BCI Competition 2003—Data Set IV: An Algorithm Based on CSSD and FDA for Classifying Single-Trial EEG

Yijun Wang*, Zhiguang Zhang, Yong Li, Xiaorong Gao, Shangkai Gao, *Senior Member, IEEE*, and Fusheng Yang

Abstract—This paper presents an algorithm for classifying single-trial electroencephalogram (EEG) during the preparation of self-paced tapping. It combines common spatial subspace decomposition with Fisher discriminant analysis to extract features from multichannel EEG. Three features are obtained based on Bereitschaftspotential and event-related desynchronization. Finally, a perceptron neural network is trained as the classifier. This algorithm was applied to the data set $\langle$self$-$paced 1s$\rangle$ of "BCI Competition 2003" with a classification accuracy of 84% on the test set.

Index Terms—Brain-computer interface (BCI), common spatial subspace decomposition (CSSD), electroencephalogram (EEG), Fisher discriminant analysis (FDA).

I. INTRODUCTION

IN RECENT years, brain-computer interface (BCI) systems based on analysis of single-trial electroencephalogram (EEG) associated with hand movements have developed rapidly. The physiological studies on movement-evoked potentials indicate that the spatio-temporal pattern of EEG differs between left and right hand movements. In the premovement period, Bereitschaftspotential (BP) can be recorded over the vertex region [1]–[3] and during the movement tasks, mu and beta rhythms are found to reveal event-related synchronization and desynchronization (ERS/ERD) over sensorimotor cortex [3]–[6].

G. Pfurtscheller *et al.* first used EEG classification based on ERS/ERD during hand movements for a BCI application [5]. B. Blankertz *et al.* designed a BCI system based on BP prior to voluntary finger movements with Fisher discriminant analysis (FDA) for classifier training, which reached >96% classification accuracy on a subject [1]. The technique of classifying single-trial EEG during finger movements has become an attractive topic in BCI research due to its advantages such as simple experimental approach, low rejection rate, high classification accuracy, short response time, and easy training for users.

In this paper, we propose an algorithm integrating both BP and ERD for classifying movement-evoked EEG during preparation of keystroke. It applies common spatial subspace decomposition (CSSD) to analyze the multichannel EEG signals for two tasks (left and right), and studies the characteristic distributions of BP and ERD in time, frequency, and space domains respectively. The algorithm was employed in off-line analysis of the data set $\langle$self-paced 1s$\rangle$ of "BCI Competition 2003." These data are single-trial EEGs recorded during voluntary self-paced tapping [7].

II. METHODOLOGY

A. Feature Consideration

1) Bereitschaftspotential (BP): Slowly decreasing potentials named movement-related cortical potentials (MRCPs) can be recorded before voluntary limb movements. BP is one of the main components of MRCPs, which can be recorded with the maximum amplitude over the vertex region.

Contralateral dominance is a significant character of BP. Fig. 1 shows the averaged trials of each channel and the spatial distribution of BP on the scalp at 140 ms and 320 ms before the key press. The declining waveforms of most channels reveal the decreasing nature of BP and the spatial distributions show the pronounced contralateral dominance. As shown in the images [Fig. 1(a)], BP of left finger movement is dominant over right vertex region and it is more obvious at the moment of -140 ms, which is closer to the actual keystroke, than at -320 ms. Moreover, even on the same channel, the responses of different tasks differ greatly in amplitudes. Thus, the remarkable difference of BP's distributions between the two tasks can be taken as an important basis for classification.

2) Event-Related Desynchronization (ERD): ERD and ERS represent the changes of the ongoing EEG activity characterized by decrease or increase of power in the given frequency bands. For actual finger movements, ERD/ERS in mu and beta rhythms are obvious over somatosensory or motor cortex, also with contralateral dominance. Preparation of movement is typically accompanied by ERD in mu and beta rhythms. ERS in beta rhythm, considered as the rebound of ERD, occurs after movement.

In our study, the epoch before actual keystroke was analyzed to predict the upcoming task, so only the character of ERD is applicable. Fig. 2 displays the power spectrum on channels C3 and
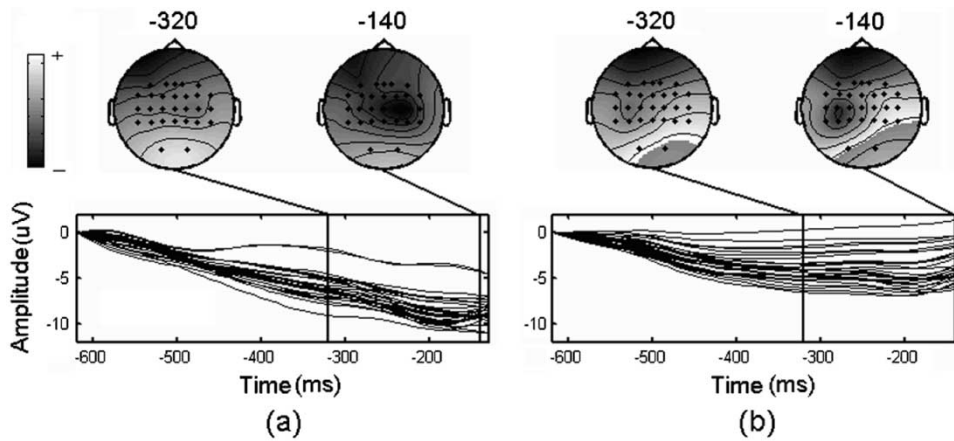
Fig. 1. Spatial distribution of BP at 320 ms and 140 ms before keystroke and averaged trials of the 28 channels corresponding to (a) left finger movements and (b) right finger movements. The amplitudes at the beginning of the epoch are normalized to zero.
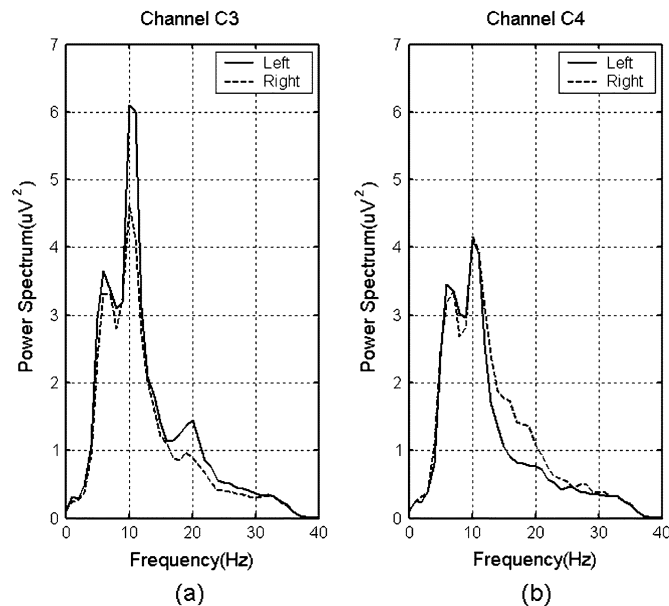

Fig. 2. Average power spectrum of EEG signals on channels C3 and C4 over motor cortex for the two tasks. Legend "—" stands for left finger movement and "- - -" stands for right finger movement.

C4 corresponding to both tasks. High-pass filtered EEG signals (cut off at 5 Hz) show difference in power distribution between the two tasks. For channel C3 [Fig. 2(a)], the power of mu and beta rhythms evoked by right hand movement is lower than that of left hand, which is consistent with the principle of contralateral dominance. Similar conclusion can be drawn from Fig. 2(b) for channel C4. The difference of power distribution derived from ERD implies another choice of feature for classification.

B. Feature Extraction

The activities specific to the tasks (left or right finger movement) are usually overwhelmed by the spontaneous EEG and other nontask activities. Here, the method of CSSD is used to extract the task-related components and eliminate the background activities. Y. Wang *et al.* successfully used it to extract signal components specific to one task from multichannel EEG of multiple tasks [8].

The multichannel EEGs are denoted as matrices X_L (left) and X_R (right) with dimension of N (channels) by T (samples). They can be modeled using multiple simultaneously activated sources as follows:

$$X_L = \begin{bmatrix} C_L & C_C \end{bmatrix} \begin{bmatrix} S_L \\ S_C \end{bmatrix} \quad X_R = \begin{bmatrix} C_R & C_C \end{bmatrix} \begin{bmatrix} S_R \\ S_C \end{bmatrix}$$

where S_L and S_R are task-related source activities corresponding respectively to left and right hand movements, S_C is the source corresponding to background activities that are common to both conditions. Assume that S_L consists of m_L sources and S_R consists of m_R sources, then C_L and C_R consist of m_L and m_R spatial patterns related with S_L and S_R respectively. A spatial pattern is an $N \times 1$ vector indicating the signal distribution over the N electrodes caused by a specific source. C_C stands for the common spatial patterns related with S_C.

The aim of CSSD is to design spatial filters that lead to the estimations of S_L and S_R. This method is based on the simultaneous diagonalization of the spatial covariance matrices of X_L and X_R. Principal component analysis (PCA) and spatial subspace analysis are performed to eliminate the common components and extract the task-related components. More details of the algorithm can be found in [8]. As the result, two spatial filters, $\mathbf{SF}_L$ and $\mathbf{SF}_R$, are constructed corresponding to the two tasks. The estimations of S_L and S_R are given by $S_L = \mathbf{SF}_L \cdot X$ and $S_R = \mathbf{SF}_R \cdot X$, where X is a single-trial EEG. It can be demonstrated that the specific temporal source activities estimated by CSSD are approximately the same as that in the spatio-temporal source model. Theoretically, for left finger movements, S_L will be much stronger than S_R and for right finger movements, the result is just the opposite.

FDA can linearly project high dimensional data to one-dimensional vector so as to facilitate the classification. It bears the advantages of easy implementation and low computational cost. A linear discriminant function is designed by use of FDA on the training set, which reduces the feature space to 1-D vector representing the proximity of the extracted feature belonging to either pattern.
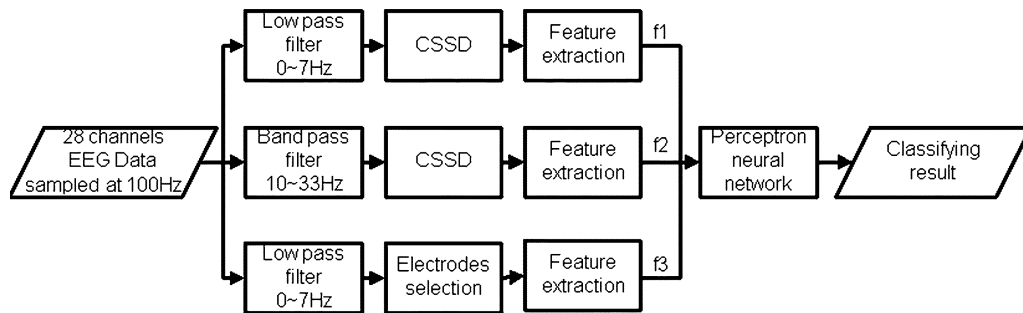
Fig. 3. Flow chart of classification algorithm, including the modules of temporal filtering, CSSD, feature extraction, electrodes selection, and perceptron neural network.
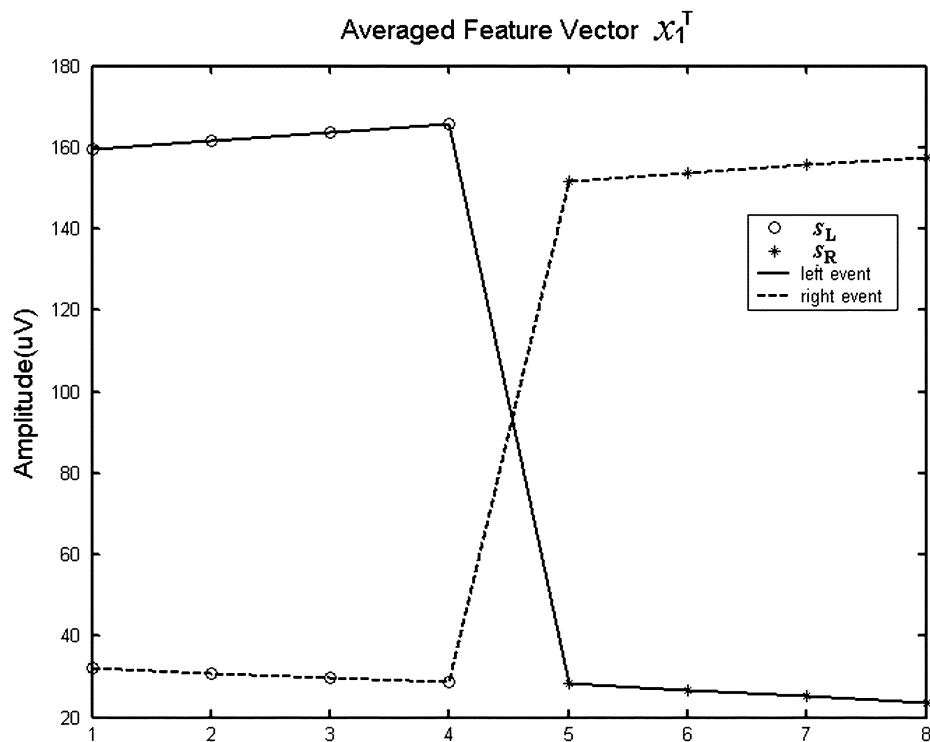

Fig. 4. Averaged feature vector x_1^T consisting of s_L and s_R, denoted as legends "o" and "*". "—" and "- - -" are used to link the elements of x_1^T corresponding to the two tasks respectively.

Fig. 3 shows the flow chart of the algorithm, which can be summarized as follows: three features $(f_i, i = 1, 2, 3)$ are extracted from the single-trial EEG, then a perceptron is trained with the extracted features as input. Note that the data have been subsampled to 100 Hz in order to reduce the computational cost. After the decimation, 50 data points are retained for each trial. For convenience, index of sample point (1 to 50) will be utilized to describe the corresponding time (-620 ms to -130 ms) infra. Our algorithm only deals with the frequency components in the range of 0 to 33 Hz, because the higher frequency components are less related with BP and ERD.

1) Extraction of Feature f_1: The first feature f_1 is derived from BP, manifested as a declining slope of EEG with contralateral dominance. A zero-phase low-pass filter (0–7 Hz) is used to preprocess the data, and then a fixed time window (sample points of 44–47) is used to intercept the segment that bears the

most obvious difference between the two tasks. Then the new data X with dimension of 28 (channels) by 4 (samples) will pass through $\mathbf{SF}_L$ and $\mathbf{SF}_R$ (with dimension of 1×28) corresponding to the most significant spatial patterns. Define an 8×1 feature vector as $x_1 = [s_L \; s_R]^T$, where $s_L = \mathbf{SF}_L \cdot X$ and $s_R = \mathbf{SF}_R \cdot X$. FDA projects x_1 to one dimension as $f_1 = W_1^T x_1 + b_1$, where the weight W_1 and the bias b_1 are determined by the training set. As shown in Fig. 4, s_L of left trials and s_R of right trials have larger amplitudes than that of the contrary patterns, ensuring an effective classification with feature f_1.

2) Extraction of Feature f_2: The second feature f_2 is derived from ERD. An analogous procedure is applied to the extraction of f_2 as that for f_1. However, different parameters are employed due to the different physiological background of ERD and BP. The parameters that significantly influence the classification
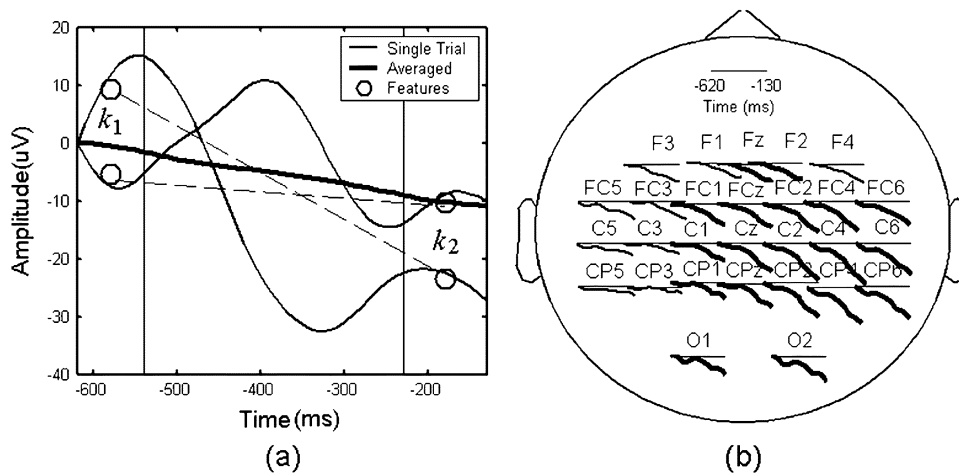
Fig. 5. (a) Averaged potential and two single trials on channel C4 of left finger movements. Legend "o" indicates the two features k_1 and k_2. (b) Differences of all the 28 channels between left and right finger movements.

result include: passband of temporal filters, time window, number of spatial patterns selected for CSSD, and feature definition on the result of spatial filtering.

It has been mentioned that ERD also shows contralateral dominance though it appears on both hemispheres. A band-pass filter (10–33 Hz) and a time window (sample points of 19–50) are worked out to preprocess the data. As three most important spatial patterns are selected to construct the spatial filters, after spatial filtering, s_{Li} and $s_{Ri}(i = 1, 2, 3)$ with dimension of 1×32 are obtained. A 192×1 feature vector is then constructed by concatenating s_{Li} and s_{Ri}, i.e., $x_2 = [s_{L1}s_{R1}s_{L2}s_{R2}s_{L3}s_{R3}]^T$. Because a high dimensional feature vector is harmful to the stability of FDA, the dimension of x_2 is reduced to 24×1 by taking the averaged absolute value of its components consecutively for every eight elements. Absolute values are employed because they are just the parameters that reflect power magnitude. FDA with x_2 results in feature f_2.

3) Extraction of Feature f_3: Like feature f_1, feature f_3 is also derived from BP. The same low-pass filter as that for f_1 extraction preprocesses the signals before the comparison of waveforms of the two tasks. Fig. 5(b) illustrates the contrast of single trials on all the 28 channels by subtracting averaged X_R from averaged X_L. As shown in the figure, differences are less obvious on F3, F1, F4, FC5, FC3, C5, C3, CP5, and CP3 than the other channels, so these channels are rejected. Two features, k_1 and k_2, are defined as the mean values of the beginning (sample points of 1–8) and ending (sample points of 41–50) portions of a single trial on one channel [see Fig. 5(a)]. Their difference is a good feature characterizing the declining trend of the waveform. All k_1 and k_2 of the remaining 19 channels are concatenated into a feature vector $x_3 = [k_{11}k_{12} \ldots k_{i1}k_{i2}]^T (i = 19)$, leading to the feature f_3 by FDA.

C. Classification

After extraction of the above three features, a perceptron, which is a fast and reliable neural network suited for simple classification problems, is used as the classifier. The

vector $p = [f_1 f_2 f_3]^T$ is fed as the input, and the output is $a = \mathrm{hardlim}(W^T p + b)$, where W and b are the weights and bias of the network determined by the training data. $\mathrm{hardlim}(x)$ denotes the hard-limit transfer function, which returns 0 or 1 corresponding to left or right finger movements respectively.

III. RESULTS

The classification accuracy is $(92.98 \pm 2.50)\%$ on the training set with the leave-one-out method. The accuracy on the test set is 84% ultimately. For comparison, similar algorithm is also applied to the data set ⟨self−paced 2s⟩ of "BCI Competition 2002." The accuracy is 98.23% due to the fact that BP is more prominent in this data set which greatly improves the effects of the features f_1 and f_3.

The results demonstrate that the proposed algorithm is reasonable and effective for classifying single-trial EEG during self-paced tapping. High classification accuracy of the algorithm makes it a feasible choice for a practical BCI system.

IV. DISCUSSIONS

A. Time Window and Frequency Window

The frequency window for temporal filtering and the time window for CSSD in the preprocessing stage should be considered carefully. The prior knowledge that BP 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, a sliding window with adjustable bandwidth and cutoff frequency is used to find the appropriate frequency band. The optimal frequency band for BP is found to be 0–7 Hz and that for ERD is 10–33 Hz. The same approach is used to determine the time windows for CSSD and the results are sample points of 44–47 for BP and 19–50 for ERD. Fig. 6 shows the relationship between different time windows and classification accuracy when classifying with feature f_2 alone. Here the test is on the training set itself without use of leave-one-out method. The ending point is fixed at point 50, and the starting point leading to the highest accuracy is point 19.
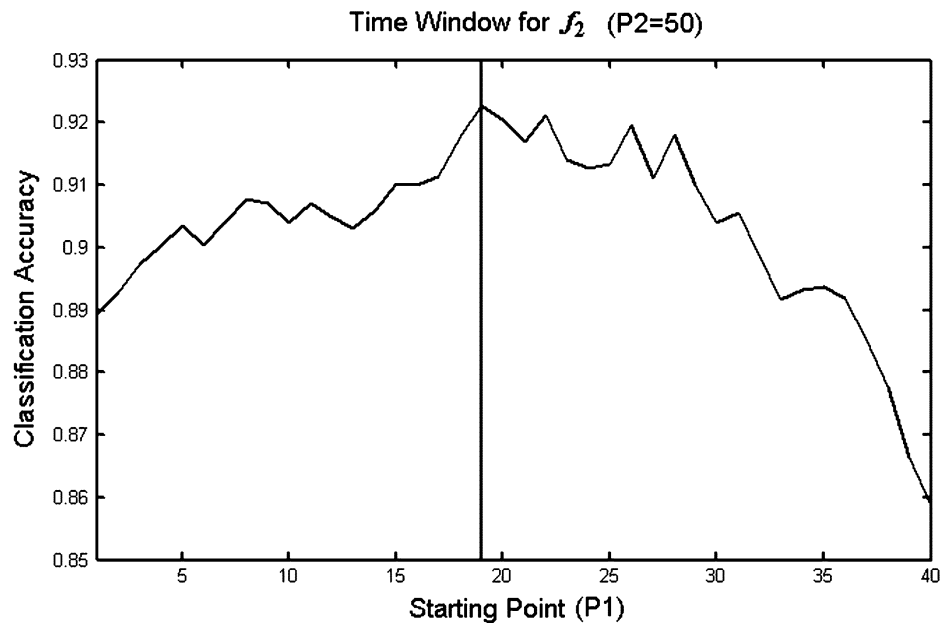
Fig. 6.   Classification accuracy based on feature f_2 with different time windows. The ending point (P2) is 50 and the best starting point (P1) is 19.

TABLE I
CLASSIFICATION ACCURACY CORRESPONDING TO SINGLE EIGENVALUE FOR
SPATIAL FILTER DESIGN

Features	Eigenvalues				
	e_1	e_2	e_3	e_4	e_5
f_1	**82**	51	53	55	52
f_2	**84**	**77**	**70**	61	59

Leave-one-out accuracy corresponding to the 5 largest eigenvalues. The

bold font letters refer to the selected eigenvalues for spatial filter design.

B. Spatial Filter Design

Proper selections of spatial patterns can markedly improve the classification accuracy. The accuracy may be degraded if insignificant spatial patterns are chosen. The selection of proper spatial patterns is determined by the magnitude of eigenvalues of the "whitening transformed" spatial covariance matrices of X_L and X_R [8]. In principle, the eigenvectors corresponding to the several largest eigenvalues should be chosen. However, in our practice, they are chosen through comparing their contributions to classification accuracy. Each one of these eigenvectors is used individually to design the spatial filter. Features are extracted from the filtering results and classified by FDA. The significant eigenvectors are determined based on their relevant classification accuracy.

Table I shows the corresponding accuracy of the five largest eigenvalues (e_1 to e_5, in descending order) in the extractions of the features f_1 and f_2. For feature f_1, the accuracy of using e_1 is 82%, much higher than those of using other eigenvalues (around 50%). Therefore, the spatial filters for feature f_1 are worked out with only the largest eigenvalues. For feature f_2, using e_1, e_2, or

e_3 individually leads to an accuracy above 70%, so all the three eigenvalues are accepted to design the spatial filters.

REFERENCES

[1] B. Blankertz, G. Curio, and K. R. Müller, "Classifying single trial EEG: Toward brain computer interfacing," in *Advances in Neural Information Processing Systems (NIPS 01)*, T. G. Diettrich, S. Becker, and Z. Ghahramani, Eds.   Cambridge, MA, 2002, vol. 14, pp. 157–164.

[2] M. Kukleta and M. Lamarche, "Steep early negative slopes can be demonstrated in pre-movement bereitschaftspotential," *Clin. Neurophysiol.*, vol. 112, pp. 1642–1649, 2001.

[3] J. A. Pineda, B. Z. Allison, and A. Vankov, "The effects of self-movement, observation, and imagination on mu rhythms and readiness potentials (RP's): Toward a brain-computer interface (BCI)," *IEEE Trans. Rehab. Eng.*, vol. 8, pp. 219–222, June 2000.

[4] 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.

[5] G. Pfurtscheller, J. Kalcher, C. Neuper, D. Flotzinger, and M. Pregenzer, "On-line EEG classification during externally-paced hand movements using a neural network-based classifier," *Electroenceph. Clin. Neurophysiol.*, vol. 99, pp. 416–425, 1996.

[6] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Classification of movement-related EEG in a memorized delay task experiment," *Clin. Neurophysiol.*, vol. 8, pp. 1353–1365, 2000.

[7] 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, pp. 1044–1051, June 2004.

[8] Y. Wang, P. Berg, and M. Scherg, "Common spatial subspace decomposition applied to analysis of brain responses under multiple task conditions: A simulation study," *Clin. Neurophysiol.*, vol. 110, pp. 604–614, 1999.

Yijun Wang was born in Fukien, China, in 1979. He received the B.E. degree in biomedical engineering from Tsinghua University, Beijing, China, in 2001. He is currently working towards the M.E. degree in the Department of Biomedical Engineering, Tsinghua University.

His research interests are brain-computer interface and biomedical signal processing.

Xiaorong Gao was born in Beijing, China, in 1963. He received the B.S. degree in biomedical engineering from Zhejiang University Hangzhou, China, in 1986, the M.S. degree in biomedical engineering from Peking Union Medical College, Peking, China, in 1989, and the Ph.D. degree in biomedical engineering from Tsinghua University, Beijing, China, in 1992.

He has been with the Department of Electrical Engineering, Tsinghua University since 1992. His current research interests are in biomedical signal processing and medical instrumentation.

Zhiguang Zhang graduated from the Department of Precision Instruments and Mechanology, Tsinghua University, Beijing, China, in 1970.

He is currently an Associate Professor with the Department of Biomedical Engineering, Tsinghua University. His research interests include biomedical image and signal processing, pattern recognition.

Shangkai Gao (SM'94) graduated from the Department of electrical engineering of Tsinghua University, Beijing, China, in 1970, and received the M.E. degree of biomedical engineering in 1982 in same department.

She is currently a Professor with the Department of Biomedical Engineering, Tsinghua University. Her research interests include biomedical signal processing and medical ultrasound.

Prof. Gao is now an Associate Editor of IEEE TRANSACTION ON BIOMEDICAL ENGINEERING.

Yong Li was born in Hunan, China, in 1976. He received the B.E. degree and Ph.D. degree in biomedical engineering from Tsinghua University, Beijing, China, in 1998 and 2003, respectively.

His research interests include biomedical signal processing and medical ultrasound.

Fusheng Yang received the B.Sc. degree of electrical engineering from Amoy University, Fukien, China, in 1949.

Since 1951, he has been a faculty member ithe the Department of Electrical Engineering, Tsinghua University, Beijing, China, where he has been Professor since 1980. He is the author of *Biomedical Signal Processing* (Beijing, China: Higher Education Press, 1989) and *Engineering Analysis of Wavelet Transform and its Application* (Beijing, China: Science Press, 1998). His current research interest lies mainly in the application of spatial analysis, time-frequency, time-scale analysis and nonlinear dynamics analysis to biomedical signals.