Progress in Natural Science
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713724104

Feature combination for classifying single-trial ECoG during motor imagery of different sessions

Wei Quigguo [ab]; Meng Fei [a]; Wang Yijun [a]; Gao Xiaorong [a]; Goa Shangkai [a]
[a] Department of Biomedical Engineering, School of Medicine, TsinghuaUniversity, Beijing, China
[b] Department of Electronic Engineering, School of Information, Nanchang University, Nanchang, China

Online Publication Date: 01 July 2007
To cite this Article: Quigguo, Wei, Fei, Meng, Yijun, Wang, Xiaorong, Gao and Shangkai, Goa (2007) 'Feature combination for classifying single-trial ECoG during motor imagery of different sessions ', Progress in Natural Science, 17:7, 851 - 858
To link to this article: DOI: 10.1080/10002007088537482
URL: http://dx.doi.org/10.1080/10002007088537482

Feature combination for classifying single-trial ECoG during motor imagery of different sessions*

Wei Qingguo[1,2], Meng Fei[1], Wang Yijun[1], Gao Xiaorong[1] and Gao Shangkai[1]**

(1. Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing 100084, China; 2. Department of Electronic Engineering, School of Information, Nanchang University, Nanchang 330029, China)

Accepted on December 4, 2006

Abstract The input signals of brain-computer interfaces (BCIs) may be either scalp electroencephalogram (EEG) or electrocorticogram (ECoG) recorded from subdural electrodes. To make BCIs practical, the classifiers for discriminating different brain states must have the ability of session-to-session transfer. This paper proposes an algorithm for classifying single-trial ECoG during motor imagery of different sessions. Three features, derived from two physiological phenomena, movement-related potentials (MRP) and event-related desynchronization (ERD), and extracted by common spatial subspace decomposition (CSSD) and waveform mean, are combined to perform classification tasks. The specific signal processing methods utilized are described in detail. The algorithm was successfully applied to Data Set I of BCI Competition III, and achieved a classification accuracy of 91% on test set.

Keywords: brain-computer interface (BCI), electrocorticogram (ECoG), session-to-session transfer, feature combination, movement-related potentials (MRP), event-related desynchronization (ERD).

A brain-computer interface (BCI) is a communication system that does not depend on the normal output pathways consisting of periphery nerves and muscles[1]. BCI transforms mental intentions into control commands by analyzing the bioelectrical brain activity. The technique can help patients totally losing volitional motor ability but having intact cognition (e.g., people with amyotrophic lateral sclerosis, cerebral palsy or locked-in syndrome) and realizing the control of external facilities, motor neuroprosthesis, wheelchair, and so on, to improve their living quality[2].

In the past decade BCI technology has got developed rapidly. However, most human BCI research focused on electroencephalographic (EEG) recordings. Due to the low signal-to-noise ratio of EEG, the classification accuracy and consequent information transfer rate of EEG-based BCIs are still low. One method to boost classification accuracy is to improve the quality of input signal of a BCI system. Compared to EEG data, electrocorticographic (ECoG) recordings, derived from surface of the cortex, have the advantages of higher signal-to-noise ratio and better spatial resolution, and thus may be used as a feasible alternative of BCI signal source[3,4].

Some researchers have already been interested in ECoG-based human BCIs. Lal et al. investigated BCIs based on ECoG data. By developing a machine learning method for feature and channel selection, they achieved a classification accuracy of 82.5% for two cognitive tasks on a subject using 1.5s data with only two channels[3]. Leuthardt and his colleagues demonstrated for the first time that ECoG activity can enable users to control a one-dimensional computer cursor rapidly and accurately. In addition, they were the first to identify ECoG signals that were related to different kinds of motor and speech imagery[4].

To promote the development and practicality of ECoG-based BCIs, the organizers of BCI Competition III provided an ECoG data of imagined movement of left small finger or tongue in which the training set and test set were recorded in two different sessions, and required the contributing algorithms to have the ability of session-to-session transfer, i.e. a classifier that was trained on the first day can correctly classify data recorded during following days[5]. Since the subject might be in a different state concerning motivation, fatigue etc., the subject's brain will show different electric activity. This brings about the difficulty in designing classifiers and therefore is the thrill in Data Set I of the BCI Competition III.

* Supported by National Natural Science Foundation of China (Grant No. 60318001) and Beijing Natural Science Foundation (Grant No. 3051001)
** To whom correspondence should be addressed. E-mail: gsk-dea@tsinghua.edu.cn

1 Methodology

It has been shown that the execution, preparation and even imagination of movements result in similar changes of brain state[6,7]. Movement-related potentials (MRP) and event-related desynchronization (ERD) are two basic electrical physiological phenomena that are activated by limb movement or imagined movement. MRP and ERD reflect different aspects of sensorimotor cortical processes, while recordings from subdural electrodes demonstrate behavior analogous to EEG data[8,9]. These electrophysiological features can be used as the basis of classifying ECoG data of imagined finger and tongue movement.

The key issue of the present competition task is that the training set and test set came from different sessions with about one week in between. Thus, the corresponding brain states may differ. Fig. 1 shows the averaged amplitude spectra of all trials from each task (motor imagery of either left little finger or tongue) in the training set and from both tasks in the test set on individual channels. Since the curve from the test set is the averaged amplitude spectra of the two tasks, it should lie between the two curves from the training set if the subject 's brain state did not change. However, as shown in this figure, the averaged amplitude spectrum of test set is far above that of the training set. This implies that the brain state of the subject varied significantly from the training session to the test session.

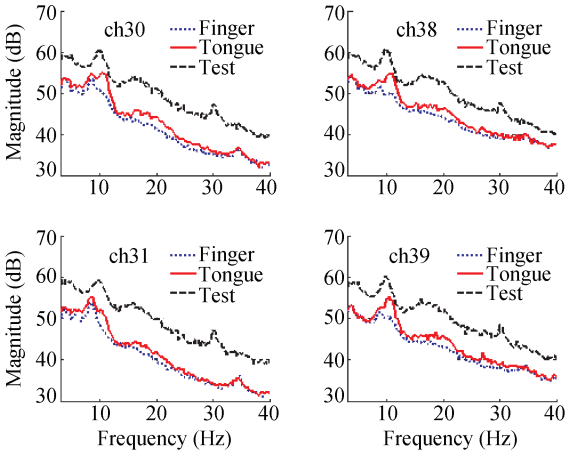

Fig. 1. The averaged magnitude spectra of all trials from each task (motor imagery of either left little finger or tongue) in the training set and from both tasks in the test set on individual channels. Since the curve from the test set is the averaged magnitude spectra of two tasks, it should lie between the two curves from the training set if the subject 's brain state did not change.

Our consideration for this problem was, if the spatial distribution of brain electric activity of the subject did not change in the two sessions, with the sole increase or decrease in magnitude, this might not pose a serious problem. Thus, the following analysis was based on this assumption and its validity should be confirmed by the result of the analysis. The impact of the magnitude change on classification can be alleviated by normalizing the features used for classification. In addition, the combination of multiple features might contribute to increasing classification accuracy.

Our procedures of classifying single-trial ECoG recorded during motor imagery of the left little finger or tongue with session-to-session transfer are summarized in Fig. 2. It consists of the following steps:

1) The multichannel ECoG was preprocessed by reducing the sampling rate, selecting a suitable reference and intercepting the acquired sequence;

2) two different frequency bands were selected to obtain MRP and ERD signals using low pass (0—3 Hz) and band pass (8—30 Hz) filters respectively;

3) three features were extracted from the above two frequency bands. Two features were derived by common spatial subspace decomposition (CSSD) and Fisher discriminant analysis (FDA), and the third one was obtained from the mean value of sample points and FDA;

4) these three features were concatenated and fed into a linear support vector machine (SVM) for classification.

Each of these steps is discussed in more detail in the following sections.

1.1 Data acquisition and preprocessing

During the BCI experiment, a subject had to perform imagined movements of either left small finger or tongue. The electrical brain activities were collected during these trials using an 8×8 ECoG platinum electride grid which was placed on the contralateral (right) motor cortex. Each trial was recorded for 3s duration and all recordings were performed with a sampling rate of 1000 Hz. After amplification the recorded potentials were stored as microvolt values. The given data set includes a training set of 278 trials with labels and an unlabelled test set of 100 trials. A detailed description of the Data Set I in BCI Competition III can be found in [3,5].

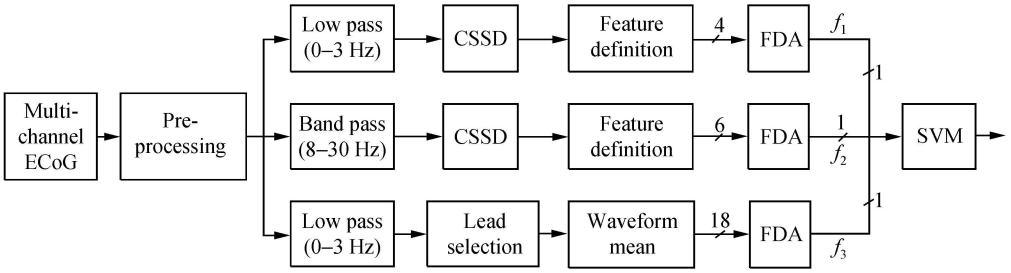

Fig. 2. Procedures for classifying single-trial ECoG during motor imagery of left little finger or tongue with session-to-session transfer.

The given time series is 3000 sample points per trial. In order to save computational load and reduce the requirement for memory, all training and test trials were downsampled to 300 points by decimation. To enhance the differences between these two tasks and reduce the effect of artifacts, common average reference (CAR)[10] was used to re-reference the resampled data.

A time window was applied to intercept the data segment from the 31st to 250th points for further analysis. As shown in Fig. 3, the classification performance depends upon the selection of data segment used for classification. Fig. 3(a) shows the relationship between classification accuracy of feature f_2 (defined in the following section) and the starting point of time window for analysis, while the ending point is fixed at 300. Fig. 3(b) illustrates the relationship between classification accuracy of f_2 and the ending point of the time window, while the starting point is fixed at 1. The original training set was randomly split into subsets for training and validation. The figure demonstrates that classification accuracy increases monotonically with the length of time window except at the beginning and ending parts of the data. Inclusion of these two parts of data (1st—30th point or 251st—30th point) would make classification accuracy decrease and unstable. Hence the time series of 220 points (31st—250th point) were chosen for feature extraction to ensure high classification accuracy.

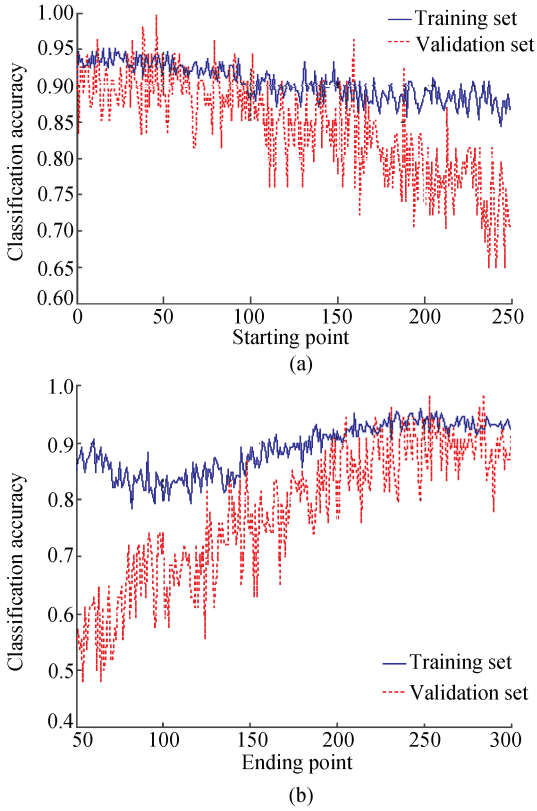

Fig. 3. (a) The relationship of classification accuracy of feature f_2 and the starting point of data (the ending point is fixed at 300); (b) the relationship of the classification accuracy of feature f_2 and the ending point of data (the starting point is fixed at 1). The original training set was randomly split into subsets for training and validation.

1.2 Feature extraction

Feature extraction is a crucial step in pattern recognition. The classification performance mainly depends on whether one extracts the most discriminative information. If a single feature does not achieve satisfactory classification results, several features can be combined to improve the classification performance provided that they are mutually independent.

Although MRP and ERD originate in similar cortical regions and share some common timing features, their magnitude and spatial distribution are different from each other[7]. This suggests that these two features are independent. Thus, their combination can provide complementary information and consequently enhance classification performance.

The following four steps were used to extract feature f_1:

1) The preprocessed data were low pass (0—3

Hz) filtered to obtain MRP signals;

2) the mean of MRP signals was removed to increase the signal difference between the two tasks;

3) two spatial filters were estimated by CSSD using training data, and applied to the data obtained from the previous step to extract source components related to each task;

4) features were defined on the basis of source components and projected onto one dimension by FDA to get feature f_1.

The procedure to extract source components is described below. Two different cognitive tasks, imagined movements of either left little finger (task A) or tongue (task B), were conducted during the experiment. Hence, there are different kinds of information embedded in the ECoG data: those specific to task A or task B, and those common to the two tasks, including noise. Thus the multi-channel ECoG signals $\boldsymbol{X}_A$ (from task A) and $\boldsymbol{X}_B$ (from task B) can be modeled as

$$\boldsymbol{X}_A = [C_A\ C_C]\begin{bmatrix} S_A \\ S_C \end{bmatrix} \tag{1}$$

$$\boldsymbol{X}_B = [C_B\ C_C]\begin{bmatrix} S_B \\ S_C \end{bmatrix} \tag{2}$$

where S_A and S_B are the sources specific to task A and task B respectively, C_A and C_B are their corresponding spatial patterns; S_C is the common source, C_C is its corresponding spatial pattern. The purpose of CSSD is to estimate two spatial filters, which can be used to extract source components S_A and S_B

$$S_A = F_A\boldsymbol{X}_A \tag{3}$$

$$S_B = F_B\boldsymbol{X}_B \tag{4}$$

where F_A and F_B are the spatial filters corresponding to task A and task B respectively. These two source components contain important information for discriminating task A and B. CSSD is based on the simultaneous diagonalization of the two spatial covariance matrices of $\boldsymbol{X}_A$ and $\boldsymbol{X}_B$. Principal component analysis (PCA) and spatial subspace analysis are applied to the two diagonalized covariance matrices to estimate two spatial filters. These two spatial filters are optimal in the sense that they extract task-related components and eliminate common components. Computation step and detailed description about CSSD can be found in [11,12].

Having derived source components, we can define features v_{1A} and v_{1B}. Assume that task A caused a relatively increased ECoG variance over the specific area of the brain, and the variance of the source component filtered by F_{1A} was greatly enhanced compared to that filtered by F_{1B}, and *vice versa*. Given a spatio-temporal signal matrix $\boldsymbol{x}$ of a multichannel ECoG with an unknown label, two runs of spatial filtering by F_A and F_B respectively were applied. Then, features v_{1A} and v_{1B} were defined as

$$v_{1A} = \log\left(\frac{\mathrm{var}(F_{1A}\boldsymbol{X})}{\mathrm{var}(F_{1A}\boldsymbol{X}) + \mathrm{var}(F_{1B}\boldsymbol{X})}\right) \tag{5}$$

$$v_{1B} = \log\left(\frac{\mathrm{var}(F_{1B}\boldsymbol{X})}{\mathrm{var}(F_{1A}\boldsymbol{X}) + \mathrm{var}(F_{1B}\boldsymbol{X})}\right) \tag{6}$$

Both v_{1A} and v_{1B} range between 0 and 1 before the logarithmic operation. Theoretically, v_{1A} will be equal to zero for trials of task B and equal to one for trials of task A. Contrary results will be obtained for v_{1B}. Note that it is the normalization that alleviates the influences of the magnitude or power distinction of brain signals from different sessions on classification accuracy. The logarithmic operation is done in order to make the distribution of elements in v_{1A} and v_{1B} more Gaussian.

When CSSD is used for extracting features, a few most important spatial patterns (i.e. columns of inverse of the spatial filter matrix) can be selected according to their contribution to classification accuracy. For the special case of extracting f_1, two spatial patterns were chosen. Therefore a feature vector of four dimensions $\boldsymbol{v}_1 = [v_{1A}^1\ v_{1B}^1\ v_{1A}^2\ v_{1B}^2]^T$ was constructed, where superscript 1 and 2 denote the index number of chosen spatial patterns, T represents transpose operation.

The procedure to extract feature f_2 is similar to that of extracting f_1, except that (i) the frequency filtering is band pass (8—30 Hz) rather than low pass (0—3 Hz); (ii) the mean of filtered signals does not need to be removed, and (iii) three spatial patterns are chosen instead of two. Thus a feature vector of six dimensions $\boldsymbol{v}_2 = [v_{2A}^1\ v_{2B}^1\ v_{2A}^2\ v_{2B}^2\ v_{2A}^3\ v_{2B}^3]^T$ was created, where superscript 1, 2 and 3 represent the index numbers of chosen spatial patterns.

The same low pass filter as that for extracting f_1 was adopted to obtain MRP signal for extracting feature f_3. Unlike the former two features, feature f_3 was constructed by choosing 18 electrodes (12, 14, 18, 21—24, 29—32, 37—40 and 46—48) and com-

puting the means of all sample points of each trial on selected electrodes. These electrodes were selected according to averaged power difference of trials between the two tasks on each electrode. Thus a feature vector of 18 dimensions $\boldsymbol{v}_3 = [m_1 m_2 \cdots m_{18}]^{\mathrm{T}}$ was first obtained for one trial, where $m_i (i = 1, 2, \cdots, 18)$ is the mean magnitude of each trial on each selected electrode. To mitigate the effect of the magnitude change of ECoG signal from different sessions on classification accuracy, $\boldsymbol{v}_3$ was normalized into $[-1, +1]$ on each dimension.

The preprocessed data were filtered forwardly and reversely to avoid phase distortion in the above frequency filtering steps. The filters were digital low pass or band pass filters of Chebysheve type I.

Features f_1, f_2 and f_3 will be obtained by reducing the dimensionality of feature vectors $\boldsymbol{v}_1$, $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$ respectively in the next section.

1.3 Dimensionality reduction

In principle, feature vectors $\boldsymbol{v}_1$, $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$ can be concatenated to make a feature vector of 28 dimensions and then fed to a classifier. However, given limited training samples, high dimension is not desired for an SVM classifier. The higher the dimension of input vector is, the poorer the generalization ability. Hence, we did not concatenate these three feature vectors directly. Instead, we further reduced their dimensions to one using FDA.

FDA is a linear discriminant analysis whose purpose is to project data from high dimension onto a line so that the new data in one-dimensional space is more manageable. To obtain good separation of the projected data, FDA maximizes the difference of the sample means between two classes and minimizes the total within-class scatter of the projected samples. Three feature vectors $\boldsymbol{v}_1$, $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$ were projected to one dimension according to $f_i = w_i^{\mathrm{T}} \boldsymbol{v}_i + b_i$, $i = 1, 2, 3$, where weights w_i and biases b_i were estimated by FDA using training data, T is the transpose operation, and f_i denotes the three features in one dimension.

1.4 Classification

Three one-dimensional features were concatenated together to form a three-dimensional feature vector $\boldsymbol{f} = [f_1 f_2 f_3]^{\mathrm{T}}$ which was normalized into $[-1\ 1]$ at each dimension and fed into an SVM for classification. A linear kernel was applied and the regularization parameter was determined by 10×10-fold model parameter cross validation. Linear classifiers were adopted because they are generally more robust and have better generalization performance than nonlinear classifiers when finite training samples are available[13,14].

SVMs are the milestone of machine learning field in the past decade, which in theory ensure optimal classification performance. Conceptually, SVMs separate data by trying to find such an optimal hyperplane that maximizes the margin between the nearest samples of the two categories. In the case when the two classes are not linearly separable, a slack variable must be given to ensure the problem solvable. This gives rise to the following SVM optimization problem

$$\begin{aligned} &\min \quad \frac{1}{2} \| w \|^2 + c \sum_i \xi_i \\ &\text{s.t.} \quad y_i (w^{\mathrm{T}} \boldsymbol{x}_i + b) \geqslant 1 - \xi_i, \quad \xi_i > 0 \ \forall i \end{aligned} \tag{7}$$

where w, b, ξ and c are weight, bias, slack variable and regularization parameter respectively, $\boldsymbol{x}_i$ is a data vector (i.e. a three-dimensional feature vector in this work) and y_i is its label. The basic idea of SVMs is discussed in detail by Vapnik[15].

2 Results

Using 10×10-fold cross validation, the classification results of single features and their combination on training set are listed in Table 1. In the table, the classification accuracy of single features was obtained by FDA, while that for feature combination was achieved by SVMs. The best classification accuracy of 95% was obtained in the case of feature combination.

Table 1. The classification accuracy and standard deviation (%, rounded) of single features and their combination on training set

Feature	f_1	f_2	f_3	$f_1 + f_2 + f_3$
Classifier	FDA	FDA	FDA	SVM
Classification accuracy	91 ± 6	90 ± 5	80 ± 7	95 ± 4

The predicted labels of test set were decided by bagging (the acronym of the "bootstrap aggregating") procedure[16]. Bagging utilizes different training subsets, each created by subsampling the original

training set, to train multiple component classifiers and the final classification decision is based on the vote of each component classifier. Specifically, 90% of training samples were drawn randomly each time to construct one SVM classifier and 100 component classifiers were constructed by repeating the subsampling. The predicted labels of test set were determined by voting of these 100 classifiers. Bagging was adopted to avoid unstable classification results caused by the performance variation of classifiers obtained from different learning set. Fig. 4 shows the classification results of the training set and test set from one SVM classifier before computing the sign (i.e. before making final decision).

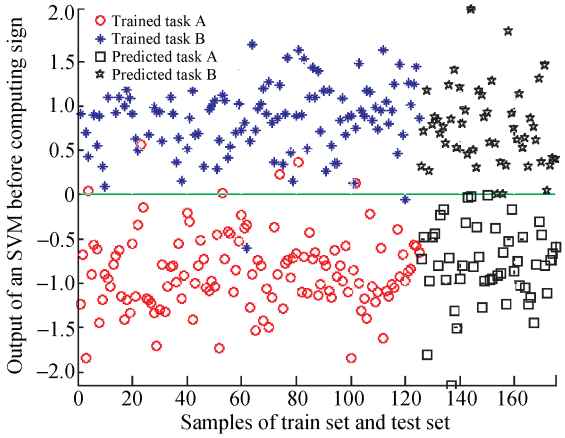

Fig. 4. The classification results of training set and test set from one SVM classifier before computing the sign (i.e. before making final decision). The cycles and asterisks represent trials of task A and task B in the training set respectively, while the squares and five-pointed stars denote trials of predicted task A and task B in the test set respectively.

The final classification accuracy of test set is 91% which was announced by the competition organizers on the website[5]. This result ranks first among 27 contributions from the whole world. Taking into account that the electrode grid was not placed for BCIs, the subject did not recover totally from implantation operation, and more importantly, the data set was derived from different sessions, the classification accuracy could be said to be high.

3 Discussion

The issue of session-to-session transfer is of great importance for BCI research and for constructing a practical BCI system. Without the ability of session-to-session transfer, BCIs have to update the algorithm from time to time, which is troublesome and will make BCIs difficult for practical use.

After the disclosure of test set labels, we re-classify the test data using single features. The classification accuracy is 73%, 87% and 55% for f_1, f_2 and f_3 respectively. The significant decrease of the accuracy rates of features f_1 and f_3 from the training set to test set (see Table 1) suggests that the features derived from MRP signal were unstable and its spatial distribution varied from training session to test session. In contrast, the slight drop of the accuracy rate of feature f_2 suggests that the feature obtained from the ERD signal is relatively stable and its spatial distribution can be thought to be basically unchanged from session to session.

Figs. 5 and 6 show the most important spatial patterns of MRP and ERD signals of the two tasks from the training set and the test set respectively. The spatial patterns are derived from the first column of the inverse of the spatial filter F_A and the last column of the reverse of the spatial filter F_B. These two figures illustrate the spatial distributions of ECoG over the 8×8 electrode grid (each square denotes an electrode). Comparing Fig. 5 (a) with (c) and (b) with (d), we found that the spatial distributions of MRP signal changed considerably in the two different sessions. In contrast, Fig. 6 shows that the spatial distribution of ERD signals is basically unvaried (the focus of spatial patterns is almost at the same electrodes).

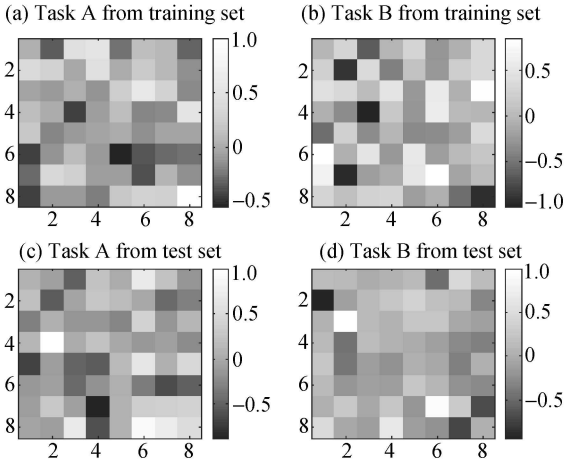

Fig. 5. The most important spatial patterns of MRP signals.

CSSD compares the variances of the source components of two tasks in different spatial patterns. If the spatial distributions of these brain signals do not change in different sessions (e.g. ERD signals), the method can give good classification result. However, if they do vary from one session to another (e.g.

MRP signals), the classification performance will decrease significantly.

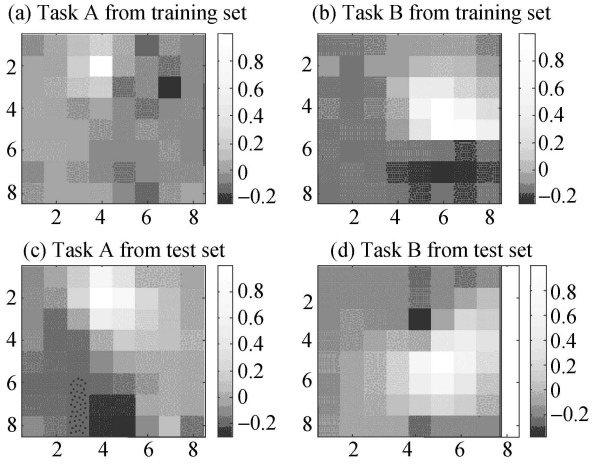

Fig. 6. The most important spatial patterns of ERD signals.

The variation of the brain signals from the training set to the test set caused the classification line of single features to shift towards one task on the test set. The label ratio of two tasks yielded by f_1 classification is 39/61, while that by f_2 classification is 60/40. Thus the classification accuracy on the test set would decrease significantly if only single features were employed. However, different features are complementary and their combination can balance the shift of classification line. The final label ratio 51/49 of the two tasks proved the benefit of feature combination. In addition, the best classification accuracy of single feature is 87% on the test set, while classification accuracy of the combination of three features is 91% which further verifies the advantage of feature combination.

In order to get the best classification result, temporal, frequency and spatial characteristics must be considered simultaneously. The selections of time window, frequency band and electrodes must meet the need of contributing to increased classification accuracy. The selection of time window has been discussed in data preprocessing section. In frequency domain, the influence of frequency band on the classification accuracy of f_1 and f_3 is trivial, while that on classification accuracy of f_2 is significant. We tested several different subbands for each feature and the optimal frequency bands were 0—3 Hz for features f_1 and f_3, and 8—30 Hz for feature f_2.

Spatial characteristics should also be treated carefully. The extraction of f_3 depended directly on electrode selection, while the choice of spatial patterns of features f_1 and f_2 had great effect on classification performance. If optimal electrodes or spatial patterns were not chosen properly, the classification accuracy would have degenerated. Features f_1 and f_2 were extracted by CSSD with all 64 channels employed. We had ever tried to choose 10 optimal electrodes for f_2 extraction and its classification accuracy on training set was 88.5%, a little poorer than that of using all channels. Because CSSD is capable of lead selecting and denoising, manual electrode selection may be unnecessary.

The high classification accuracy demonstrates that ECoG-based BCIs have great potential of application and the problem of session-to-session transfer may not be an intractable one.

Acknowledgments The authors would like to thank the competition organizers for providing data. They would also like to extend special thanks to Prof. Yang Fusheng for a very informative discussion and his valuable comments, and to Dr. Hong Bo for carefully editing the manuscript.

References

1 Wolpaw JR, Birbaumer N, McFarland WJ, et al. Brain-computer interface technology: A review of the first international meeting. IEEE Trans. Rehabil Eng, 2000, 8(2): 164—173
2 Wolpaw JR, Birbaumer N, McFarland DJ, et al. Brain-computer interface for communication and control. Clin Neurophysiol, 2002, 113: 767—791
3 Lal TN, Hinterberger T, Widman G, et al. Methods towards invasive human brain computer interfaces. In: Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2005, 17: 737—744
4 Leuthardt EC, Schalk G, Wolpaw JR, et al. A brain-computer interface using electrocorticographic signals in humans. J Neural Eng, 2004, 1: 63—71
5 http://ida.first.fraunhofer.de/projects/bci/competition_iii
6 Porro CA, Francescato MP, Diamond ME, et al. Primary motor and sensory cortex activation during motor performance and motor imagery: a functional resonance imaging study. J Neurosci, 1996, 16: 7688—7698
7 Ehrsson HH, Geyer S and Naito E. Imagery of voluntary movement of fingers, toes, and tongue activates corresponding body-part-specific motor representations. J Neurophysiol, 2003, 90: 3304—3316
8 Toro C, Deuschl G, Thatcher R, et al. Event-related desynchronization and movement-related cortical potentials on the ECoG and EEG. Electroenceph Clin Neurophysiol, 1994, 93: 380—389
9 Babiloni C, Carducci F, Cincotti F, et al. Human movement-related potentials vs. desynchronization of EEG alpha rhythm: A high-resolution EEG study. NeuroImage, 1999, 10: 658—665
10 McFarland DJ, McCane LM, Stephen SV, et al. Spatial filter selection for EEG-based communication. Electroenceph Clin Neurophysiol, 1997, 103: 386—394

11 Wang Y, Berg P and Scherg M. Common spatial subspace decomposition applied to analysis of brain responses under multiple task conditions: a simulation study. Electroenceph Clin Neurophysiol, 1999, 110: 604—614
12 Muller-Gerking J, Pfurtscheller G and Flyvbjerg H. Designing optimal spatial filters for single-trial EEG classification in a movement task. Electroenceph Clin Neurophysiol, 1999, 110: 787—798
13 Muller KR, Anderson CW and Birch GE. Linear and nonlinear methods for brain-computer interfaces. IEEE Trans Neural Syst Rehab Eng, 2003, 11: 165—169
14 Blankertz B, Dornhege G, Schafer C, et al. Boosting bit rates and error detection for the classification of fast-paced motor commands based on single-trial EEG analysis. IEEE Trans Neural Syst Rehab Eng, 2003, 11: 127—131
15 Vapnik VN. The Nature of Statistical Learning Theory. New York: Springer-Verlag, 2000
16 Duda RO, Hart PE and Stork DG. Pattern Classification. 2nd ed. New York: John Wiley & Sons, 2001