Proceedings of the 2005 IEEE
Engineering in Medicine and Biology 27th Annual Conference
Shanghai, China, September 1-4, 2005

Common Spatial Pattern Method for Channel Selection in Motor Imagery Based Brain-computer Interface

Yijun Wang, Shangkai Gao, Xiaorong Gao
Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing, China

Abstract—**A brain-computer interface (BCI) based on motor imagery (MI) translates the subject's motor intention into a control signal through classifying the electroencephalogram (EEG) patterns of different imagination tasks, e.g. hand and foot movements. Characteristic EEG spatial patterns make MI tasks substantially discriminable. Multi-channel EEGs are usually necessary for spatial pattern identification and therefore MI-based BCI is still in the stage of laboratory demonstration, to some extent, due to the need for constantly troublesome recording preparation. This paper presents a method for channel reduction in MI-based BCI. Common spatial pattern (CSP) method was employed to analyze spatial patterns of imagined hand and foot movements. Significant channels were selected by searching the maximums of spatial pattern vectors in scalp mappings. A classification algorithm was developed by means of combining linear discriminant analysis towards event-related desynchronization (ERD) and readiness potential (RP). The classification accuracies with four optimal channels were 93.45% and 91.88% for two subjects.**

Keywords—**Brain-computer interface, motor imagery, common spatial pattern, channel selection**

I. Introduction

Brain-computer interfaces (BCIs) translate brain signals into a control signal without using muscles or peripheral nerves. Among a variety of noninvasive BCI methods, electroencephalogram (EEG) method has short time constants, less environmental limits, and requires inexpensive equipment. Therefore, most practical BCI systems use EEG signals as inputs. Visual evoked potentials , P300 evoked potentials , slow cortical potentials, event-related synchronization and desynchronization, and mu and beta rhythms are widely used signals in EEG-based BCIs [1][2].

In recent years, BCI systems based on classifying single trial EEG of motor imagery have been developed rapidly. The physiological studies on motor imagery indicate that the spatial distribution of EEG differs between different imagined movements. First, brain oscillations at mu (8-12Hz) and beta (18-26Hz) rhythms display specific areas of event-related desynchronization (ERD) corresponding to each imagery state. ERD represents the changes of the ongoing EEG activity characterized by decrease of power in the given frequency bands. While the ERD of imagined hand movements appears over somatosensory areas controlateral to the movement, the foot area of ERD localizes on the central area between both hemispheres [3]. On the other hand, lateral readiness potential (LRP), which is one slowly decreasing potential, can be recorded with the maximum amplitude over the motor cortex contralateral to the involved hand movements, whereas readiness potentials preceding foot movements show no lateralization [4].

G. Pfurtscheller et al. first used EEG classification based on ERD during imagined motor actions for a BCI application [5]. An adaptive autoregressive (AAR) model was used for feature extraction. H. Ramoser et al. designed optimal spatial filters by the method of common spatial pattern (CSP) for filtering single-trial EEG during imagined hand movements [6]. B. Blankertz et al. significantly boosted MI-based BCI performances through the combination of classifiers towards different physiological phenomena, i.e. ERD and LRP [7].

So far, MI-based BCI is an independent system with higher classification accuracy. For example, the average accuracy of classifying imagined left and right hand movements was above 90% [6][7]. However, MI-based BCI is still not used for real-life application. One important factor is that most available algorithms focus on analyzing multi-channel EEG, which needs relatively expensive equipment, inconvenient recording preparation, and complicated calculation. To design a practical MI-based BCI, one way is to select fewer channels for application.

In this paper, we present a method for channel selection in MI-based BCI. Optimal channels were selected by searching the maximums of spatial pattern vectors derived from CSP method. And towards the few-channel EEG classification, a simplified algorithm fit for on-line application was proposed and validated.

II. Methodology

A. Data description

The data sets were provided by Fraunhofer FIRST, Intelligent Data Analysis Group, and Campus Benjamin Franklin of the Charité - University Medicine Berlin (http://ida.first.fraunhofer.de/projects/bci/competition_iii/desc_IVa.html). The aim was to discriminate trials of different motor imagery tasks. EEG data were recorded from two healthy subjects. Subjects sat in a comfortable chair with arms resting on armrests. Visual cues appeared on a computer screen for 3.5 seconds. During this period, the subjects were performing three motor imagery tasks: left hand, right hand, and right foot. The presentation of target cues was intermitted by periods of random length (1.75-2.25 seconds) in which the subject could relax. Only trials for the classes "right hand" and "right foot" were provided for analysis. The data set for each subject consists of 160 trials (80 hand and 80 foot).

The recording was made using a 128 channel Ag/AgCl electrode cap. 118 EEG channels were measured at positions of the extended international 10/20-system. Signals were band-pass filtered between 0.05 and 200 Hz and then digitized at 1000 Hz. All signals were down-sampled to 100Hz for analysis.

B. Data preprocessing

Data preprocessing includes time window intercepting and band-pass temporal filtering. The prior knowledge that LRP appears in the lower frequency band and ERD occurs in mu and beta frequency bands can be utilized to make initial estimates of the parameters. In practice, the parameters were determined through time-frequency analysis on contrast of EEG power between two classes.

C. Common Spatial patterns

Common spatial patterns (CSP) method was firstly suggested for classification of multi-channel EEG during imagined hand movements by H. Ramoser [6]. The main idea is to use a linear transform to project the multi-channel EEG data into low-dimensional spatial subspace with a projection matrix, of which each row consists of weights for channels. This transformation can maximize the variance of two-class signal matrices. CSP method is based on the simultaneous diagonalization of the covariance matrices of both classes.

Details of the algorithm are described as follows with the example of classifying single-trial EEG during hand and foot movements. $\boldsymbol{X}_{\mathrm{H}}$ and $\boldsymbol{X}_{\mathrm{F}}$ denote the preprocessed EEG matrices under two conditions (hand and foot) with dimensions N×T, where N is the number of channels and T is the number of samples per channel. The normalized spatial covariance of the EEG can be represented as:

$$R_{\mathrm{H}} = \frac{X_{\mathrm{H}} X_{\mathrm{H}}^{T}}{\mathrm{trace}(X_{\mathrm{H}} X_{\mathrm{H}}^{T})} \qquad R_{\mathrm{F}} = \frac{X_{\mathrm{F}} X_{\mathrm{F}}^{T}}{\mathrm{trace}(X_{\mathrm{F}} X_{\mathrm{F}}^{T})} \; . \tag{1}$$

$\boldsymbol{X}^{\mathrm{T}}$ is the transpose of $\boldsymbol{X}$ and trace($\boldsymbol{A}$) computes the sum of the diagonal elements of $\boldsymbol{A}$. The averaged normalized covariance $\overline{R_{\mathrm{H}}}$ and $\overline{R_{\mathrm{F}}}$ are calculated by averaging over all the trials of each group. The composite spatial covariance can be factorized as

$$R = \overline{R_{\mathrm{H}}} + \overline{R_{\mathrm{F}}} = U_0 \Sigma U_0^{T} \tag{2}$$

where $\boldsymbol{U}_0$ is the matrix of eigenvectors and $\boldsymbol{\Sigma}$ is the diagonal matrix of eigenvalues. The whitening transformation matrix

$$P = \Sigma^{-1/2} U_0^{T} \tag{3}$$

transforms the average covariance matrices as

$$S_{\mathrm{H}} = P \overline{R_{\mathrm{H}}} P^{T} \qquad S_{\mathrm{F}} = P \overline{R_{\mathrm{F}}} P^{T} \; . \tag{4}$$

$\boldsymbol{S}_{\mathrm{H}}$ and $\boldsymbol{S}_{\mathrm{F}}$ share common eigenvectors and the sum of corresponding eigenvalues for the two matrices will always be one,

$$S_{\mathrm{H}} = U \Sigma_{\mathrm{H}} U^{T} \quad S_{\mathrm{F}} = U \Sigma_{\mathrm{F}} U^{T} \quad \Sigma_{\mathrm{H}} + \Sigma_{\mathrm{F}} = I \; . \tag{5}$$

The eigenvectors with the largest eigenvalues for $\boldsymbol{S}_{\mathrm{H}}$ have the smallest eigenvalues for $\boldsymbol{S}_{\mathrm{F}}$ and vice versa. The transformation of whitened EEG onto the eigenvectors corresponding to the largest eigenvalues in $\boldsymbol{\Sigma}_{\mathrm{H}}$ and $\boldsymbol{\Sigma}_{\mathrm{F}}$ is optimal for separating variance in two signal matrices. The projection matrix $\boldsymbol{W}$ is denoted as

$$W = U^{T} P \; . \tag{6}$$

With the projection matrix $\boldsymbol{W}$, the original EEG can be transformed into uncorrelated components

$$Z = WX \; . \tag{7}$$

$\boldsymbol{Z}$ can be seen as EEG source components including common and specific components of different tasks. The original EEG $\boldsymbol{X}$ can be reconstructed by

$$X = W^{-1} Z \tag{8}$$

where $\boldsymbol{W}^{-1}$ is the inverse matrix of $\boldsymbol{W}$. The columns of $\boldsymbol{W}^{-1}$ are spatial patterns, which can be considered as EEG source distribution vectors. The first and last columns of $\boldsymbol{W}^{-1}$ are the most important spatial patterns that explain the largest variance of one task and the smallest variance of the other.

D. Channel selection

T. N. Lal et al. used the feature selection algorithms based on the training of support vector machines for the purpose of selecting channels for classifying EEG during mental tasks [8]. Their studies showed that the number of channels could be reduced significantly without decreasing the classification rate, and an appropriate number of channels could have nearly the same accuracy as that acquired by all channels. By reducing the number of channels, the recording preparation can be more convenient, as well as the calculation time can be enormously decreased.

Spatial patterns derived from CSP method can be seen as EEG source distribution vectors. We assume that only the first and last spatial patterns are related with the specific sources of the two tasks. Then, the two channels corresponding to the maximal coefficients of spatial pattern vectors may be the channels most correlated with the task specific sources. We applied this idea to search for two pairs of optimal channels, corresponding to CSP analyses on ERD and RP respectively. The resulting channels agreed well with the expected underlying cortical activity distributions during the MI tasks.

Fig.1(a) displays the two most important spatial patterns based on ERD of one subject. The ERD of imagined right hand movements appears dominantly over left hemisphere, the foot area localizes on the central area between both hemispheres. Therefore, right hand imagery causes a relatively increased EEG power over the foot area, and foot imagery has an increased EEG variance over the right hand area. That is to say, the spatial pattern of one task essentially presents the ERD distribution of the other task. Fig.1(b) shows the spatial patterns derived from RP, which are obviously different from those of ERD. To some extent, it suggests that ERD and RP are independent physiological phenomena with different spatial distributions.

The spatial pattern of hand imagery is given as $\boldsymbol{SP}_{\mathrm{H}}$, and $\boldsymbol{SP}_{\mathrm{F}}$ for foot. The optimal channels can be determined through searching the maximums of the absolute value of $\boldsymbol{SP}_{\mathrm{H}}$ and $\boldsymbol{SP}_{\mathrm{F}}$:

$$\mathrm{CH_H}=\mathrm{find}(|\boldsymbol{SP}_{\mathrm{H}}|==\mathrm{Max}(|\boldsymbol{SP}_{\mathrm{H}}|))$$

$$\mathrm{CH_F}=\mathrm{find}(|\boldsymbol{SP}_{\mathrm{F}}|==\mathrm{Max}(|\boldsymbol{SP}_{\mathrm{F}}|))$$

where find() finds indices of elements. For this subject, the optimal channels for ERD are C3 and CPz, and Cz and FCz for RP (See Fig.1). These four channels are used for feature extraction and classification procedures.

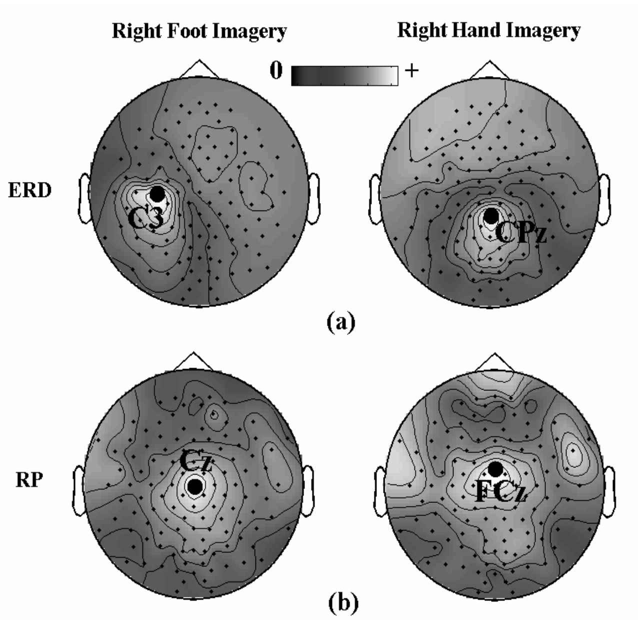

Fig.1. Spatial mappings of the two spatial patterns corresponding to right hand and foot imagery tasks, based on (a) ERD and (b) RP respectively. The thick dot " • " indicates the selected optimal channel.

E. Feature extraction

Average time course features over all the trials of each group can contribute to considering the feature extraction method. Readiness potential is one component of movement-related potentials (MRPs), and it can be extracted by averaging techniques. However, ERD is not phased-locked and thus will be eliminated by averaging [3]. Fortunately, it can be detected by power analysis. Time course ERD can be presented as power samples obtained through calculating the absolute value of the amplitude samples.

As shown in Fig.2, average time course ERDs on the optimal channels (C3 and CPz) display different patterns under two conditions. On C3, hand ERD shows much lower amplitude. But on CPz, foot ERD is slightly lower. Fig.3 presents average RPs on Cz and FCz. Amplitude of foot RP is lower than hand RP on both channels. It may suggest that RP of foot imagery is more significant than that of hand imagery.

The amplitude differences of time course ERD and RP between hand and foot movements ensure the substantial discriminability of these two states. Features corresponding to ERD and RP can be defined as [9][10]:

$$f_1^{\mathrm{ERD}}=\mathrm{mean}(|\boldsymbol{x}_{\mathrm{C3}}^{\mathrm{ERD}}|) \qquad f_2^{\mathrm{ERD}}=\mathrm{mean}(|\boldsymbol{x}_{\mathrm{CPz}}^{\mathrm{ERD}}|)$$

$$f_1^{\mathrm{RP}}=\mathrm{mean}(\boldsymbol{x}_{\mathrm{Cz}}^{\mathrm{RP}}) \qquad f_2^{\mathrm{RP}}=\mathrm{mean}(\boldsymbol{x}_{\mathrm{FCz}}^{\mathrm{RP}})$$

where $\boldsymbol{x}$ denotes preprocessed EEG signal of one channel and mean($\boldsymbol{x}$) calculates the mean value of $\boldsymbol{x}$.

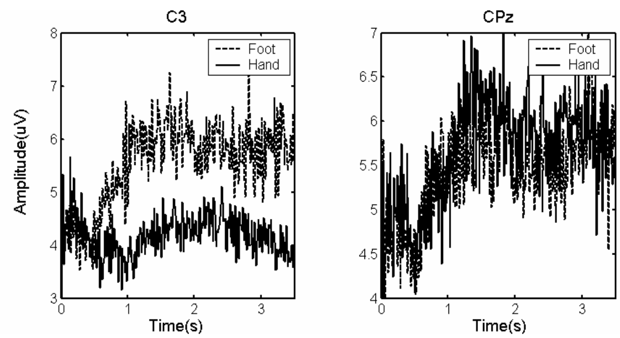

Fig.2. Average time course ERDs on C3 and CPz. "--" denotes foot imagery and "—" denotes hand imagery.

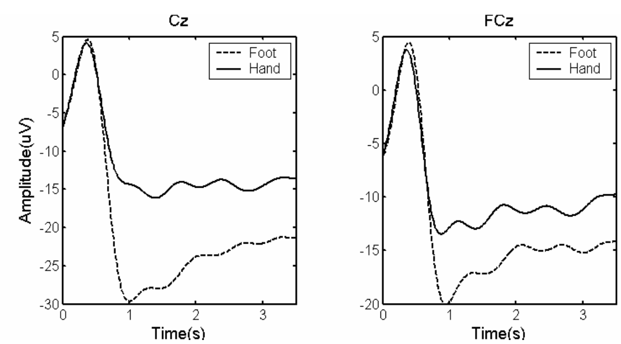

Fig.3. Average time course RPs on Cz and FCz. "--" denotes foot imagery and "—" denotes hand imagery.

F. Classification and cross-validation

A well known linear classification method, Fisher Discriminant (FD), is used as the classifier. The feature vectors $\boldsymbol{p}^{\mathrm{ERD}} = [f_1^{\mathrm{ERD}}\ f_2^{\mathrm{ERD}}]^{\mathrm{T}}$ and $\boldsymbol{p}^{\mathrm{RP}} = [f_1^{\mathrm{RP}}\ f_2^{\mathrm{RP}}]^{\mathrm{T}}$ are defined for training of two classifiers. The outputs are calculated as:

$$a^{\mathrm{ERD}}=(\boldsymbol{W}^{\mathrm{ERD}})^{\mathrm{T}}\boldsymbol{p}^{\mathrm{ERD}}+b^{\mathrm{ERD}} \qquad a^{\mathrm{RP}}=(\boldsymbol{W}^{\mathrm{RP}})^{\mathrm{T}}\boldsymbol{p}^{\mathrm{RP}}+b^{\mathrm{RP}}$$

where $\boldsymbol{W}$ and b are the weights and offset determined by the training data. a^{ERD} and a^{RP} denote the classification results based on ERD and RP respectively. The final decision is made by summarizing the two weighted outputs as:

$$a = \mathrm{sgn}(w^{\mathrm{ERD}}a^{\mathrm{ERD}}+w^{\mathrm{RP}}a^{\mathrm{RP}}).$$

The weights w^{ERD} and w^{RP} are defined as:

$$w^{ERD}=(2acc^{ERD}-1)^{m} \qquad w^{RP}=(2acc^{RP}-1)^{m}$$

where *acc* is the classification accuracy obtained from the training stage and *m* is a control parameter used to emphasize the feature with higher accuracy.

A 10×10-fold cross-validation is used to estimate the classification accuracy.

III. Results

The classification accuracy rates of two subjects are listed in Table 1, which consists of three parts. First, “4-chn” refers to only making use of the four optimal channels to do classification. The combination of ERD and RP resulted in a high accuracy for the two subjects (93.45% and 91.88%). Second, “8-chn” denotes applying the 8 optimal channels (including additive 4 channels corresponding to the second maximums of spatial pattern vectors) for classification. The additive channels contributed to increase the classification accuracy of some percent, but decreasing the convenience of system operation at the same time. For these two subjects, 8-channel results have already been very close to those of much more channels. Finally, “s-to-s” means classification using subject-to-subject 4-channel configuration. For subject S2, the accuracy decreased from 91.88% to 73.38%, displaying the advantage of subject specific channel selection.

TABLE I
CLASSIFICATION ACCURACY RATES (%) OF 2 SUBJECTS

Subjects	**Accuracy ±std (%)**		
4-chn	ERD	RP	Combined
S1	90.14±0.22	88.41±0.32	93.45±0.23
S2	84.69±0.61	84.00±0.32	91.88±0.66
8-chn	ERD	RP	Combined
S1	91.82±0.37	92.32±0.45	96.68±0.31
S2	85.06±0.69	87.06±0.66	93.25±0.49
s-to-s	ERD	RP	Combined
S1	84.09±0.48	77.23±0.34	88.18±0.21
S2	71.13±0.57	64.12±1.15	73.38±0.73

IV. Discussion

ERD and RP are independent in frequency domain, so the combination algorithm can lead to a significant performance gain. Moreover, different feature extraction methods may also result in relatively independent features for the same physiological background. From this point of view, the other methods such as an autoregressive model (AR) should be combined to do the classification [7].

Summarizing, we have developed an effective channel selection approach based on common spatial patterns method. The feature extraction and classification method based on linear discriminant analysis may be applicable for on-line analysis with an easy calculation. The present method may help to construct a practical MI-based BCI. In addition, continuous classifier outputs could be easily used as the feedback to realize an asynchronous BCI, which allows the user to operate the system at any moment [11].

Acknowledgment

This project is supported by National Natural Science Foundation of China (#60318001) and the Key Project of Chinese Ministry of Education (NO.1041185). And we are grateful to Klaus-Robert Müller, Benjamin Blankertz and Gabriel Curio for providing their data.

References

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, T. M. Vaughan, “Brain-computer interfaces for communication and control,” *Clin. Neurophysiol.*, vol.113, pp.767-791, 2000.

[2] B. Blankertz, K. R. Müller, G. Curio, T. M. Vaughan, G. Schalk, J. R. Wolpaw, A. Schloegl, C. Neuper, G. Pfurtscheller, T. Hinterberger, M. Schroeder, and N. Birbaumer, “The BCI Competition 2003: progress and perspectives in detection and discrimination of EEG single trials”, *IEEE Trans. Biomed. Eng.*, vol.51, no.6, pp.1044-1050, 2004.

[3] G. Pfurtscheller, and F. H. Lopes da Silva, “Event-related EEG/MEG synchronization and desynchronization: basic principles,” *Clin. Neurophysiol.*, vol.110, pp.1842-1857, 1999.

[4] K. B. E. Böcker, C. H. M. Brunia, and P. J. M. Cluitmans, “A spatio-temporal dipole model of the readiness potential in humans. II. Foot movement,” *Electroencephalogr. Clin. Neurophysiol.*, vol.91, pp.286-294, 1994.

[5] G. Pfurtscheller, C. Neuper, A. Schlogl, and K. Lugger, “Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters,” *IEEE Trans. Rehab. Eng.*, vol.6, no.3, pp.316-325, 1998.

[6] H. Ramoser, J. M. Gerking, and G. Pfurtscheller, “Optimal spatial filtering of single trial EEG during imagined hand movement,” *IEEE Trans. Rehab. Eng.*, vol. 8, no.4, pp.441-446, 2000.

[7] G. Dornhege, B. Blankertz, G. Curio, and K. R. Müller, “Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms,” *IEEE Trans. Biomed. Eng.* vol.51, no.6, pp.993-1002, 2004.

[8] T. N. Lal, M. Schroder, T. Hinterberger, J. Weston, M. Bogdan, N. Birbaumer, and B. Scholkopf, “Support vector channel selection in BCI,” *IEEE Trans. Biomed. Eng.* vol.51, no.6, pp.1003-1010, 2004.

[9] B. Blankertz, G. Curio, and K. R. Müller, “Classifying single trial EEG: towards brain computer interfacing,” In: T. G. Diettrich, S. Becker, and Z. Ghahramani (eds.), *Advances in Neural Inf. Proc. Systems* (*NIPS* 01), vol.14, pp.157-164, 2002.

[10] Y. Wang, Z. Zhang, Y. Li, X. G, S. G and F. Y, “BCI Competition Data Set IV: an algorithm based on CSSD and FDA for classifying single trial EEG”, *IEEE Trans. Biomed. Eng.*, vol.51, no.6, pp.1081-1086, 2004.

[11] G. Townsend, B. Graimann, and G. Pfurtscheller, “Continuous EEG classification during motor imagery-simulation of an asynchronous BCI,” *IEEE Trans. Rehab. Eng.*, vol. 12, no.2, pp.258-265, 2004.