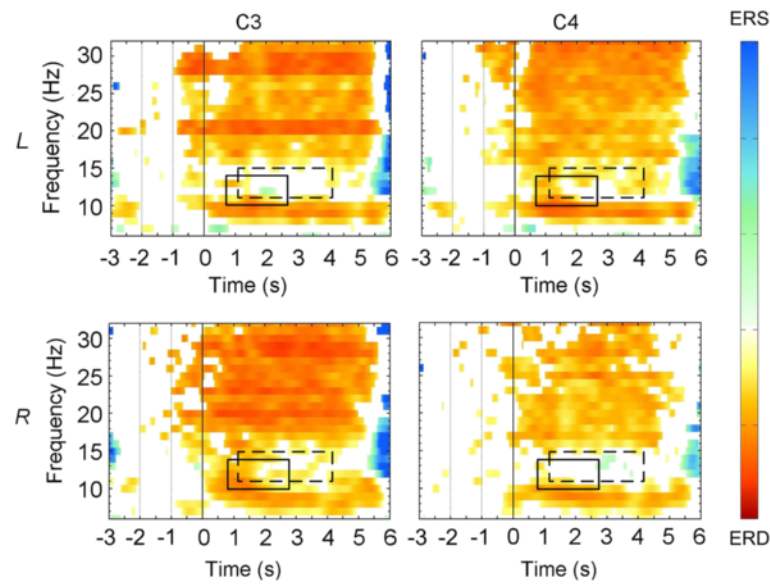


Figure 6 ERD/ERS for subject 6. These time-frequency maps display significant ERD (red) and ERS (blue) for subject 6 (a typical example) in BCI competition IV IIb. The areas in the rectangles are the time-frequency ROI selected by the proposed methods based on TFDF (solid line) and F score (dashed line), respectively.

than FBCSP on this dataset until now [30]. The second place winner employed common spatial subspace decomposition (CSSD) with frequency band and time segment selections [30]. The third place winner applied CSP on spectrally filtered neural time series prediction preprocessing (NTSPP) signals [30]. Note that these methods involved frequency and/or time optimization process(es). The results of session-to-session transfers for all methods are provided in Table 4. The κ values and the number of electrodes ($\#E$) used in the classification are given. TFDF generates the best mean κ value ($\bar{\kappa} = 0.62$) among all methods in the independent evaluation. Although the improvements of κ values yielded by TFDF compared to the first place winner ($\bar{\kappa} = 0.60$, $p = 0.12$) and the second place winner ($\bar{\kappa} = 0.58$, $p = 0.18$) are not statistically significant due to the limited number of subjects, TFDF outperforms the first place winner for six out of nine subjects (except subjects 4, 5, and 8), and the second place winner for six out of nine subjects (except subjects 4, 7, and 9) too. The mean kappa value obtained by the F score ($\bar{\kappa} = 0.60$) is relatively lower than the one obtained by TFDF (not significantly, $p = 0.29$), but comparable to the first place winner and higher than the second place winner (not significantly, $p = 0.52$). It has to be mentioned that the F score yields the best κ values for most subjects (four subjects) among all methods. Further examination of the results show that the poor performance for subject 3 led to a remarkable decrease in mean performance of the F score. In fact, the performances for subject 3

are much poorer than those for the other subjects for all methods, so that results averaged over all subjects might not be representative. Both time-frequency criteria (TFDF and F score) yield better performances than the third place winner ($\bar{\kappa} = 0.46$, both $p < 0.01$) and those obtained by broad band (8 to 30 Hz) EEG segments (0.5 to 2.5 s) with CSP ($\bar{\kappa} = 0.41$, both $p = 0.01$) and without CSP ($\bar{\kappa} = 0.53$, both not significant by $p > 0.05$, even if TFDF and F score outperform it for seven and five out of nine subjects, respectively). Thus, both criteria are promising for seeking optimal time-frequency patterns to improve classification performance of BCIs based on a few bipolar channels.

As all of the first three BCI competition winners have used all three bipolar channels (C3, Cz, and C4) provided by the dataset, our methods not only generate good performances but also use less channels, which indicates that they may also be helpful for channel reduction. This potential contribution is validated in the next subsection.

4.2 Electrode reduction for dataset IIa

In this dataset, the time-frequency ROIs are estimated by the two criteria, respectively, for nine possible channel combinations of C3 and C4 (see Figure 5, C3-C4: $a-a$, $a-p$, $a-l$, $p-a$, $p-p$, $p-l$, $l-a$, $l-p$, $l-l$). The scatter plot of classification accuracies (Acc) obtained by using the TFDF vs. those obtained by using the F score in time-frequency optimization for all channel combinations and all subjects is shown

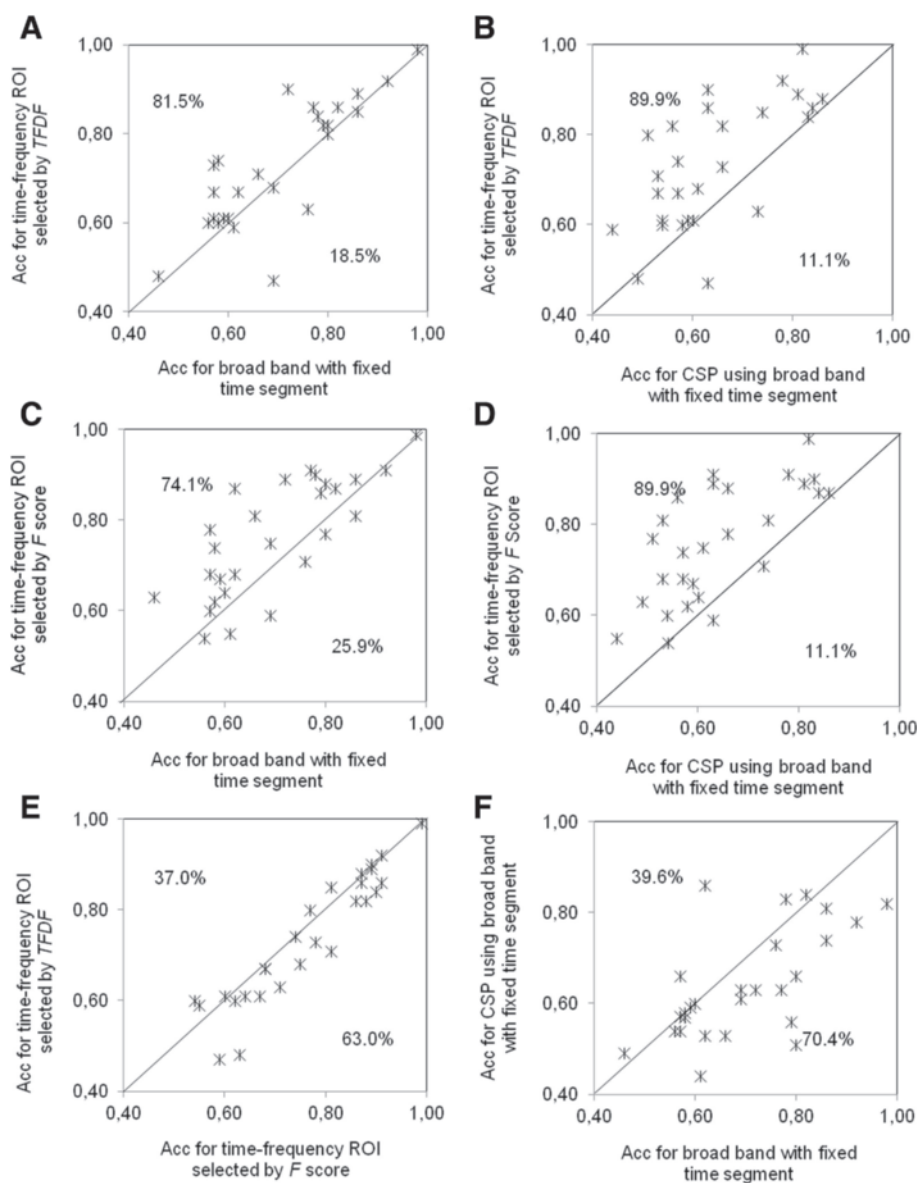


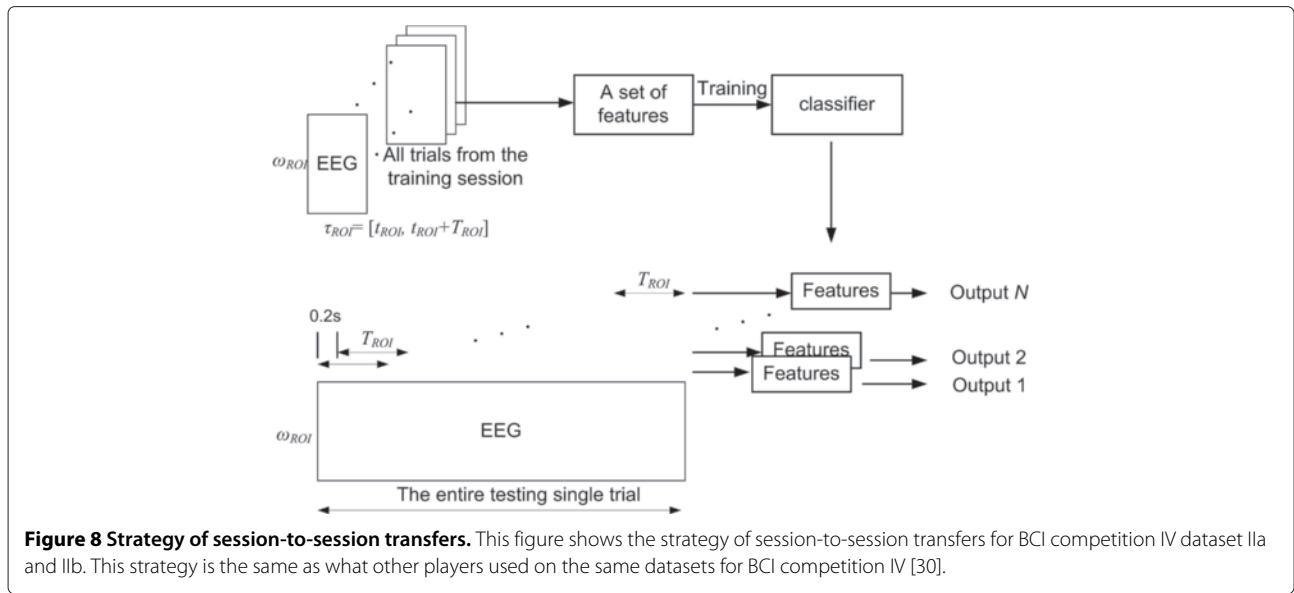
Figure 7 Comparison of method performances in the cross-validation on dataset IIb. This figure shows comparisons of method performances in the cross-validation on dataset IIb. **(A, B)** Scatter plots of classification accuracies (Acc) obtained by using TFDf in time-frequency selection vs. those obtained by broad band (8 to 30 Hz) EEG in a fixed time segment (0.5 to 2.5 s) without and with CSP, respectively. **(C, D)** Scatter plots of Acc obtained by using *F* score in time-frequency selection vs. those obtained by broad band (8 to 30 Hz) EEG in a fixed time segment (0.5 to 2.5 s) without and with CSP, respectively. **(E)** Scatter plots of Acc obtained by using TFDf vs. those obtained by using *F* score in time-frequency selection. **(F)** Scatter plot of Acc obtained by broad band (8 to 30 Hz) EEG in the fixed time segment (0.5 to 2.5 s) with CSP filtering vs. those obtained by the same EEG but without CSP filtering. For the points above the diagonal in each scatter plot, the method in y-axis outperforms the method in x-axis in the cross-validation on the corresponding training session.

in Figure 9. Using the *F* score generates higher accuracy than using the TFDf for most cases (74.1%).

The optimal channel combinations are selected by comparing the classification accuracies (choosing the best one) among different combinations in the 10×10 -fold cross-validation. Optimal channel combinations of C3-C4 and the corresponding estimated time-frequency ROI

for different criteria and different subjects are listed in Table 5.

In session-to-session transfers, the optimal channel combinations are used. The classifier is obtained from the whole training session using the BP features from the corresponding $(\omega^* \times \tau^*)$. The ω^* -bandpass filtered EEG segments with the same time length as τ^* (i.e., T)



are obtained from each entire single-trial of testing data via a 0.2-s step sliding window to generate continuous classification outputs (see Figure 8).

As this study focuses on the two-class (right vs. left hand) problem, it is difficult to compare with BCI competition winners' results (reported based on a four-class problem including tongue and feet motor imagery data) on this dataset. Here, we compared the results obtained by our method with those obtained by FBCSP^a [17], sparse CSP (SCSP) [33], and classic CSP, respectively. Note that FBCSP is believed to be an effective method that well solves the frequency and/or time optimization [14], which has achieved the best classification performance on at least two datasets including dataset IIa in BCI competition IV [30]. SCSP is an optimized CSP that selects the least number of channels in CSP-based classification under a constraint of classification accuracy. SCSP has generated better performances than other channel reduction methods (based on the usual Fisher ratio, mutual information, SVM, CSP coefficients) and the regularized CSP (RCSP) on BCI competition IV dataset IIa for the right

vs. left hand problem (for details, see [33]). The comparisons of classification results and the number of electrodes ($\#E$) used in classification between different methods are given in Table 6. As other researchers provided their classification results as classification accuracy values (Acc, defined in section 'Improving classification performance for dataset IIb') for the right vs. left hand problem on this dataset, we also provide Acc values for the sake of comparison. Table 6 shows that all methods generate better mean performances than the classical CSP algorithm with all 22 monopolar channels (mean classification accuracy, $\overline{\text{Acc}} = 77.26\%$), indicating the interest of time-frequency selection and electrode reduction. Our method based on F score ($\overline{\text{Acc}} = 79.67\%$) yields slightly better results than FBCSP ($\overline{\text{Acc}} = 79.17\%$) and SCSP ($\overline{\text{Acc}} = 79.07\%$) but using far less electrodes on this dataset: our method used only the two bipolar channels C3 and C4 (equivalent to four monopolar channels); FBCSP used all 22 monopolar channels, and SCSP used 8.55 monopolar channels in average [33]. Further examination of individual results shows that our method based on F score generates the

Table 3 Comparison between our method and the first three winners on BCI competition IV dataset IIb

	Time-frequency selection	Electrodes used	Features	Classifier
Our method	Selected by TDF or F score	C3, C4	BP	LDA
First winner [17]	Mutual information-based selection	C3, Cz, C4	FBCSP	Naïve Bayes Parzen Window classifier
Second winner [30]	Selected by classification performance in cross-validation	C3, Cz, C4	CSSD	LDA
Third winner [30]	Selected by a heuristic search and a selection criterion based on overall classification accuracy in cross-validation.	C3, Cz, C4	CSP	Using the best classifier among three variants of LDA and two variants of SVM

Table 4 Performances of different methods in session-to-session transfers on BCI competition IV dataset IIb

	#E	κ values for subjects									Mean κ
		1	2	3	4	5	6	7	8	9	
TFDF	4	<i>0.44</i>	0.24	<i>0.25</i>	0.93	0.86	<i>0.70</i>	0.55	0.85	0.75	0.62
F score	4	0.39	<i>0.25</i>	0.13	0.93	<i>0.88</i>	0.63	0.55	<i>0.88</i>	<i>0.78</i>	0.60
Without CSP	6	0.40	0.24	0.18	0.94	0.39	0.66	0.52	0.81	0.68	0.53
With CSP	6	0.28	0.13	0.11	0.47	0.56	0.13	0.58	0.76	0.67	0.42
FBCSP (first) [17]	6	0.40	0.21	0.22	<i>0.95</i>	0.86	0.61	0.56	0.85	0.74	0.60
CSSD (second) [30]	6	0.43	0.21	0.14	0.94	0.71	0.62	<i>0.61</i>	0.84	<i>0.78</i>	0.58
NTSPP (third) [30]	6	0.19	0.12	0.12	0.77	0.57	0.49	0.38	0.85	0.61	0.46

Best scores are indicated in italics.

best Acc for most subjects (four subjects), indicating that it is the most effective on this dataset. Although the mean classification result of our method based on TFDF ($\overline{\text{Acc}} = 78.00\%$) is slightly lower than those of FBCSP and SCSP, the differences are not statistically significant ($p > 0.05$). Comparing individual performances, our method based on TFDF outperforms FBCSP and SCSP for five out of nine subjects. Moreover, our method based on TFDF also employs less electrodes than FBCSP and SCSP in the classification. Thus, the method based on TFDF still meets the goal of electrode reduction without a significant drop of classification accuracy. Generally speaking, our method based on both criteria can effectively select time-frequency ROI for BCI classification based on only a few channels and therefore contributes to electrode reduction.

Let us mention that four electrodes (two bipolar channels) are the smallest set of electrodes required for a good

performance in left vs. right hand motor imagery discrimination on these data. Although using a common reference (e.g., Cz) for C3 and C4 can further reduce the number of electrodes to three, this monopolar setting will significantly deteriorate the classification performances ($p < 0.05$, see Figure 10). This result, to some extent, also indicates that the bipolar setting is more effective than the monopolar setting in a BCI with only few channels.

5 Discussion

A possible widespread use of BCI is limited by many issues, such as the inter-subject variability and the number of electrodes used. Individual differences of brain pattern will deteriorate the performance of BCI when using a general parameter setting, such as features from a broad frequency band (8 to 30 Hz), for all subjects. In this paper, subject-specific time-frequency characteristics are captured by the proposed method to solve this problem, so as to increase the inter-subject robustness. By this subject-specific time-frequency optimization, our method improves the performance of BCI, in particular when only a few channels of data are available. As our method is applied with only two bipolar channels, it also reduces the number of electrodes required in a BCI system.

In the proposed method, two alternative criteria, TFDF and F score, are proposed for measuring the discrimination power of each possible time-frequency region. Both criteria have their novelties and contributions.

Different from classical criteria, TFDF measures the discrimination power of features based on neurophysiologic phenomena (task-relevant ERD), on which a motor imagery BCI relies, considering both discriminative and common modulations instead of only the statistical distribution of features. Like the CSP algorithm, the method based on TFDF is effective to extract discriminative spatial patterns for motor imagery BCI. However, the proposed method requires less electrodes than the CSP algorithm and its variants. Please note that the TFDF can be adopted in other motor imagery BCI problems by placing the electrodes on the task-relevant areas (e.g., using C3 and Cz

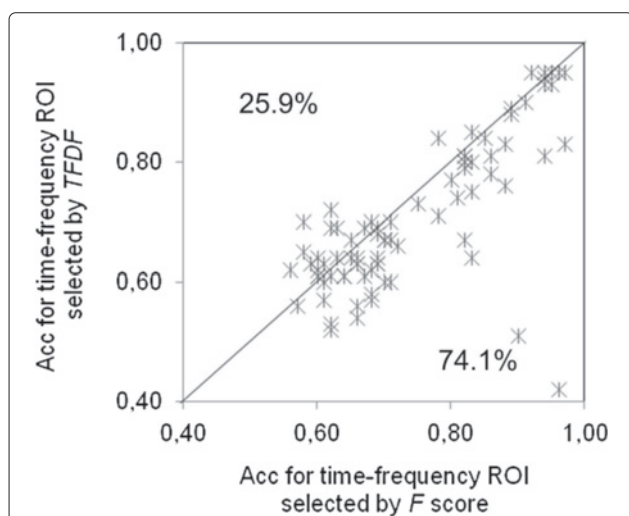


Figure 9 Comparison between performances of TFDF and F score in the cross-validation on dataset IIa. This figure shows the comparison between performances of TFDF and F score in the cross-validation on dataset IIa. F score generated better performance than TFDF in most cases.

Table 5 Optimal channel combinations of C3-C4 and the selected time-frequency ROI on BCI competition IV dataset IIa

Subject	C3-C4		ω^* (Hz)		τ^* (s)	
	TFDF	<i>F</i>	TFDF	<i>F</i> score	TFDF	<i>F</i> score
1	<i>a-a</i>	<i>a-l</i>	12 to 16	11 to 15	1.3 to 3.8	0.5 to 3.0
2	<i>l-l</i>	<i>a-l</i>	10 to 14	22 to 26	1.1 to 3.1	0.5 to 3.5
3	<i>a-a</i>	<i>a-a</i>	10 to 14	11 to 15	0.7 to 2.7	0.7 to 3.2
4	<i>p-p</i>	<i>a-a</i>	21 to 25	11 to 19	0.5 to 2.5	1.1 to 3.1
5	<i>p-l</i>	<i>p-a</i>	25 to 29	26 to 30	1.5 to 3.5	1.3 to 4.3
6	<i>l-l</i>	<i>l-p</i>	22 to 26	23 to 27	0.5 to 2.5	0.7 to 2.7
7	<i>a-a</i>	<i>a-l</i>	19 to 23	18 to 22	0.7 to 2.7	1.1 to 4.1
8	<i>l-l</i>	<i>l-p</i>	8 to 12	8 to 12	0.7 to 2.7	0.7 to 3.2
9	<i>a-l</i>	<i>p-a</i>	11 to 15	10 to 18	0.5 to 2.5	0.7 to 2.7

for discrimination between foot and right hand motor imagery).

The *F* score is a data-driven criterion and easy to compute. As an improvement to the Fisher criterion, which is typically used to measure the discriminative power of a single feature in BCI [20], the *F* score provides an effective measure for evaluating the discriminative power of a group of features (a multi-dimensional feature vector), in particular for time-frequency selection in motor imagery-based BCI. As the *F* score does not require any prior knowledge of neurophysiology, it might be possible to extend its applications on other problems outside the BCI field.

The comparison between the two criteria shows that the *F* score generated better cross-validation performances than the TFDF on both datasets (see Figures 7E and 9). As the *F* score tends to select the time-frequency region by minimizing the overlapping area between two classes (see Figure 3), it is not surprising to have these results when the testing and training data are from the same session and recorded during the same day. However, as we mentioned, a real BCI problem can be more complicated than the cross-validation within one session. When the testing and training data are recorded in two separate days, the unpredictable data evaluation may happen due to slight shift of electrode positions and impedances

and possible changes in the subject's motivation level [25]. This phenomenon gives TFDF a chance to outperform the *F* score for some subjects (such as subjects 1, 3, and 6 on dataset IIb and subject 4 on dataset IIa), since TFDF selects the time-frequency region not just based on the statistical distribution of features but more on task-relevant ERD, whose frequency characteristics may not change a lot between different days for the same subject.

Although TFDF and *F* score alternatively outperformed each other on two different datasets, both of them generated better individual performances than the state-of-the-art methods in most subjects for both datasets as we mentioned in the 'Experimental results' session. Thus, generally speaking, our method, either based on TFDF or *F* score, is robust to session-to-session transfers. As a result, the training data only need to be collected one time for learning the subject-specific time-frequency region and training the classifier, and then the parameters can be used on the same subject for a long-term on-line classification. The time for collecting the training data, on the one hand, depends on the amount of training data required for well describing the different classes; on the other hand, it is affected by the time needed for skin preparation and electrode placing. As the proposed method uses less electrodes than other methods, it will

Table 6 Performances of different methods in session-to-session transfers on BCI competition IV dataset IIa

Method	#E	Acc (%) for subjects									Mean Acc
		1	2	3	4	5	6	7	8	9	
TFDF	4	87.23	66.20	97.81	68.97	64.44	69.44	68.57	96.27	83.08	78.00
<i>F</i> score	4	89.36	69.01	97.81	66.38	66.67	72.22	68.57	97.01	90.00	79.67
FBCSP ^a	22	94.44	52.77	93.05	65.97	88.19	60.41	70.13	94.44	93.05	79.17
SCSP [33]	8.55	91.66	60.41	97.14	70.83	63.19	61.11	78.47	95.13	93.75	79.07
CSP [34]	22	83.51	56.53	97.50	70.00	54.50	62.49	84.50	95.57	90.77	77.26

^aThe results of FBCSP on this dataset for right vs. left hand problem are provided by the BCI lab at Institute for Infocomm Research, Singapore, using all 22 monopolar channels and the Naïve Bayes Parzen Window classifier. Best scores are indicated in italics.

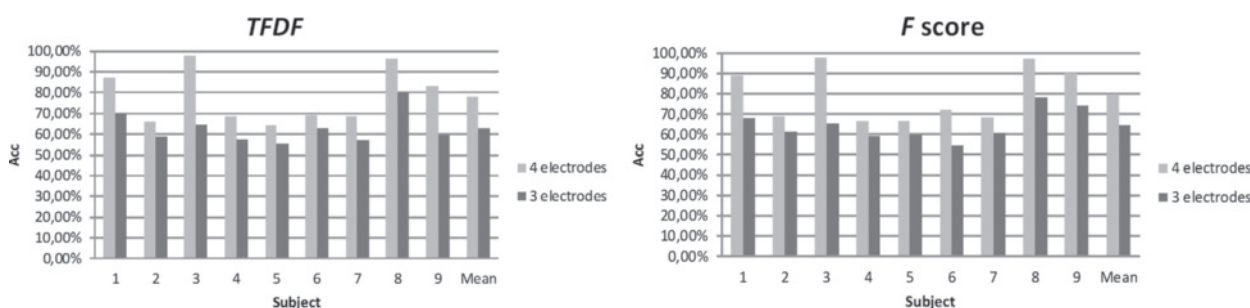


Figure 10 Comparison between using four electrodes and using three electrodes on dataset IIa. This figure shows the comparison between performances of using four electrodes (two bipolar channels) and using three electrodes (two monopolar channels using Cz as common reference) for TFDF and F score, respectively, in session-to-session transfers on dataset IIa. Reducing the number of electrodes to three significantly deteriorates the classification performance ($P < 0.05$).

save time during skin preparation and electrode placing. The amount of training data required is affected by the dimensionality of features, D_f , and the artifacts, since the classifier will fail to give a good performance when the ratio of useful training trials (Num_{tr}) to the dimensionality of features, Num_{tr}/D_f , is too small [19]. Note that most artifacts in BCI are from EOG, which can be removed by the EOG removal technique, we used in this work to maintain the amount of useful training trials [29]. Our method reduces the dimensionality of features by using less channels of data, so from this view, it will, to some extent, either improve the classifier training by increasing Num_{tr}/D_f or save the time for collecting the training data if keeping the same Num_{tr}/D_f as other methods.

Algorithm computational complexity is also an important factor when considering a real application of BCI. Before further discussion on this issue, we first need to distinguish the computational complexity of off-line analysis and online classification since the importance of these two computational complexities are not at the same level. A time-consuming off-line analysis may not be a real problem for some BCI applications [33], but the speed of online classification does affect the usage of a BCI system. Note that the time-frequency optimization is an off-line analysis; thus, its computational complexity may not be a key issue. Nevertheless, compared to FBCSP [14], this off-line analysis is indeed inexpensive in terms of computational cost in our method, neither involving mutual information calculation nor eigenvalue decomposition. In the online classification, band power features are directly extracted from the optimal time-frequency area in our method. This feature extraction step in our method does not involve matrix multiplication, which is needed in CSP-based methods [7]. Furthermore, the computational complexity of the on-line classification is proportional to the dimensionality of feature, D_f , when using a classifier like LDA [35]. Our method only uses two channel data,

i.e., $D_f = 2$, which is not larger than any CSP-based methods ($D_f = 2p$, where p is the number of paired CSP filters). Therefore, thanks to its simplicity, our method is inexpensive in terms of computational complexity for BCI usage.

Last but not least, the proposed method may also contribute to reduce the hardware cost in a BCI system since less electrodes are required for a good classification.

6 Conclusions

Previous studies have shown the importance of subject-specific parametrization in the preprocessing steps of motor imagery BCI, especially in terms of time segment and frequency band selection. This contribution, through the proposition of a novel algorithm and two new criteria, addresses subject-specific time-frequency optimization for motor imagery BCIs. Experimental results show that our method can achieve comparable or even better classification results than state-of-the-art methods, using fewer electrodes (only four active electrodes for composing two bipolar channels around C3 and C4 for discriminating right and left hand motor imagery). Thus, this work, on the one hand, provides an effective method for time-frequency optimization; on the other hand, it shows the potential contribution of time-frequency optimization to electrode reduction that is reducing the number of electrodes required in a BCI system without deteriorating its classification performance. Thanks to its simplicity and few electrodes requirement, the proposed method is promising for portable BCI systems. Additionally, two novel criteria were also proposed in this paper to overcome the disadvantage of classical measures often used in BCIs, in particular for time-frequency optimization. In this paper, the applications of our method focused on the most popular motor imagery BCI task: the discrimination between right and left hand motor imagery data. In the future, this study should be extended

to address time-frequency optimization for multi-class motor imagery BCIs.

Endnote

^aThe results of FBCSP on this dataset for right vs. left hand problem are provided by the BCI lab at Institute for Infocomm Research, Singapore, using all 22 monopolar channels and the Naïve Bayes Parzen Window classifier.

Competing interests

The authors declare that they have no competing interests.

Authors' information

Yuan Yang has finished his PhD in Telecom ParisTech and is currently with Delft University of Technology.

Acknowledgements

This work was partially supported by grants from China Scholarship Council and Orange Labs. Authors would like to thank Dr. O. Kyrgyzov (Whist Lab, France) for some useful discussions, Ms. M. Arvaneh and Dr. K. K. Ang (I²R BCI Lab in A*STAR, Singapore) for providing the results of FBCSP and helpful suggestions in experimental comparison.

Author details

¹Institut Mines-Telecom, Telecom ParisTech/CNRS LTCI, 46 rue Barrault, Paris 75013, France. ²Université de Versailles St-Quentin, Vélizy 78140, France. ³Orange Labs R&D, Issy les Moulineaux 92130, France. ⁴Whist Lab, Paris 75020, France.

Received: 2 October 2013 Accepted: 5 March 2014

Published: 26 March 2014

References

- JR Wolpaw, N Birbaumer, DJ McFarland, G Pfurtscheller, TM Vaughan, Brain-computer interfaces for communication and control. *Clin. Neurophysiol.* **113**(6), 767–791 (2002)
- N Weiskopf, F Scharnowski, R Veit, R Goebel, N Birbaumer, K Mathiak, Self-regulation of local brain activity using real-time functional magnetic resonance imaging (fMRI). *J. Physiol.-Paris.* **98**(4–6), 357–373 (2004)
- R Sitaram, H Zhang, C Guan, M Thulasidas, Y Hoshi, A Ishikawa, K Shimizu, N Birbaumer, Temporal classification of multichannel near-infrared spectroscopy signals of motor imagery for developing a brain-computer interface. *NeuroImage.* **34**(4), 1416–1427 (2007)
- G Pfurtscheller, FH Lopes da Silva, Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin. Neurophysiol.* **110**(11), 1842–1857 (1999)
- M Naeem, C Brunner, G Pfurtscheller, Dimensionality reduction and channel selection of motor imagery electroencephalographic data. *Comput. Intell. Neurosci.* **2009**, 1–8 (2009)
- J Müller-Gerking, G Pfurtscheller, H Flyvbjerg, Designing optimal spatial filters for single-trial EEG classification in a movement task. *Clin. Neurophysiol.* **110**(5), 787–798 (1999)
- B Blankertz, R Tomioka, S Lemm, M Kawanabe, KR Müller, Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal Proc. Mag.* **25**, 41–56 (2008)
- R Leeb, F Lee, C Keirath, R Scherer, H Bischof, G Pfurtscheller, Brain-computer communication: motivation, aim, and impact of exploring a virtual apartment. *IEEE Trans. Neural Syst. Rehabil. Eng.* **15**(4), 473–482 (2007)
- B Lou, B Hong, X Gao, S Gao, Bipolar electrode selection for a motor imagery based brain-computer interface. *J. Neural Eng.* **5**, 342–349 (2008)
- JW Osselton, Acquisition of EEG data by bipolar unipolar and average reference methods: a theoretical comparison. *Electroencephalogr. Clin. Neurophysiol.* **19**(5), 527–528 (1965)
- F Shاربrough, GE Chatrian, RP Lesser, H Lüders, M Nuwer, TW Picton, American electroencephalographic society guidelines for standard electrode position nomenclature. *Electroencephalogr. Clin. Neurophysiol.* **8**, 200–202 (1991)
- G Pfurtscheller, C Neuper, D Flotzinger, M Pregenzer, EEG-based discrimination between imagination of right and left hand movement. *Electroencephalogr. Clin. Neurophysiol.* **103**(6), 642–651 (1997)
- N Yamawaki, C Wilke, Z Liu, B He, An enhanced time-frequency-spatial approach for motor imagery classification. *IEEE Trans. Neural Syst. Rehabil. Eng.* **14**(2), 250–254 (2006)
- KK Ang, ZY Chin, H Zhang, C Guan, Mutual information-based selection of optimal spatial-temporal patterns for single-trial EEG-based BCIs. *Pattern Recogn.* **45**, 2137–2144 (2011)
- NF Ince, S Arica, A Tewfik, Classification of single trial motor imagery EEG recordings with subject adapted non-dyadic arbitrary time-frequency tilings. *J. Neural Eng.* **3**, 235 (2006)
- B Wu, F Yang, J Zhang, Y Wang, X Zheng, W Chen, A frequency-temporal-spatial method for motor-related electroencephalography pattern recognition by comprehensive feature optimization. *Comput. Biol. Med.* **42**, 253–263 (2012)
- KK Ang, ZY Chin, C Wang, C Guan, H Zhang, Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b. *Front. Neurosci.* **6**, 39 (2012)
- Y Yang, S Chevallier, J Wiart, I Bloch, 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)* (IEEE, San Diego, 2012), pp. 2744–2747
- F Lotte, M Congedo, A Lécuyer, F Lamarche, B Arnaldi, A review of classification algorithms for EEG-based brain-computer interfaces. *J. Neural Eng.* **4**, R1–R13 (2007)
- KR Müller, M Krauledat, G Dornhege, G Curio, B Blankertz, Machine learning techniques for brain-computer interfaces. *Biomed. Tech. (Berl.)* **49**, 11–22 (2004)
- MJ Peters, HJ Wieringa, The influence of the volume conductor on electric source estimation. *Brain Topogr.* **5**(4), 337–345 (1993)
- C Vidaurre, N Kramer, B Blankertz, A Schlögl, Time domain parameters as a feature for EEG-based brain-computer interfaces. *Neural Netw.* **22**(9), 1313–1319 (2009)
- G Pfurtscheller, C Brunner, A Schlögl, FH Lopes da Silva, Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks. *NeuroImage.* **31**, 153–159 (2006)
- PL Nunez, R Srinivasan, AF Westdorp, RS Wijesinghe, DM Tucker, RB Silberstein, PJ Cadusch, EEG coherency I: statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales. *Electroencephalogr. Clin. Neurophysiol.* **103**(5), 499–515 (1997)
- B Blankertz, K Müller, D Krusienski, G Schalk, J Wolpaw, A Schlögl, G Pfurtscheller, J Millán, M Schroder, N Birbaumer, The BCI competition III: validating alternative approaches to actual BCI problems. *IEEE Trans. Neural Syst. Rehabil. Eng.* **14**(2), 153–159 (2006)
- C Brunner, M Naeem, R Leeb, B Graimann, G Pfurtscheller, Spatial filtering and selection of optimized components in four class motor imagery EEG data using independent components analysis. *Pattern Recogn. Lett.* **28**(8), 957–964 (2007)
- B Graimann, JE Huggins, SP Levine, G Pfurtscheller, Visualization of significant ERD/ERS patterns in multichannel EEG and ECoG data. *Clin. Neurophysiol.* **113**, 43–47 (2002)
- A Schlögl, C Brunner, BioSig: a free and open source software library for BCI research. *Computer.* **41**(10), 44–50 (2008)
- A Schlögl, C Keirath, D Zimmermann, R Scherer, R Leeb, G Pfurtscheller, A fully automated correction method of EOG artifacts in EEG recordings. *Clin. Neurophysiol.* **118**, 98–104 (2007)
- M Tangermann, KR Müller, A Aertsen, N Birbaumer, C Braun, C Brunner, R Leeb, C Mehring, KJ Miller, GR Müller-Putz, G Nolte, G Pfurtscheller, H Preissl, G Schalk, A Schlögl, C Vidaurre, S Waldert, B Blankertz, Review of the BCI competition IV. *Front. Neurosci.* **6**, 55 (2012)
- B Blankertz, BCI competition IV: results (2008). <http://www.bbci.de/competition/iv/results/>. Accessed 26 September 2011
- A Schlögl, J Kronegg, JE Huggins, S Mason, G Dornhege, Evaluation criteria for BCI research, in *Toward Brain-Computer Interfacing* (The MIT Press, Cambridge, 2007), pp. 327–342
- M Arvaneh, C Guan, KK Ang, C Quek, Optimizing the channel selection and classification accuracy in EEG-based BCI. *IEEE Trans. Biomed. Eng.* **58**(6), 1865–1873 (2011)

34. Y Yang, S Chevallier, J Wiart, I Bloch, 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)* (ESANN, Bruges, 2012), pp. 109–114
35. D Cai, X He, J Han, Training linear discriminant analysis in linear time, in *IEEE 24th International Conference on Data Engineering (ICDE 2008)* (IEEE, Cancun, 2008), pp. 209–217

doi:10.1186/1687-6180-2014-38

Cite this article as: Yang et al.: Time-frequency optimization for discrimination between imagination of right and left hand movements based on two bipolar electroencephalography channels. *EURASIP Journal on Advances in Signal Processing* 2014 **2014**:38.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
