Biomedical Signal Processing and Control 5 (2010) 174–180

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Technical note

# A frequency-weighted method combined with Common Spatial Patterns for electroencephalogram classification in brain–computer interface

Guangquan Liu, Gan Huang, Jianjun Meng, Xiangyang Zhu*

*State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University, Shanghai 200240, China*

ARTICLE INFO

*Article history:*
Received 4 May 2009
Received in revised form
25 December 2009
Accepted 23 February 2010
Available online 21 March 2010

*Keywords:*
Brain–computer interface (BCI)
Electroencephalogram (EEG)
Common Spatial Patterns (CSP)
Frequency-weighted method (FWM)

ABSTRACT

Common Spatial Patterns (CSP) has been proven to be a powerful and successful method in the detection of event-related desynchronization (ERD) and ERD based brain–computer interface (BCI). However, frequency optimization combined with CSP has only been investigated by a few groups. In this paper, a frequency-weighted method (FWM) is proposed to optimize the frequency spectrum of surface electroencephalogram (EEG) signals for a two-class mental task classification. This straightforward method computes a weight value for each frequency component according to its importance for the discrimination task and reforms the spectrum with the computed weights. The off-line analysis shows that the proposed method achieves an improvement of about 4% (averaged over 24 datasets) in terms of cross-validation accuracy over the basic CSP.



## 1. Introduction

People with severe neuromuscular disorders or suffering from a locked-in syndrome need alternative methods for communication and control. Brain–computer interfaces (BCIs) are systems that allow their users to communicate with a computer program or control a mechanical device directly by intent rather than by neuromuscular passway [1]. Existing literature suggests that both actual movements [2] and imaginary movements [3,4] of different body parts can cause circumscribed attenuation of mu and beta rhythms at corresponding cortex locations, and this attenuation is called event-related desynchronization (ERD) [3]. Meanwhile, an enhancement of mu and beta rhythm at other cortex locations can be observed, which is termed as event-related synchronization (ERS) [3]. Since different movement intents can be reflected by different ERD/ERS patterns [5,6], ERD/ERS based BCIs are widely studied by many groups [7–9].

As mentioned above, different body parts are related to different locations in the motor and somatosensori cortex [5]. For example, movement or imagery movement of left hand will cause ERD in right motor cortex and ERS in left motor cortex, and vice versa. Therefore, spatial information in multi-channel EEG recordings is important for discriminating different intent patterns. Since raw EEG recordings have a poor spatial resolution due to volume conduction, various spatial filters have been used to extract spatial information from raw EEG recordings, such as Common Average Reference (CAR), bipolar reference, small Laplacian reference, large Laplacian reference [10,11], Principal Components Analysis (PCA), Independent Components Analysis (ICA) [12,13], and Common Spatial Patterns (CSP) [10,11,14]. CSP, which is a supervised method aiming at finding the most discriminative components for two-class tasks, has been reported to be a powerful and successful algorithm for ERD/ERS detection and ERD based BCIs [5].

Besides spatial information, frequency information is also important for the discriminating task, since the effects of different activations on EEG recordings are not reflected to the same degree in all frequency bands. In general, the mu rhythm has a frequency band of 8–12 Hz and beta rhythm 18–25 Hz, but these frequency bands can vary with subjects and mental states of the subject [15,16]. In CSP method, the criteria of separability is the ratio of variance of the decomposed components, but the specific frequency band of the decomposed components are not considered. In order to take full advantage of both frequency and spatial information, frequency band selection work or other frequency optimization work has to be done before CSP is applied. However, most studies on CSP either used an unspecifically selected broad band [11] or manually selected among several predefined narrow bands [10].

Recently, some automatical and flexible frequency filtering methods combined with CSP have been proposed [17–20]. In [17], an extension of CSP named Common Spatio Spectral Pattern (CSSP) was proposed to achieve a simultaneous spatial and spectral optimization. This method automatically computes a suitable FIR filter by feeding a time-delay embedded signal into a CSP procedure.

* Corresponding author.
*E-mail address:* mexyzhu@sjtu.edu.cn (X. Zhu).

1746-8094/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bspc.2010.02.004

However the flexibility of the frequency filter is very limited. In [18], a Common Sparse Spectral Spatial Pattern (CSSSP) was proposed, which solves the problem of flexibility in CSSP by computing a $T$-length sparse FIR filter. However, the CSSSP needs extensive parameter tuning, and the complexity of the optimization problem leads to computational inefficiency. In [19,20] a spectrally weighted CSP (SPEC-CSP) was proposed, which uses an iterative procedure to achieve optimization of both spatial and spectral filters. The need of extensive parameter tuning and iterative computation limits the real application of this method.

In this paper, we propose a straightforward frequency-weighted method (FWM) to achieve automatic frequency optimization before CSP is applied. Our method is different from the above mentioned methods in the following ways: (a) this method does not try to achieve a simultaneous optimization of spectral and spatial filters, but optimizes these two separately. By this tradeoff, the computational complexity is reduced. (b) This method weights the frequency spectrum directly in frequency domain rather than in temporal domain. In this way, the complex optimization problem of filter coefficients can be avoided. (c) This method does not need an iterative computation, but just a straightforward computing procedure. This property makes it easy to realize in the future online application, although in this paper only an off-line analysis is done.

The paper is organized as follows: Section 2 describes the CSP algorithm, some existing frequency optimizing methods, and the proposed method; Section 3 describes our experimental setup and the datasets; Section 4 reports the analysis results and makes some discussion; Section 5 concludes the paper.

## 2. Methods

### 2.1. Common Spatial Pattern

CSP is a supervised spatial filtering method for two-class discrimination problems, which finds directions that maximize variance for one class and at the same time minimize variance for the other class. Mathematically, CSP is realized by simultaneous diagonalization of the covariance matrices for the two classes [21].

CSP algorithm is described as follows: Let $\mathbf{X}_d^i$, $d \in \{1, 2\}$ denote the zero-mean EEG recordings of trial $i$, class $d$, and its dimension is $N \times T$, with $N$ the number of channels and $T$ the number of samples in time. The normalized spatial covariance of this trial can be written as

$$\mathbf{R}_d^i = \frac{\mathbf{X}_d^i \mathbf{X}_d^{i\dagger}}{\mathrm{trace}(\mathbf{X}_d^i \mathbf{X}_d^{i\dagger})} \tag{1}$$

where $\dagger$ means the conjugate transpose of a matrix (transpose for real matrix) and trace$(\cdot)$ means the sum of elements on the diagonal of a matrix. For each class, the normalized covariance matrices are averaged over trials,

$$\mathbf{R}_d = \langle \mathbf{R}_d^i \rangle, \quad d \in \{1, 2\} \tag{2}$$

The sum of the two averaged normalized covariances is

$$\mathbf{R}_c = \mathbf{R}_1 + \mathbf{R}_2 \tag{3}$$

$\mathbf{R}_c$ can be decomposed as

$$\mathbf{R}_c = \mathbf{U}_c \mathbf{\Lambda}_c \mathbf{U}_c^\dagger \tag{4}$$

where $\mathbf{U}_c$ is the matrix of eigenvectors and $\mathbf{\Lambda}_c$ is the diagonal matrix of eigenvalues. Then the whitening transformation matrix can be written as

$$\mathbf{P} = \sqrt{\mathbf{\Lambda}_c^{-1}} \mathbf{U}_c^\dagger \tag{5}$$

$\mathbf{R}_c$ can be whitened by $\mathbf{P}$ as follows

$$\mathbf{I} = \mathbf{P}\mathbf{R}_c\mathbf{P}^\dagger \tag{6}$$

where $\mathbf{I}$ denotes the identity matrix. It can be easily seen that, if we transform the $\mathbf{R}_1$ and $\mathbf{R}_2$ as

$$\mathbf{S}_d = \mathbf{P}\mathbf{R}_d\mathbf{P}^\dagger, \quad d \in \{1, 2\} \tag{7}$$

then $\mathbf{S}_1$ and $\mathbf{S}_2$ share common eigenvectors, and the corresponding eigenvalues for the two matrices sum up to 1, i.e., if

$$\mathbf{S}_1 = \boldsymbol{B}\mathbf{\Lambda}_1\mathbf{B}^\dagger \tag{8}$$

then

$$\mathbf{S}_2 = \boldsymbol{B}\mathbf{\Lambda}_2\mathbf{B}^\dagger \quad \text{and} \quad \mathbf{\Lambda}_1 + \mathbf{\Lambda}_2 = \mathbf{I} \tag{9}$$

This means that the eigenvector with largest eigenvalue for $\mathbf{S}_1$ has the smallest eigenvalue for $\mathbf{S}_2$ and vice versa. In this way the two groups are best discriminated. The projection matrix is

$$\mathbf{W} = \mathbf{B}^\dagger\mathbf{P} \tag{10}$$

The dimension of $\mathbf{W}$ is $N \times N$. The rows of $\mathbf{W}$ are called spatial filters, and the columns of $\mathbf{W}^{-1}$ are called spatial patterns. With $\mathbf{W}$, the EEG recordings of trial $i$ can be decomposed as

$$\mathbf{Z}^i = \mathbf{W}\mathbf{X}^i \tag{11}$$

After the decomposition, the components most suitable for discrimination are the first and last few rows of $\mathbf{Z}$. We used the normalized log-variances of these components as features. A Linear Discriminant Analysis (LDA) classifier was used to classify these features.

### 2.2. Existing frequency optimizing methods

#### 2.2.1. Manual selection

In [10], six different bands were studied, namely alpha (8–12 Hz), lower alpha (8–10 Hz), upper alpha (10–12 Hz), beta (19–26 Hz), gamma (38–42 Hz), and broad band (8–30 Hz). It was reported that the broad band (8–30 Hz) showed the best result.

#### 2.2.2. Heuristic selection algorithm

In [5], a heuristic process of frequency band selection was provided. This algorithm gives each band a score for discrimination and the band with the highest score is selected. The score used is the correlation coefficient between band power and the class label. As mentioned in [5], the algorithm works best if only few channels are used.

#### 2.2.3. CSSP

In [17], an extension of CSP named CSSP was proposed and showed an improvement over CSP. This method delays the original signal by a time delay $\tau$, and treats the delayed signal as a new channel, resulting in a doubled number of channels. Then all these channels are put into a CSP procedure. In this way, a simple FIR filter is automatically computed as well as the spatial filters. This method achieves simultaneous optimization of spatial and frequency filters. However, with only one time delay, the flexibility of the frequency filter is very limited. On the other hand, if more than one time delays are used, the number of parameters will also increase, and the optimization problem of the filter coefficients will become more complex. In [17] it was concluded that the choice of only one time delay is the most effective in most situations. Another challenge of this method is to tune the value of time delay $\tau$.

#### 2.2.4. CSSSP

In [18], a method named CSSSP was presented, which allows simultaneously optimizing of an arbitrary FIR filter within the CSP analysis. This method trains an FIR filter with a fixed length $T$, i.e., $T-1$ time delays are used. In order to restrict the complexity of the frequency filter, the solution for the filter coefficients is enforced to be sparse, i.e., only a few non-zero entries. The sparsity of the filter coefficients is controlled by a regularization constant $C$, which has to be tuned by cross validation. The main difficulty in this work is the computational inefficiency of the optimization problem.

#### 2.2.5. SPEC-CSP

In [19,20] a spectrally weighted CSP was proposed. This method is a generalization of the CSP algorithm, incorporating non-homogeneous weighting of the cross-spectrum matrices. This method alternately updates the spectral weights and the spatial projection. The spectral filter optimized from the training data is then applied to the test data. It was reported that this method achieves comparable performance to CSSP and CSSSP, but with far less computational cost. However, the need of extensive parameter tuning and iterative computation limits its real application.

### 2.3. The proposed frequency-weighted method

The proposed FWM is an automatical frequency optimization method, which does not need iterative computing and therefore can be easily realized. In this method the weight vector is computed directly from the training data, but not tuned depending on the classification accuracy. The weight vector is computed in frequency domain rather than in time domain. We will show that this method can improve the classification performance of CSP.

Let $\mathbf{x}$ denote EEG signal in a time window of interest from one channel, and the Fourier transformation of $\mathbf{x}$ is

$$\mathbf{y} = \mathrm{FFT}(\mathbf{x}) \tag{12}$$

where $\mathrm{FFT}(\cdot)$ means the Fourier transformation. The basic idea of FWM is to find a weight vector $\mathbf{w}$, which weights different frequency components according to their importance for the discrimination task:

$$\mathbf{y}_w(k) = \mathbf{w}(k) \times \mathbf{y}(k) \tag{13}$$

where $k$ is the index of frequency. The weighted frequency spectrum $\mathbf{y}_w$ is then transformed back to time domain as

$$\mathbf{x}_w = \mathrm{IFFT}(\mathbf{y}_w). \tag{14}$$

where $\mathrm{IFFT}(\cdot)$ means the inverse Fourier transformation. Then the new signals from different channels are put into CSP for feature extraction. The main work here is to find a suitable weight vector $\mathbf{w}$. In our work, the method used to compute $\mathbf{w}$ is based on Fisher's Linear Discriminative Analysis (LDA) [22], which finds the projecting direction where the generalized Rayleigh quotient of the between-class scatter matrix to the within-class scatter matrix is maximized.

A brief description of the LDA algorithm is given as follows: If $\mathbf{z}_d$, $d \in \{1, 2\}$ is a vector representing one trial from class $d$, and $\mathbf{z}_d$ follows a normal distribution, then let $\bar{\mathbf{z}}_d$ denote the averaged value of $\mathbf{z}_d$ over trials. The scatter matrix of class $d$ is

$$\mathbf{S}_d = \sum (\mathbf{z}_d - \bar{\mathbf{z}}_d)(\mathbf{z}_d - \bar{\mathbf{z}}_d)^\dagger \tag{15}$$

Then the within-class scatter matrix is defined as

$$\mathbf{S}_W = \mathbf{S}_1 + \mathbf{S}_2 \tag{16}$$

and the between-class scatter matrix as

$$\mathbf{S}_B = (\bar{\mathbf{z}}_1 - \bar{\mathbf{z}}_2)(\bar{\mathbf{z}}_1 - \bar{\mathbf{z}}_2)^\dagger \tag{17}$$

In terms of $\mathbf{S}_W$ and $\mathbf{S}_B$, the criterion function $J(\cdot)$ of separability is defined as

$$J(\mathbf{w}) = \frac{\mathbf{w}^\dagger \mathbf{S}_B \mathbf{w}}{\mathbf{w}^\dagger \mathbf{S}_W \mathbf{w}} \tag{18}$$

The $\mathbf{w}$ which maximizes $J(\cdot)$ is the most discriminative direction for the two classes. It can be proven that the $\mathbf{w}$ that maximizes $J(\cdot)$ is

$$\mathbf{w} = \mathbf{S}_W^{-1}(\bar{\mathbf{z}}_1 - \bar{\mathbf{z}}_2) \tag{19}$$

In order to find the most discriminative frequency components, we use the logarithm of amplitude of each frequency component as vector $\mathbf{z}$, i.e.,

$$\mathbf{z} = \log|\mathbf{y}| \tag{20}$$

where $|\cdot|$ means absolute value and $\mathbf{y}$ has the same meaning as in Eq. (12). The transformation to logarithmic values makes the distribution approximately normal. As an example, Fig. 1 shows the

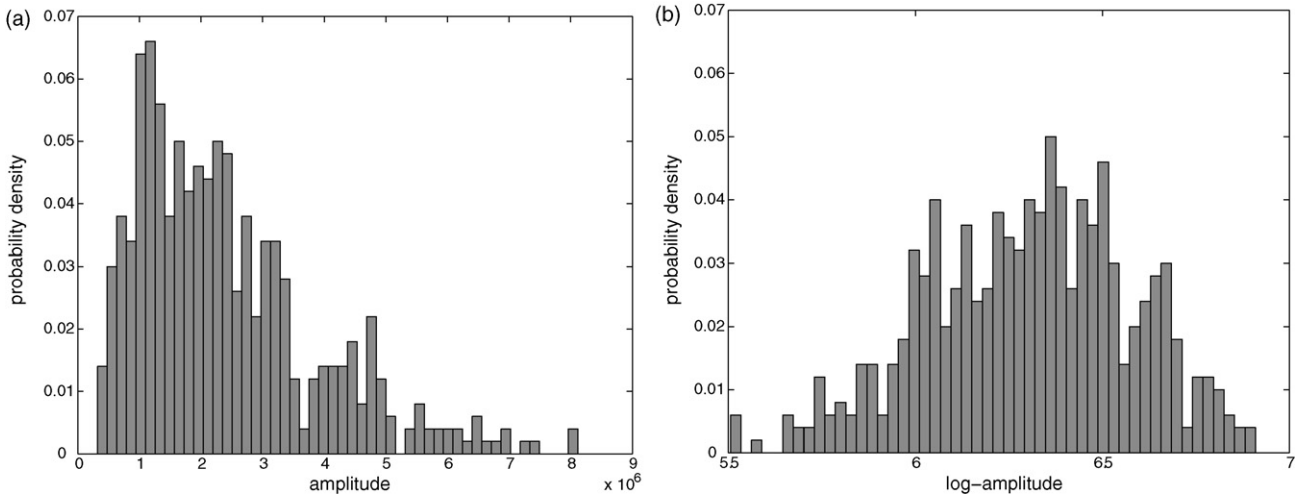


**Fig. 1.** Distribution of (a) the amplitude and (b) log-amplitude at 11 Hz, for one of the subjects in our experiment, where $k$ means kurtosis, and $s$ means skewness. For normal distribution, $k$ should be 3, and $s$ should be 0. From the figure we can see that $k$ and $s$ for (b) are nearer to 3 and 0 than those of (a). In other words, the transformation to logarithmic values makes the distribution approximately normal, which is an important hypothesis in LDA algorithm. (a) $k = 4.16$, $s = 1.14$ and (b) $k = 2.67$, $s = -0.25$.

distribution of the original amplitude and log-amplitude at 11 Hz for one of the subjects in our experiment.

There are some remarks that should be stated: (a) the **w** we find is just a weight vector indicating the importance of different frequency components, but not a projecting direction. This means, only its absolute value is meaningful to us. (b) Theoretically, the correlation coefficient between different frequency components should be zero. Therefore, the scatter matrix $\mathbf{S}_d$ is supposed to be diagonal. However empirical evidence shows that the estimation of $\mathbf{S}_d$ is not diagonal. In computation, we only consider its diagonal elements. (c) Since the frequency spectrum of a real time series is conjugately symmetric around the Nyquist frequency, only the first half of the spectrum is used to calculate the weight vector **w**. (d) Since the frequency spectrum estimated by Fourier transformation appears oscillatory, a smoothing action is taken after the weight vector is computed. The weight values in a 1-Hz window centered at a particular frequency are averaged to get the new weight value for this frequency. For example, the weight values from 9.5 to 10.5 Hz are averaged to get the new weight value for 10 Hz.

## 3. Experiment and dataset

### 3.1. Our experimental setup and the SJTU data

#### 3.1.1. Subjects

Five right-handed subjects (four male and one female, age 23–27 years) took part in the experiment. None of them had an experience of BCI experiment before. The volunteers were paid for their participation.

#### 3.1.2. Procedure

The subjects were seated in a comfortable armchair about 2 m in front of a computer monitor. They were instructed to keep still and avoid blinking during a trial. At the beginning of each trial, i.e., second 0, the screen was blank. At second 1, a fixation cross appeared in the center of the screen. At second 2, an arrow pointing to either left or right was added to the cross indicating the imagination of left or right hand movement. The arrowed cross was shown until second 5. During the time period from second 2 to second 5, the subject had to imagine left or right hand movement according to the cue. The two kinds of movements were decided by the subject herself/himself, e.g., patting a ball or pulling a brake. At second 5, a feedback of this single trial was provided by moving the arrowed cross to the left or right side of the screen, according to the classifier output. After a random interval varying from 1.5 to 2.5 s, the next trial began. The sequence of left and right trials was randomized and the chance for each class was flat. In each run, 10 left and 10 right trials were performed. There were five runs in each session and three sessions for each subject. Runs in one session were performed on the same day. Data in one session were termed as one dataset, including 100 trials (50 left and 50 right). All data were saved for later analysis.

It should be noted that although a feedback was provided in the experiments, the online feedback was computed by a basic CSP in an 8–30 Hz broad band without FWM, and an LDA classifier. In other words, in this paper, we only do an off-line analysis of the proposed method, but do not check its online performance. The basic CSP and the LDA classifier were calibrated before the online experiment on a very short session containing 15 left and 15 right trials. These 30 trials were not provided with feedback, and are not used for analysis in this paper.

#### 3.1.3. Recordings

EEG signals were recorded using a SynAmps system (Neuroscan, USA). Signals from 21 channels over central and related motor areas were used for classification. The grounding electrode was mounted on the forehead and reference electrodes on the left and right mastoids. The electrodes were placed according to the extended 10/20-system [23,24] (see Fig. 2). Horizontal and vertical EOGs were recorded for the purpose of artifact detection, and were not used for classification. The EEG were first filtered by the recording system in a 5–30 Hz frequency band, and the sampling rate was 1000 Hz. Before feature extracting and classifying, the signals were down-sampled to 200 Hz and re-filtered by a 50-order FIR filter in 8–30 Hz frequency band. By the high-pass filtering, low-frequency components of EOG artifacts were also removed [25].

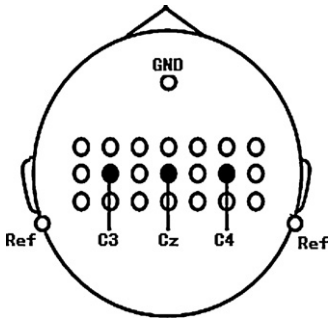


**Fig. 2.** Placement of the 21 EEG electrodes. The three electrodes painted black are C3, Cz, and C4, respectively. 'GND' means ground electrode, and 'Ref' means reference electrodes.

In the remaining part of this paper, we will refer to these data as the SJTU data.

### 3.2. The Graz data

For the purpose of validation, we also applied the proposed method to data set IIa of BCI competition IV, which was provided by Graz University of Technology, Austria. It consists of EEG data from nine subjects, or nine datasets. The task of the experiment is four-class motor imagery, namely, left hand, right hand, both feet, and tongue. Considering that our method is for two-class discrimination, we only use the data of left and right hand.

Details of the data description and experimental setup can be found in [12,13]. In the remaining part of this paper, we will refer to these data as the Graz data.

## 4. Results and discussion

The performance of three methods are compared: (1) FWM combined with CSP (FWM-CSP); (2) CSP in a broad band of 8–30 Hz, without FWM (basic-CSP); (3) CSP in the best frequency band among 49 narrow bands, without FWM (Best-band).

The 49 narrow bands in the Best-band method are: 11 narrow bands between 8 and 30 Hz (width 2 Hz, no overlapping), 10 overlapping bands between 8 and 29 Hz (width 3 Hz, overlapping 1 Hz), 10 overlapping bands between 8 and 30 Hz (width 4 Hz, overlapping 2 Hz), 9 overlapping bands between 8 and 29 Hz (width 5 Hz, overlapping 3 Hz), 9 overlapping bands between 8 and 30 Hz (width 6 Hz, overlapping 4 Hz).

In this work we only do an off-line analysis, and the performances are measured in terms of a 10-fold cross-validation error rate. In each fold of the cross-validation, the weight vector **w** in FWM, the decomposition matrix **W** in CSP, and the linear classifier, are all calculated from the calibration data, and the validation data are treated as unlabelled.

The signals used for calculation were recorded from the time point when the arrow on the cross appeared to the current time. For each trial, a classification output is given at every time point. The lowest cross-validation error rate along the time axis, as well as the integrated error rate along the time axis are reported.

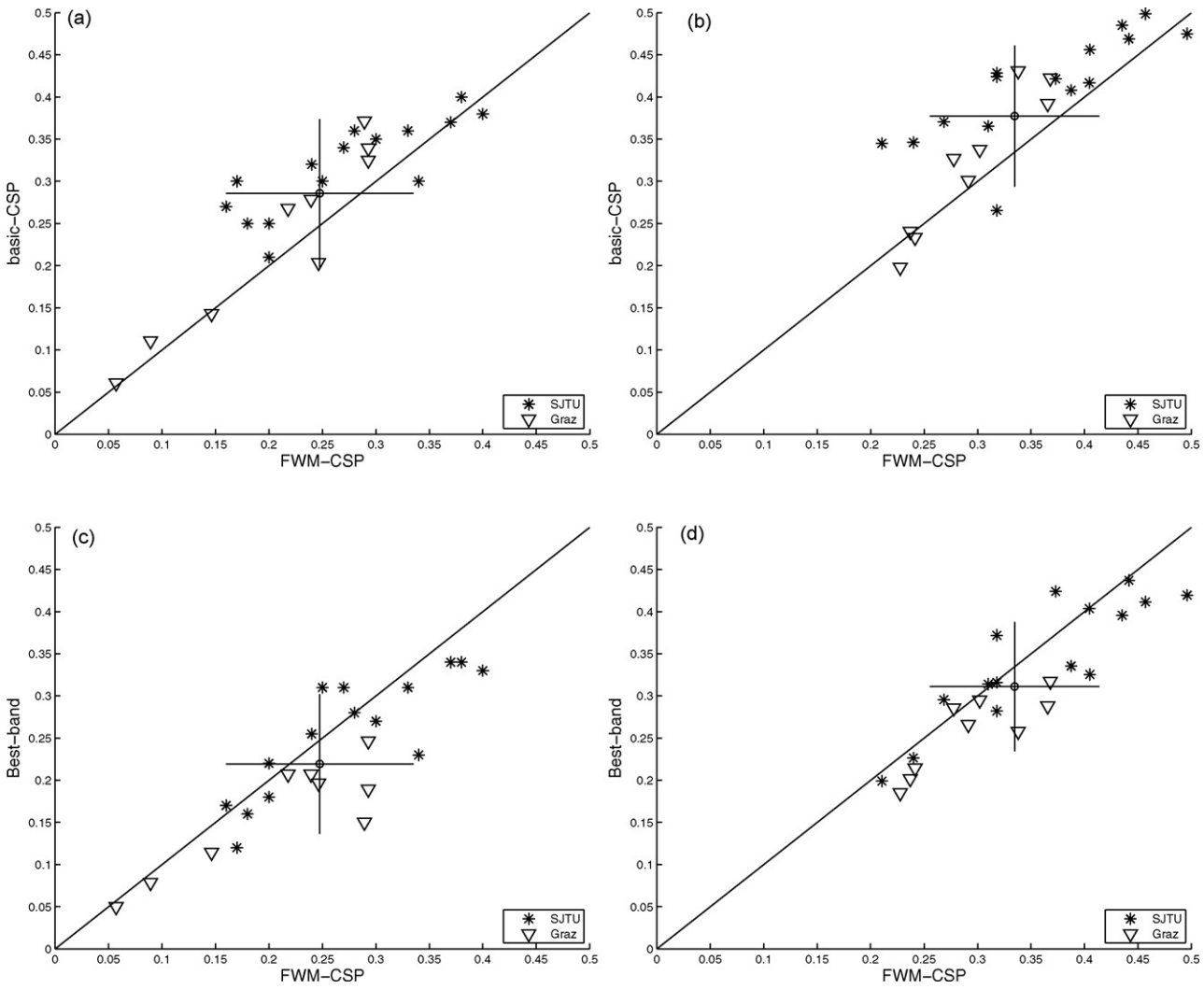


**Fig. 3.** Cross-validation error rates of FWM-CSP versus CSP and best band. (a) FWM-CSP versus basic CSP, lowest error rate; (b) FWM-CSP versus basic CSP, integrated error rate; (c) FWM-CSP versus best band, lowest error rate; (d) FWM-CSP versus best band, integrated error rate. Each point in the plot represents a dataset. A point above the diagonal means a dataset where FWM-CSP outperforms the other method. The circle and cross in the plot indicate the mean and standard deviation of error rates.

### *4.1. Results*

The error rates of FWM-CSP versus the other two methods are shown in Fig. 3. Each point in the plot represents a dataset, with the horizontal axis representing the error rate of FWM-CSP, and the vertical axis representing the error rate of basic-CSP (the first row) or Best-band (the second row). Plots in the left column are the lowest error rates and those in the right column are the integrated error rates. All the error rates are averaged over the 10-fold cross-validation. The mean and standard deviation of error rates over all the datasets, including SJTU data and Graz data, are indicated as a circle and cross in the plot.

From Fig. 3 we can see that, for most datasets, FWM-CSP shows a significant improvement when compared to basic-CSP. The error rate averaged over 24 datasets for FWM-CSP is about 4% lower than those for basic-CSP. However, when compared to the best band found by an enumeration approach, FWM-CSP shows a higher error rate for more than half of the datasets. This means that, the proposed FWM can actually improve the performance of CSP but it still can not achieve the global best level.

Table 1 shows the best frequency bands for each dataset in the SJTU data. As we can see, the best band for most subjects shows a variability between sessions. In other words, the best band optimized from the previous session may be not the best selection for the later session. This makes the band selection work quite difficult in real application.

As an example, Fig. 4 shows the shape of the weight vector between 8 and 30 Hz for each method, data from session 2, subject 4, SJTU data. The curves were averaged over all trials in the session. For basic-CSP, all frequency components get a flat weight. For Best-band, the selected band gets a weight of non-zero while components beyond the band get a weight of zero. For FWM-CSP, different frequency components get a flexible weight according to their separability. Note that for this dataset, the best band selected

**Table 1**
Best frequency band (in Hz) for each dataset (SJTU data). 'sub*i*–*j*' means dataset of session *j*, subject *i*.

| | | | | | | |
|---|---|---|---|---|---|---|
| Data | sub1–1 | sub1–2 | sub1–3 | sub2–1 | sub2–2 | sub2–3 |
| Best band | 18–24 | 12–15 | 8–12 | 26–28 | 8–12 | 14–18 |
| Data | sub3–1 | sub3–2 | sub3–3 | sub4–1 | sub4–2 | sub4–3 |
| Best band | 10–15 | 10–16 | 24–27 | 12–16 | 10–14 | 12–17 |
| Data | sub5–1 | sub5–2 | sub5–3 | | | |
| Best band | 8–13 | 10–15 | 10–16 | | | |

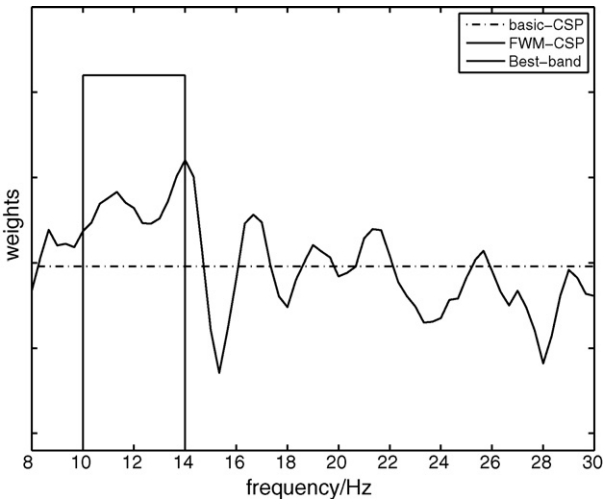


**Fig. 4.** Weight vector got by each method, averaged over 10-fold cross validation, data from C4 channel, session 2, subject 4, SJTU data. For this dataset, the best band is 10–14 Hz, where the weights got by FWM also have a higher value.

is 10–14 Hz (see Table 1), and the weight vector estimated by FWM also has a higher value in this band. In Fig. 4 the vertical axis does not have a scale because for a weight vector the only thing that makes sense is its shape.

### 4.2. Discussion

As mentioned before, the proposed method does not achieve the global best level. There may be several reasons for this: (a) the estimation of frequency spectrum based on Fourier transformation is rather oscillatory, which influences the estimation precision of the weight vector **w**. (b) Although the logarithm transformation makes the distribution of amplitude approximately normal, with limited trial samples, the real distribution is actually not normal, which can make the LDA algorithm not optimal. (c) With this FWM, the optimization of frequency is independent from the optimization of spatial projection, i.e., it is not a simultaneous optimization. This is a tradeoff we made between classification accuracy and computational cost.

As shown in Fig. 3, the best narrow band achieves better result than the 8–30 Hz broad band, which disagrees with the result reported in [10]. The reason for this disagreement may be that, in [10] the five narrow bands were selected empirically and kept unchanged, whereas in this paper the narrow band was selected among 49 narrow bands for each subject and each dataset.

## 5. Conclusion

Optimization of EEG in spatial and frequency domain is important for ERD/ERS based BCI systems. CSP is a powerful and successful algorithm for the optimization in spatial domain. However, existing literature on the optimization of frequency combined with CSP is limited. In this paper, we proposed a frequency optimization method named FWM which automatically and flexibly optimizes the frequency components before CSP is applied to the EEG signals. We investigated the performance of this method on both EEG data recorded in our laboratory and data set IIa from BCI competition IV. The results show that the proposed method achieves a considerable improvement over the basic CSP applied on a broad band. On the other hand, the proposed method does not achieve the global optimal level when compared with the best narrow band found by an enumeration approach. However, the best band for one subject is not constant, therefore the selection work is very difficult. In this work only off-line analysis is done, our future work will focus on the online application of the proposed method.

## Acknowledgments

This work was jointly supported by National Natural Science Foundation (Grant No. 50525517) and Science and Technology Commission of Shanghai Municipality (Grant Nos. 08JC1412100 and 09JC1408400). The authors would like to thank Dr. Dingguo Zhang for his generous help.

## References

[1] J. Wolpaw, N. Birbaumer, D. McFarland, G. Pfurtscheller, T. Vaughan, Brain–computer interfaces for communication and control, Clinical Neurophysiology 113 (6) (2002) 767–791.
[2] G. Pfurtscheller, A. Aranibar, Evaluation of event-related desynchronization (ERD) preceding and following voluntary self-paced movement, Electroencephalography and Clinical Neurophysiology 46 (2) (1979) 138–146.
[3] G. Pfurtscheller, C. Brunner, A. Schlögl, F. Lopes da Silva, Mu rhythm (de) synchronization and EEG single-trial classification of different motor imagery tasks, NeuroImage 31 (1) (2006) 153–159.
[4] G. Pfurtscheller, C. Neuper, D. Flotzinger, M. Pregenzer, EEG-based discrimination between imagination of right and left hand movement, Electroencephalography and Clinical Neurophysiology 103 (6) (1997) 642–651.
[5] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, K. Muller, Optimizing spatial filters for robust EEG single-trial analysis, IEEE Signal Processing Magazine 25 (1) (2008) 41–56.
[6] D. McFarland, L. Miner, T. Vaughan, J. Wolpaw, Mu and beta rhythm topographies during motor imagery and actual movements, Brain Topography 12 (3) (2000) 177–186.
[7] G. Pfurtscheller, C. Neuper, C. Guger, W. Harkam, H. Ramoser, A. Schlogl, B. Obermaier, M. Pregenzer, Current trends in Graz brain–computer interface (BCI) research, IEEE Transactions on Rehabilitation Engineering 8 (2) (2000) 216–219.
[8] B. Blankertz, G. Dornhege, S. Lemm, M. Krauledat, G. Curio, K. Müller, The Berlin brain–computer interface: machine learning based detection of user specific brain states, Journal of Universal Computer Science 12 (6) (2006) 581–607.
[9] J. Wolpaw, D. McFarland, T. Vaughan, Brain–computer interface research at the Wadsworth Center, IEEE Transactions on Rehabilitation Engineering 8 (2) (2000) 222–226.
[10] J. Müller-Gerking, G. Pfurtscheller, H. Flyvbjerg, Designing optimal spatial filters for single-trial EEG classification in a movement task, Clinical Neurophysiology 110 (5) (1999) 787–798.
[11] H. Ramoser, J. Muller-Gerking, G. Pfurtscheller, Optimal spatial filtering of single trial EEG during imagined handmovement, IEEE Transactions on Rehabilitation Engineering 8 (4) (2000) 441–446.
[12] M. Naeem, C. Brunner, R. Leeb, B. Graimann, G. Pfustscheller, Separability of four-class motor imagery data using independent component analysis, Journal of Neural Engineering 3 (2006) 208–216.
[13] 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 Recognition Letters 28 (8) (2007) 957–964.
[14] Z. Koles, J. Lind, A. Soong, Spatio-temporal decomposition of the EEG: a general approach to the isolation and localization of sources, Electroencephalography and Clinical Neurophysiology 95 (4) (1995) 219–230.
[15] C. Andrew, G. Pfurtscheller, On the existence of different alpha band rhythms in the hand area of man, Neuroscience Letters 222 (2) (1997) 103–106.
[16] G. Pfurtscheller, A. Stancák, G. Edlinger, On the existence of different types of central beta rhythms below 30 Hz, Electroencephalography and Clinical Neurophysiology 102 (4) (1997) 316–325.
[17] S. Lemm, B. Blankertz, G. Curio, K. Muller, Spatio-spectral filters for improving the classification of single trial EEG, IEEE Transactions on Biomedical Engineering 52 (9) (2005) 1541–1548.
[18] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, K. Muller, Combined optimization of spatial and temporal filters for improving brain–computer interfacing, IEEE Transactions on Biomedical Engineering 53 (11) (2006) 2274–2281.
[19] R. Tomioka, G. Dornhege, G. Nolte, K. Aihara, K. Muller, Optimizing spectral filters for single trial EEG classification, Lecture Notes in Computer Science 4174 (2006) 414–423.
[20] R. Tomioka, G. Dornhege, G. Nolte, B. Blankertz, K. Aihara, K. Müller, Spectrally Weighted Common Spatial Pattern Algorithm for Single Trial EEG Classification, Department of Mathematical Engineering, University of Tokyo, Tokyo, Japan, Technical Reports, vol. 40, 2006.

[21] K. Fukunaga, Introduction to Statistical Pattern Recognition, Academic Press, 1990.
[22] R. Duda, P. Hart, D. Stork, Pattern classification, 2nd edn, Wiley, New York, 2001.
[23] H. Jasper, The ten-twenty electrode system of the international federation in electroencephalography and clinical neurophysiology, EEG Journal 10 (1958) 371–375.
[24] R. Oostenveld, P. Praamstra, The five percent electrode system for high-resolution EEG and ERP measurements, Clinical Neurophysiology 112 (4) (2001) 713–719.
[25] M. Fatourechi, A. Bashashati, R. Ward, G. Birch, EMG and EOG artifacts in brain computer interface systems: a survey, Clinical Neurophysiology 118 (3) (2007) 480–494.