IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON IEEE RELIABILITY SOCIETY SECTION

Received December 20, 2019, accepted January 7, 2020, date of publication January 20, 2020, date of current version January 28, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2967814

Common Cross-Spectral Patterns of Electroencephalography for Reliable Cognitive Task Identification

HAIFENG LI[1], (Member, IEEE), SHANG FENG[1], (Senior Member, IEEE), LIN MA[1], ZHONGLIANG XU[1], RUIFENG XU[2], AND TZYY-PING JUNG[3], (Fellow, IEEE)

[1]School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China
[2]School of Computer Science and Technology, Harbin Institute of Technology (Shenzhen), Shenzhen 518000, China
[3]Institute for Neural Computation, University of California San Diego, San Diego, CA 92093, USA

Corresponding author: Haifeng Li (lihaifeng@hit.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFC0806800, in part by the National Natural Science Foundation of China under Grant 61671187, in part by the Shenzhen Foundational Research Funding under Grant JCYJ20180507183608379, in part by the Shenzhen Key Laboratory of Innovation Environment Project under Grant ZDSYS201707311437102, in part by the Funds for National Scientific and Technological Development, in part by the Project funded by the China Postdoctoral Science Foundation under Grant 2019M663095, and in part by the Applied Basic Research Programs under Grant CJN13J004.

ABSTRACT A current popular feature extraction method of classifying cognitive states and task engagements from electroencephalographs (EEG) is common spatial patterns (CSP). However, the classical CSP only focuses on the correlation between the signals and ignores all characteristics of the signals in the time domain and the frequency domain. In this paper we propose Common cross-spectral patterns (CCSP) a novel EEG feature extraction method for combining spectral and spatial patterns based on cross-spectral density (CSD) to overcome the disadvantages of classical CSP. In CCSP method. the cross-power matrices (CPMs) are extracted to measure the spatial correlation of each task in the band of interest. Then, an orthogonal linear transformation is constructed by simultaneously diagonalizing the CPMs of two tasks. Finally, each band's logarithmic power of the transformed signals is extracted for the support vector machine (SVM) classifier. The experiment results on multiple datasets showed that the CCSP algorithm is fully applicable to multi-channel EEG for reliable multi-cognitive-task identification.

INDEX TERMS Brain-computer interfaces, biomedical signal processing, common spatial patterns, cognitive task identification, reliability, electroencephalography.

I. INTRODUCTION

The human brain is always at a particular cognitive state or in transition from one state into another one, and any changes in brain activities will be reflected in scalp electroencephalograph (EEG). It is of great significance to properly measure and analyze brain activities for the research of human cognition, brain computer interfaces (BCIs) and the clinical diagnosis of various nervous and mental diseases.

This study centers on the analysis of EEG recorded during cognitive tasks, which traces back to the lateralization of brain function and covert mental activity research in the 1970s [1], [2]. Motor imagery activities that modulate sensorimotor oscillations in EEG [3] can be detected by the BCI to infer a user's intent [4]–[8]. Therefore, motor imagery was a cognitive task research focus on the BCI Competition II, III and IV [9]–[11] from the start of the 21st century.

The feature extraction algorithm plays a vital role in cognitive task recognition. Since the training process is time-consuming and demanding for the subject, the training sets are relatively small compared to the data collected in real applications. Thus, it is essential for the feature extraction algorithm to make the most of the limited training data. Due to the spectral properties of EEG, various power spectrum analyses are the most popular feature extraction methods, including the Fourier transform [12], the wavelet transform [13], [14], and others. Because different band activities have certain physiological and pathological significance,

The associate editor coordinating the review of this manuscript and approving it for publication was Zhaojun Li.

how to divide different bands is a key problem. Keirn and Aunon [12] created a feature set consisting of the asymmetry ratios and the power values for each electrode at four frequency bands: δ (0∼3 Hz), θ (4∼7 Hz), α (8∼13 Hz), and β (14∼20 Hz). His results indicate that it is possible to accurately distinguish between any two tasks being investigated. The high-frequency band power has attracted increasing attention. Palaniappan [15] and Zhang *et al.* [16] respectively used spectral power and asymmetric ratios from the γ (24∼37 Hz) band, the high-frequency band (40∼100 Hz) and lower-frequency bands, and mental task classification performance was improved.

However, there is a strong correlation between different channels of EEG in mental tasks. Apparently, spectral power approaches that assume that all the electrodes are mutually independent have significant limitations [17]. A more practical solution is principal component analysis (PCA) of the spatial correlation between signals [18], [19]. In a binary classification scenario, a more appropriate spatial filtering method is common spatial patterns (CSP) [20], [21]. CSP uses the simultaneous diagonalization of covariance matrices of two populations to convert two observation sets of possibly correlated variables into two sets of values of linearly uncorrelated variables and can ensure the maximum difference of two populations in the variance of the converted variables.

CSP is commonly used for signal analysis in the time domain. It was first introduced by Zoltan J. Koles to extract the features of variance from multi-channel CSP to discriminate normal subjects from the patients with excellent results in feature selection and classification performance [17], [22]. The source localization method of CSP decomposition is significantly better than that of PCA-based source localization for the background CSP from patient [23], [24]. Another application of CSP in EEG analysis is the classification of movement tasks [25], [26] and movement imagination [27], [28]. EEG from movement imagination tasks were generally filtered from 8∼30 Hz before the CSP transformation since this broad frequency range contains all μ and β frequency components related to movement.

However, CSP is sensitive only to the spatial covariance matrices of the two classes of tasks engagement from EEG, and all time and frequency information is lost. A simple solution is to filter EEG in the frequency bands of interest in the time domain for the CSP transformation. Muller divided α (8∼12 Hz), lower α (8∼10 Hz), upper α (10∼12 Hz), β (19∼26 Hz), γ (38∼42 Hz), and a broad band of 8∼30 Hz [25]. Novi *et al.* [29] and Ang *et al.* [30] selected an equivalent bandwidth filter bank. Zhang et al. constructed a filter bank with center frequencies at a constant interval in the logarithm frequency domain and a uniform Q-factor (bandwidth-to-center frequency) [31]. These methods were proven to have superior performance than the conventional CSP for motor imagery classification.

Besides motor imagery, recent studies showed that spectral domain features of EEG can also be utilized for mental tasks classification [32], mental states monitoring [33] and cognitive decision prediction [34], [35]. Previous study has also shown that spatial dynamic was also part of EEG spectral features in cognitive tasks analysis [36]. As a result, a feature extraction method that combine spatial features and spectral features would be a better choice for EEG based cognitive tasks classification.

In this paper, we propose common cross spectral patterns (CCSP), which is a spatial-spectral transformation method as a feature extraction method for EEG signals. The method first uses the cross-power spectrum[38], [39] to measure the spectral dependence of the signals. Then, the CSP transformation is constructed by using the cross-power matrices of the band of interest. Finally, it extracts the band-power features. This method reduces redundant information by using the CSP transformation. Compared with the CSP in the temporal domain, the CCSP replace time temporal filtering with fast Fourier transformation (FFT), which can effectively reduce the calculation complexity and time delay, and has a more flexible band implementation strategy.

The main goal of this paper is to employ CSP to extract more effective band-power features from EEG signals in order to classify signals according to cognitive task engagement. The rest of this paper is organized as follows. Section 2 describes the proposed method and algorithm optimization. Section 3 describes the experiments and results and is followed by the conclusions in Section 4.

II. METHOD

The architecture of the cognitive task recognition system based on CCSP with support vector machine (SVM) is shown in Figure 1.

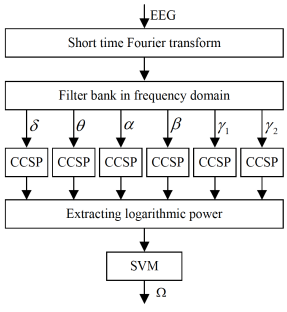


FIGURE 1. Cognitive task recognition system based on CCSP, where Ω is the class of task.

A. OPTIMIZING SPATIAL DISTRIBUTION OF VARIANCES BASED ON SIMULTANEOUS DIAGONALIZATION

The CSP method is a linear orthogonal transformation method that maximizes the variance difference of the transformed random variable between two populations. The transformation is constructed by the simultaneous diagonalization of the covariance matrices of two populations.

1) COVARIANCE MATRIX

The covariance matrix is a measure of how much each pair of variables of a random vector change together. The covariance matrix (Σ) of the random vector X is defined as

$$\Sigma = \mathbb{V}(X) = \mathbb{E}\left[(X-\mu)(X-\mu)^*\right] = \mathbb{E}\left(XX^*\right) - \mu\mu^* \quad (1)$$

where μ is the mean of X, and X^* denotes the complex conjugate transpose of the matrix. Every covariance matrix is Hermitian symmetric and positive semidefinite. Evidently, the diagonal elements contain the variances of individual variables and the off-diagonal elements contain the covariance between two variables. According to the linear combination property theorem, the covariance $\sigma^2$of the random variable $y = \omega^* X$ can be obtained by (2).

$$\sigma^2 = \mathbb{V}(\boldsymbol{y}) = \mathbb{V}(\boldsymbol{\omega}^*\mathbf{X}) = \boldsymbol{\omega}^*\boldsymbol{\Sigma}\boldsymbol{\omega} \quad (2)$$

2) COVARIANCE MATRIX PROPERTIES

Suppose that the observation vectors are $X|\,\Omega_1 \backsim N(0, \Sigma_1)$ and $X|\,\Omega_2 \backsim N(0, \Sigma_2)$ where $\Omega_1$and Ω_2 are two populations. Let us find a vector ω that maximizes the difference between the covariance $\mathbb{V}(\omega^* X|\,\Omega_2)$ and the covariance $\mathbb{V}(\omega^* X|\,\Omega_1)$ and is subject to the condition $\mathbb{V}(\omega^* X|\,\Omega_2) = 1$ (constant). This optimization problem can be expressed as (3).

$$\begin{cases} \mathbf{argmax}_{\boldsymbol{\omega}} \left|\boldsymbol{\omega}^*\boldsymbol{\Sigma}_2\boldsymbol{\omega} - 1\right| \\ \boldsymbol{\omega}^*\boldsymbol{\Sigma}_1\boldsymbol{\omega} = 1 \end{cases} \quad (3)$$

We introduce a Lagrange multiplier λ, and the Lagrange function is defined by (4):

$$\mathbf{L}(\omega, \lambda) = \omega^*\Sigma_2\omega - 1 + \lambda\left(\omega^*\Sigma_1\omega - 1\right) \quad (4)$$

The partial derivatives of $\mathbf{L}(\boldsymbol{\omega}, \boldsymbol{\lambda})$ are described in (5):

$$\begin{cases} \dfrac{\partial L}{\partial \omega} = 2\Sigma_2\omega - 2\lambda\Sigma_1\omega = 0 \\ \dfrac{\partial L}{\partial \lambda} = \omega^*\Sigma_1\omega - 1 = 0 \end{cases} \quad (5)$$

Let $\boldsymbol{\Sigma_2}$ and $\boldsymbol{\Sigma_1}$be positive definite Hermitian matrices. Then, the result is (6).

$$\begin{cases} \boldsymbol{\Sigma_1^{-1}\Sigma_2\omega} = \boldsymbol{\lambda\omega} \\ \boldsymbol{\omega^*\Sigma_1\omega} = \mathbf{1} \end{cases} \quad (6)$$

Evidently, the solution to the problem is the eigenvector of the quotient matrix ($\boldsymbol{\Sigma_1^{-1}\Sigma_2}$) of two covariance matrices after normalization with respect to $\boldsymbol{\Sigma_1}$, and $\boldsymbol{\lambda}$ is the eigenvalue of $\boldsymbol{\Sigma_1^{-1}\Sigma_2}$.

Because $\boldsymbol{\Sigma_1^{-\frac{1}{2}}}$ and $\boldsymbol{\Sigma_1^{\frac{1}{2}}}$ are positive definite Hermitian matrices, note that $\boldsymbol{\Sigma_1^{\frac{1}{2}}\Sigma_1^{-1}\Sigma_2\Sigma_1^{-\frac{1}{2}}} = \boldsymbol{\Sigma_1^{-\frac{1}{2}}\Sigma_2\Sigma_1^{-\frac{1}{2}}} = \left(\boldsymbol{\Sigma_1^{-\frac{1}{2}}\Sigma_2\Sigma_1^{-\frac{1}{2}}}\right)^*$. Therefore, $\boldsymbol{\Sigma_1^{-1}\Sigma_2}$ is similar to a positive definite Hermitian matrix ($\boldsymbol{\Sigma_1\Sigma_2^{-1}} \sim \boldsymbol{\Sigma_1^{-\frac{1}{2}}\Sigma_2\Sigma_1^{-\frac{1}{2}}}$). According to the spectral theorem for Hermitian matrices, $\boldsymbol{\Sigma_1^{-\frac{1}{2}}\Sigma_2\Sigma_1^{-\frac{1}{2}}}$ is unitarily diagonalizable, and hence, $\boldsymbol{\Sigma_1^{-1}\Sigma_2}$ must also be diagonalizable by similarity.

$$\Sigma_1^{-1}\Sigma_2 W = W\Lambda \quad (7)$$

where $\boldsymbol{\Lambda}$ is the eigenvalue matrix of $\boldsymbol{\Sigma_1^{-1}\Sigma_2}$. Each column ($\boldsymbol{\omega}$) of $\boldsymbol{W}$ is its eigenvector corresponding to $\boldsymbol{\lambda}$.

3) THE CLASSIFICATION SIGNIFICANCE OF EIGENVALUE

From (6), we have the following

$$\begin{aligned} \Sigma_2\,\omega &= \Sigma_1\omega\,\lambda \\ \omega^*\Sigma_2\,\omega &= \omega^*\Sigma_1\omega\,\lambda \\ \omega^*\Sigma_2\,\omega &= \lambda \end{aligned} \quad (8)$$

Hence, after the linear transformation $\boldsymbol{y = \omega^* X}$, $\boldsymbol{\Omega_1}$ is subject to the standard normal distribution($\mathbf{y}|\boldsymbol{\Omega_1} \backsim \boldsymbol{N(0, \Sigma_1)}$), and $\boldsymbol{\Omega_2}$ is subject to normal distribution ($\mathbf{y}|\boldsymbol{\Omega_1} \backsim \boldsymbol{N(0, \Sigma_1)}$).

We can further use the probability of a correct Bayes decision to measure the separability between two populations. The classification boundary point ($\boldsymbol{y}$) of an equal probability density can be calculated as follows.

$$\begin{aligned} &\mathbf{p}\left(\mathbf{y}|\,\boldsymbol{\Omega_1}\right)\boldsymbol{P}\left(\boldsymbol{\Omega_1}\right) \\ &\quad = \boldsymbol{p}\left(\mathbf{y}|\,\boldsymbol{\Omega_2}\right)\boldsymbol{P}\left(\boldsymbol{\Omega_2}\right) \\ &\mathbf{lnp}\left(\mathbf{y}|\,\boldsymbol{\Omega_1}\right)\boldsymbol{P}\left(\boldsymbol{\Omega_1}\right) + \mathbf{ln}\boldsymbol{P}\left(\boldsymbol{\Omega_1}\right) \\ &\quad = \mathbf{ln}\boldsymbol{p}\left(\mathbf{y}|\,\boldsymbol{\Omega_2}\right) + \mathbf{ln}\boldsymbol{P}\left(\boldsymbol{\Omega_2}\right) \\ &\mathbf{ln}\boldsymbol{P}\left(\boldsymbol{\Omega_1}\right) - \frac{\boldsymbol{y^2}}{\mathbf{2}} \\ &\quad = \mathbf{ln}\boldsymbol{P}\left(\boldsymbol{\Omega_2}\right) - \frac{\mathbf{1}}{\mathbf{2}}\mathbf{ln}\boldsymbol{\lambda} - \frac{\boldsymbol{y^2}}{\boldsymbol{2\lambda}} \\ &\mathbf{y} \\ &\quad = \pm\sqrt{\frac{\boldsymbol{\lambda}\left(\mathbf{2ln}\boldsymbol{P}\left(\boldsymbol{\Omega_2}\right) - \mathbf{2ln}\boldsymbol{P}\left(\boldsymbol{\Omega_1}\right) - \mathbf{ln}\boldsymbol{\lambda}\right)}{\mathbf{1}-\boldsymbol{\lambda}}} \end{aligned} \quad (9)$$

$\mathbf{p}\left(\mathbf{y}|\,\boldsymbol{\Omega_1}\right)$ and $\boldsymbol{p}\left(\mathbf{y}|\,\boldsymbol{\Omega_2}\right)$ are conditional probability density functions of $\boldsymbol{\Omega_1}$ and $\boldsymbol{\Omega_2}$, respectively. $\boldsymbol{P}\left(\boldsymbol{\Omega_1}\right)$ and $\boldsymbol{P}\left(\boldsymbol{\Omega_2}\right)$ are the prior probability of $\boldsymbol{\Omega_1}$ and $\boldsymbol{\Omega_2}$, respectively**.** Therefore, the correct probability of Bayesian decision is as (10), as shown at the bottom of this page, where $\boldsymbol{\Phi}(\boldsymbol{y})$ is the standard normal distribution function, and $\mathbf{y} \geq \mathbf{0}$.

Evidently, when the prior probabilities of two classes are equal, $\boldsymbol{P}(\boldsymbol{\lambda})$ is an even function in logarithmic coordinates ($\mathbf{ln}\boldsymbol{\lambda}$). The variation of $\boldsymbol{P}(\boldsymbol{\lambda})$ is illustrated in Figure 2 in the case of $\boldsymbol{\lambda} > \mathbf{1}$. The figure shows that the eigenvectors of

$$\boldsymbol{P}(\boldsymbol{\lambda}) = \begin{cases} \boldsymbol{P}\left(\boldsymbol{\Omega_1}\right)\left(\mathbf{2}\boldsymbol{\Phi}\left(\boldsymbol{y}\right) - \mathbf{1}\right) + \mathbf{2}\boldsymbol{P}\left(\boldsymbol{\Omega_2}\right)\left(\mathbf{1} - \boldsymbol{\Phi}\left(\frac{\boldsymbol{y}}{\sqrt{\boldsymbol{\lambda}}}\right)\right) & (\boldsymbol{\lambda} > \mathbf{1}) \\ \boldsymbol{P}\left(\boldsymbol{\Omega_2}\right)\left(\mathbf{2}\boldsymbol{\Phi}\left(\frac{\boldsymbol{y}}{\sqrt{\boldsymbol{\lambda}}}\right) - \mathbf{1}\right) + \mathbf{2}\boldsymbol{P}\left(\boldsymbol{\Omega_1}\right)\left(\mathbf{1} - \boldsymbol{\Phi}\left(\boldsymbol{y}\right)\right) & (\boldsymbol{\lambda} \leq \mathbf{1}) \end{cases} \quad (10)$$

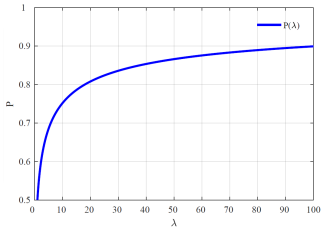


FIGURE 2. Probability of being correct (P) as a function of the eigenvalue (λ) when $\mathbf{P}(\Omega_1) = \mathbf{P}(\Omega_2)$.

larger eigenvalues should be selected when $\boldsymbol{\lambda > 1}$. Instead, those of smaller eigenvalues should be selected when $\boldsymbol{\lambda \leq 1}$.

B. COMMON CROSS SPECTRAL PATTERNS

Obviously, the above simultaneous diagonalization method only depends on the covariance matrices of two populations, while it loses all the time-frequency information. To more effectively measure the spectral and spatial correlations of signals, the power spectral density matrix is introduced to CSP.

1) CROSS-SPECTRAL DENSITY

In the spectral analysis of the time series, the cross-spectral density (CSD, or the cross-power spectrum) is used as part of the frequency-domain analysis of the cross-correlation between two time series. The cross-spectral density describes how the covariance of two data sets is distributed over the frequency domain. The cross-spectral density between the time series $x[n]$ and $y[n]$ is defined by

$$P_{xy}\left(e^{i\omega}\right) = \int_{m=-\infty}^{\infty} \tau_{xy}[m]\, e^{-i\omega m} \tag{11}$$

where $\tau_{xy}[m] = \int_{n=-\infty}^{\infty} x[n+m]\overline{y[n]}$ is the cross-correlation function of $x[n]$and $y[n]$. P_{xy} is a complex number. The cross-spectral density can be directly expressed as the product of the discrete Fourier transformations of those time series according to the time domain correlation theorem.

$$P_{xy}\left(e^{i\omega}\right) = X(e^{i\omega})\overline{Y(e^{i\omega})} \tag{12}$$

2) THE BAND CROSS-POWER MATRIX OF MULTIVARIATE TIME SERIES

The frequency domain characteristics of a multivariate time series $\mathbf{x}[\boldsymbol{n}]$ may be summarized by the power spectral density matrix (PSDM) in (11) [37].

$$\boldsymbol{P}\left(\boldsymbol{e}^{i\omega}\right) = \begin{bmatrix} \boldsymbol{P_{11}}\left(\boldsymbol{e}^{i\omega}\right) & \boldsymbol{P_{12}}\left(\boldsymbol{e}^{i\omega}\right) & \cdots & \boldsymbol{P_{1M}}\left(\boldsymbol{e}^{i\omega}\right) \\ \boldsymbol{P_{21}}\left(\boldsymbol{e}^{i\omega}\right) & \boldsymbol{P_{22}}\left(\boldsymbol{e}^{i\omega}\right) & \cdots & \boldsymbol{P_{2M}}\left(\boldsymbol{e}^{i\omega}\right) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{P_{M1}}\left(\boldsymbol{e}^{i\omega}\right) & \boldsymbol{P_{M2}}\left(\boldsymbol{e}^{i\omega}\right) & \cdots & \boldsymbol{P_{MM}}\left(\boldsymbol{e}^{i\omega}\right) \end{bmatrix} \tag{13}$$

where the diagonal elements contain the spectra of individual channels and the off-diagonal elements contain the cross-spectra. The matrix is a Gramian matrix.

However, we are more interested in the correlation of a particular frequency band. The band cross-power matrix (BCPM) is defined as follows.

$$\boldsymbol{P} = \int_{\boldsymbol{\omega_1}}^{\boldsymbol{\omega_2}} \left(\boldsymbol{P}\left(\boldsymbol{e}^{i\omega}\right) + \boldsymbol{P}\left(\boldsymbol{e}^{-i\omega}\right)\right) \boldsymbol{d\omega} \tag{14}$$

Since $\mathbf{x}[\mathbf{n}]$ is real-valued, its discrete time Fourier transformation is conjugate symmetric as $\boldsymbol{P_{ij}}\left(\boldsymbol{e}^{-i\omega}\right) = \overline{\boldsymbol{P_{ij}}\left(\boldsymbol{e}^{i\omega}\right)}$. Therefore, (12) can be rewritten by only using the positive frequency spectrum (or negative frequency spectrum) as follows.

$$\boldsymbol{P} = 2\mathrm{Re}\left(\int_{\boldsymbol{\omega_1}}^{\omega_2} \boldsymbol{P}\left(\boldsymbol{e}^{i\omega}\right) \boldsymbol{d\omega}\right) \tag{15}$$

where **Re** denotes the real part of a complex number. However, we suggest the following equation to calculate the cross-power matrix of the frequency band of interest.

$$\boldsymbol{P} = \int_{\boldsymbol{\omega_1}}^{\boldsymbol{\omega_2}} \boldsymbol{P}\left(\boldsymbol{e}^{i\omega}\right) \boldsymbol{d\omega} \tag{16}$$

The matrix is complex-valued, which comprehensively considers the effect of real and imaginary parts.

3) ESTIMATION OF THE BAND CROSS-POWER MATRIX

In a real-world application, how to estimate the spectral density is one of the key issues for calculating the cross-power matrix. It is well-known that CSP varies across time. Therefore, a feasible scheme is the periodogram method. The cross-spectral density is estimated by the single measurement FFT series. Then, it computes the band cross-power matrix of the single measurement. Finally, it averages the multiple repeat measurements as the band cross-power matrix of the population, as shown in Figure 3.

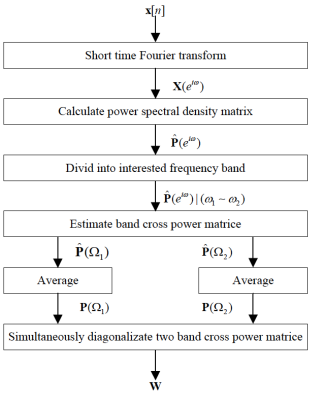


FIGURE 3. Calculative flow for the linear transformation vector (ω) of CCSP.

According to (12), the $\omega_1 \sim \omega_2$ band cross-power matrix (P) of each frame is estimated using the FFT series

as follows.

$$\widehat{\boldsymbol{P}_{ij}} = \frac{1}{N}\int_{n=Nf_1/fs}^{Nf_2/fs}\left(\overline{\boldsymbol{X}_i[n]}\boldsymbol{X}_j[n] + \overline{\boldsymbol{X}_i[-n]}\boldsymbol{X}_j[-n]\right) \tag{17}$$

where $X_i[n]$ and $X_j[n]$ are respectively the FFT series of the ith and jth channel signal, fs is sampling rate, $f_1 \sim f_2$ are the bandpass Hertz frequencies corresponding to $\omega_1 \sim \omega_2$ and $\widehat{P_{ij}}$ is the estimation of the i, j entry of $\hat{P}$. When the signal is real-valued, the band cross-power matrix of each frame can be estimated according to (15) and (16).

To eliminate magnitude variations between samples, the band cross-power matrix is calculated by averaging the normalized band cross-power matrix over all the frames of each population.

$$\boldsymbol{P} = \mathbb{E}\left(\frac{\hat{P}}{Trace(\hat{P})}\right) \tag{18}$$

where $Trace(\hat{P})$ *is* the sum of the diagonal elements of $\hat{P}$.

4) THE EXTRACTION OF LOGARITHMIC BAND POWER OF CCSP COMPONENTS

After the transformation$Y[n] = \omega^* X[n]$, we use the logarithmic band power of each frame (CCSP component) as a feature for recognition in the band of interest because the feature exhibits an approximately normal distribution.

The feature can be directly computed by averaging the FFT series.

$$f = ln\left(\frac{1}{N}\int_{n=Nf_1/fs}^{Nf_2/fs}|\boldsymbol{Y}[n]|^2\right) \tag{19}$$

When the estimation of the band cross-power matrix of each frame is known (i.e., training phase), a simple and effective method is to use the linear transformation property of the covariance matrix (see (2)). It can be calculated as follows:

$$f = ln(\omega^* \widehat{\Sigma}\omega) \tag{20}$$

Its complex multiplication times is BC (C + 1), and the direct calculation times is N(C + 1) (B is the filter band number, C is the channel number, and N is the signal length of each frame). Usually, BC << N (the indirect calculation) is more efficient. When the band cross-power matrix of each frame is unknown, its needs the NC (C + 1) = 2 times complex multiplication. Therefore, the direct calculation method should be proposed. The direct and indirect calculation processes for the features are shown in Figure 4.

III. EXPERIMENT AND RESULTS

We designed several experiments to validate the proposed CCSP feature extraction method and analyze the effect of CCSP component selection on spatial features.

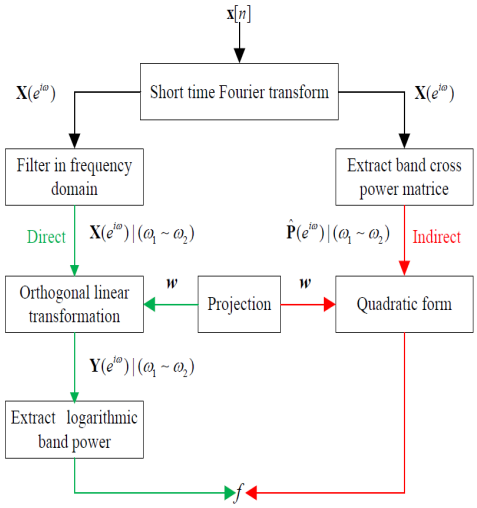


FIGURE 4. Linear transformation effects of PFCCSP. Two extraction flows of the CCSP features. Direct indicates the direct calculation of the logarithmic band power of CCSP (Green lines), while Indirect indicates the indirect calculation (Red lines).

A. CCSP PERFORMANCE COMPARED WITH OTHER METHODS

To evaluate the performance of CCSP, we used the 1989 Keirn and Aunon data set of the Brain-Computer Interfaces Laboratory in Colorado State University [38]. The data set contains 5 kinds of mental tasks performed by different age groups with eyes open and closed. In that study, a total of ten different experimental conditions were applied to verify the universality of the method.

1) 1989 KEIRN AND AUNON SET

Seven healthy volunteers participated in that study. Subject 1 and Subject 2 were employees of a university, and Subject 3 through Subject 7 were college students. Table 3-A1 lists the subjects' information (age, gender and handedness) and completed sessions. The subjects were seated in an Industrial

Subject	Sex [a]	Age	Handedness[b]	Condition [c]
S1	M	48	L	a, b
S2	M	39	R	a
S3	M	20-30	R	a, b
S4	M	20-30	R	a, b, c
S5	F	20-30	R	a, b
S6	M	20-30	R	a, b
S7	M	20-30	R	a

[a] M is male, and F is female;
[b] L is left handedness, R is right handedness;
[c] a is eye-open condition, b is eye-closed condition, and c is the second eye-open condition.

Acoustics Company sound-controlled booth with dim lighting and noiseless fans for ventilation. Electrodes were placed at C3, C4, P3, P4, O1, and O2 as defined by the 10-20 placement. Recordings were made with reference to electrically linked mastoids A1 and A2. The EOG was recorded between the forehead above the left brow line and

another on the left cheekbone. The electrodes were connected through a bank of Grass 7P511 amplifiers and bandpass filtered from 0.1-100 Hz. Data were recorded at a sampling rate of 250 Hz with a Lab Master 12-bit A/D converter mounted in an IBM-AT computer.

Data from all of the electrodes were recorded for 10 s during each task, and each task was repeated five times per session. The EEG signals for the subjects performing five different mental tasks were analyzed. Subjects performed five trials of each task in one day. They returned to do a second set of five trials on another day. There were a total of five distinct tasks, and each task was performed under both eyes open and eyes closed conditions. Therefore, a total of ten different experimental conditions were investigated. The following is a description of the tasks performed by each subject.

***B**aseline Measurements (**B**).* There was no mental task to be performed here. The subject was told to simply relax and to think of nothing in particular. This was done with both eyes opened and eyes closed. This task was used as a control and as a baseline measure of the EEG.

***M**ultiplication Solving (**M**).* The subject was given nontrivial multiplication problem to solve and, as in all of the tasks, was instructed to not vocalize or make overt movements while solving the problems. The problems were non-repeated and were designed so that an immediate answer was not apparent. The subject can verify at the end of the task whether or not he had arrived at a solution, The task would last for 10s.

*Geometric Figure **R**otation (R)* The subject was given 30s to study a drawing of a complex three-dimensional block figure, after which the drawing was removed, and the subject was instructed to visualize the object being rotated on an axis. The EEG was recorded during the mental rotation period.

*Mental **L**etter Composing (L)* The subject was instructed to mentally compose a letter to a friend or relative without vocalizing. Since the task was repeated several times, the subject was told to try to pick up where he left off in the previous task.

*Visual **C**ounting (C)* The subject was asked to imagine a blackboard and to visualize numbers being sequentially written on the board, with the previous number being erased before the next number was written. The subjects were further instructed to not verbally read the numbers but just to visualize them and to pick up counting from the previous task rather than starting over each time.

2) BINARY CLASSIFICATION OF TASKS

Because the CSP method is only suitable for a binary classification problem, we selected all the combinations of any two tasks under each condition for classification. Because each kind of mental task only contains 5 samples, the Leave-One-Out method was applied to obtain the unique recognition rate. To compare the performance of every feature extraction method, all the feature extraction methods adopted the logarithmic band power, and the effectiveness of various features is evaluated by the recognition rate of the SVM classifier.

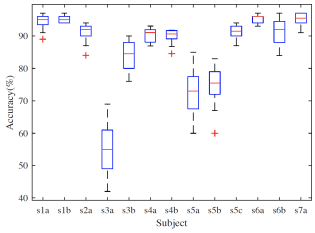


FIGURE 5. Average accuracy of each subject based on the FBP features across all task pairs in different kinds of frame shifts.

3) THE CONTRIBUTING FACTORS TO THE RECOGNITION PERFORMANCE

There are many factors affecting the recognition rate besides the feature extraction method itself. To objectively evaluate the quality of each feature extraction method, the following factors are considered in the paper.

Subject: The trials that were oversampled and empty trials were removed. Each subject's recognition rates of the frequency domain band power (FBP) cross-task pairs for the SVM are shown in Figure 5 under 32 kinds of frame shifts (8, 16,..., 256 point). Apparently, s3a represented an extreme subject sample, and it was therefore excluded from subsequent analyses in order to avoid the illusion of total recognition performance only because of the great recognition rate decrease of s3a.

Frequency Band Division: Frequency division is one of the key problems of feature extraction. To expediently compare the results with previous results, we adopted the following 6 frequency bands that are widely accepted in clinical practice: δ (2-4 Hz), θ (4-7 Hz), α (8-15 Hz), β (16-31 Hz), $\gamma 1$ (32-55 Hz) and $\gamma 2$ (65-100 Hz). The γ band is divided into $\gamma 1$ and $\gamma 2$ to eliminate the artifact of 60 Hz commercial power.

Frame and Vote: EEG data are usually split into a number of fixed-length frames. Long frames cause long delays in results, while short frames affect the spectral resolution of the low-frequency band. The frame length was set to 256 points with the time delay of approximately 0.5 s. Therefore, the low-frequency band of δ, θ contained at least $3\sim4$ frequency points under the sampling rate of 250 Hz. Frame shift is the main factor affecting the recognition rate. Too large of a frame shift causes too irregular recognition results and seriously affects the recognition performance, while too short of a frame shift leads to long processing and training time because of excessive redundant samples. The recognition rate of the SVM classifier on the logarithmic band power (FBP) with the increase in the frame shift is shown in Figure.6.

4) COMPARISON BETWEEN CSP, FBCSP, CCSP, AND FBP

In the FBP (Logarithmic band power in the frequency domain), each channel of the EEG was short-time Fourier transformed, and then the logarithmic power of each frame was extracted in the band of interest. In CSP (Common spatial

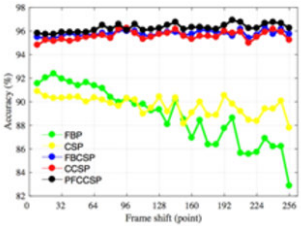


FIGURE 6. Average accuracy based on the features of FBPs, CSP, FBCSP, CCSP and PFCCSP in different kinds of frame shifts.

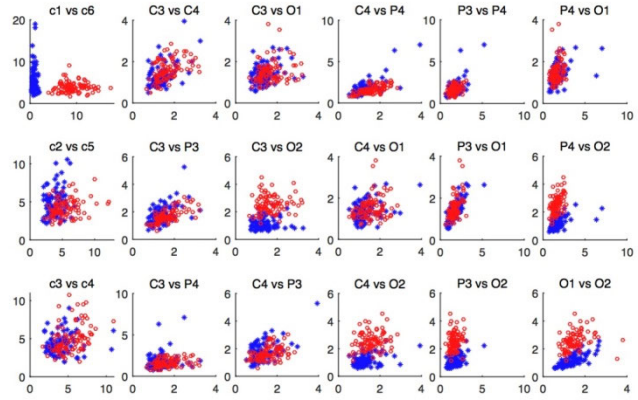


FIGURE 7. Linear transformation effects of PFCCSP. The distribution of the logarithmic band power (16-31 Hz) of the CSP recorded for subject 7 during complex problem multiplication and solving and geometric figure rotation. Blue * represents a sample of complex problem multiplication task and red circle represents a sample of geometric figure rotation. C3, C4, P3, P4, O1 and O2 are electrode positions. c1, c2, c3, c4, c5, and c6 are the components after the transformation sorted by eigenvalues. Among them, c1 corresponds to the minimum eigenvalue, and c6 corresponds to the maximum eigenvalue. The first column on the left shows the sample distribution of the components after the transformation, and the rest shows the sample distribution of any two electrodes.

patterns), the EEG was notched filtered from 55-65 Hz in order to eliminate the artifact of commercial power. Then, it was transformed using CSP, and it finally extracted the logarithmic power of all components of each frame. In the FBCSPs (Common spatial patterns in filter-bank), the EEG was filtered into bands of interest and transformed using CSP. Finally, it extracted the logarithmic power of all components of each frame. In CCSP (Common cross spectral patterns), each channel of the EEG was short-time Fourier transformed. Then, the frequency spectrum was transformed using CCSP in the band of interest. Finally, it extracted the logarithmic power of all components of each frame. In PFCCSP (Positive frequency common cross-spectral patterns), its algorithm is the same as the CCSP, except it only accounts for the positive frequency spectrum of each EEG frame according to 14. The effect of PFCCSP's linear transformation is shown in Figure 7. Obviously, the two components corresponding to the maximum and minimum eigenvalues have the strongest separability, which are respectively distributed along their own coordinate axis.

We used paired t-test to validate that the CCSP method and PFCCSP method had an influence on recognition rates. the degrees of freedom were corrected by the Greenhouse-Geisser method. The result is listed in Table 3-A4. There is a significant main effect of Method on recognition rates. And a significant interaction between Method and Frame shift on the recognition rates. Because of a significant correlation in the data determined by the Mauchly sphericity test, the application of CCSP can indeed improve the recognition performance. The significant interaction is mainly caused by the improvement of FBP recognition rates of CSP, considering that the frequency information has 4.615∼5.202% increases and are significantly higher than that of FBPs. The average recognition rate of PFCCSP is indeed higher than that of CCSP and FBCSP, and the p-value 0.074 for FPCCSP-FBCSP is close to the significant level.

Pair	Mean	t	df	p
CSP-FBP	0.739	0.501	11	0.626
CCSP - CSP	0.282	0.975	11	0.001*
FBCSP - FBP	4.826	4.374	11	0.001**
CCSP-FBP	4.615	4.584	11	0.001**
PFCCSP - FBP	5.202	5.171	11	<0.001**
PFCCSP - CCSP	0.587	2.625	11	0.024*
PFCCSP - FBCSP	0.377	1.975	11	0.074

Paired t-test on the recognition rate of FBPs, CSP, FBCSP, CCSP and PFCCSP. * denotes $p< 0.05$, ** denotes $p< 0.01$

Figure 6 shows the recognition rate comparison among five different methods. A three-way repeated measures ANOVA was used to evaluate the effects of Method (5 methods), Frame shift (32 different frame shifts), and Task pair (10 task pairs) on each subject's recognition rate. The results are listed in Table 1.

TABLE 1. Three-way repeated-measures ANOVA.

Factor	F	df	rdf	p	$\partial(\eta^2)$
M	16.105	4,44	1.959,21.553	0.001**	0.594
F	2,518	15,165	2.882,31.698	0.078	0.186
T	0.459	9,99	3.634,39.972	0.748	0.04
M×F	3.813	60,660	4.763,51.405	0.006**	0.257

M, F and T respectively represent the factors of Method, Frame shift and Task pair. * denotes $p< 0.05$, and ** denotes $p< 0.01$.

TABLE 2. Two-way repeated-measures ANOVA.

Factor	F	df	rdf	p	$\partial(\eta^2)$
F	2.374	31,341	4.640,51.040	0.056	0.177
T	0.228	9,99	3.756,41.311	0.912	0.020
F×T	1.105	279,3069	6.326,69.582	0.369	0.091

Two-way repeated measures ANOVA on the recognition rate of PFCCSP using the Greenhouse-Geisser method to correct the degrees of freedom. F and T respectively represent the factors of Frame shift and Task pair.

5) BETTER RESPONSE TIME AND UNIVERSALITY OF CCSP

The recognition rate of PFCCSP fluctuates very little with the frame shift from Figure 6. Thus, the recognition rates of frame shifts between 8 and 256 were selected for the two-way repeated measures ANOVA (Frame shift (32 different frame shifts), Task pair (10 task pairs)). The result is in Table 2.

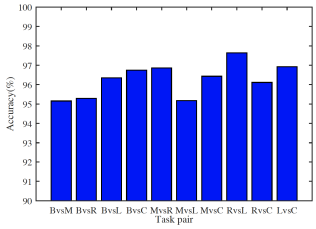


FIGURE 8. Average accuracy of the PFCCSP features in different task pairs.

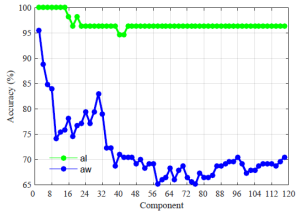


FIGURE 9. Recognition performance of selecting CCSP components.

Similarly, the Mauchly sphericity test shows that there is a significant correlation in the data, and the degrees of freedom need to be corrected by the Greenhouse-Geisser method. There only exists a tending to be a significant main effect of Frame shift from the statistical analysis results in Table 2. The recognition rates of each kind of frame shift are between 94.813% to 96.239%. Obviously, the logarithmic band power of the PFCCSP component is not sensitive to frame shifts. Therefore, we can use larger frame shifts to reduce unnecessary processing time, which brings better response time. The recognition rate of each task pair is between 95.156% and 97.639%, as shown in Figure 8. Because the significance of Task pair ($p = 0.912$) is greater than 0.9, we can assume that the PFCCSP method has the same recognition performance on all the tasks. Therefore, the method has a certain universality. Table 3 shows the average recognition rates in each subject across all task pairs under the 200 (points) frame shift. Compared with the results of Z.A. Keirn's method, the average recognition rates of the majority of the subjects were improved.

TABLE 3. Average accuracy across tasks of each subject.

Subject and Condition	PFCCSP	FBCSP	Z. Keirn [12]
s1a	**99**	99	99
s1b	**98**	98	98
s2a	99		
s3b	**97**	97	97
s4a	97.1	**96.5**	**98**
s4b	**96.9**	96.8	96.9
s5a	92	91.5	
s5b	91	92	97.5
s5c	**98**	97.5	97
s6a	**98**	97.7	97
s6b	**100**	98	96.5
s7a	97	96.5	

The frame shift is 200 points.

B. EXPERIMENT ON COMPONENT SELECTION IN CCSP

According to Equation 10, the separability of the CCSP component is closely related to the corresponding eigenvalue. We chose the dataset IVa of the BCI Competition III [10] to verify the effect of CCSP component selection on the spatial features and show the time-invariant spatial filtering pattern of CCSP.

1) DATASET DESCRIPTION

The dataset we used to test Component selection capability is Dataset IVa from BCI COMPETITION III. This dataset was recorded from five healthy subjects. Subjects sat in a comfortable chair with their arms resting on armrests. This dataset contains only data from the 4 initial sessions without feedback. Visual cues at 3.5 s indicated which of the following 3 motor imagery tasks the subject should perform: (L) left hand, (R) right hand, or (F) right foot. The presentation of target cues was intermitted by periods of random length from 1.75 to 2.25 s during which the subject could relax. There were two types of visual stimuli: (A) targets were indicated by letters appearing behind a fixation cross (which might nevertheless induce little target-correlated eye movements), and (B) a randomly moving object indicated targets (inducing target-uncorrelated eye movements). For subjects al and aw, 2 sessions of both types were recorded, while 3 sessions of stimulus type (1) and 1 session of stimulus type (A) were recorded for each of the other subjects. The recordings were made using BrainAmp amplifiers and a 128 channel Ag/AgCl electrode cap from ECI. Moreover, 118 EEG channels were measured at positions of the extended international 10/20-system. Signals were bandpass-filtered between 0.05 and 200 Hz and then digitized at 1000 Hz with 16-bit (0.1 μV) accuracy.

2) DATASET PREPROCESSING AND RECOGNITION RESULTS

Subjects al (training data 80%) and aw (training data 20%) were selected to test the effect of selecting PFCCSP components on different ratios of training data to testing data. The data were down-sampled to 100 Hz, and the data after visual cues 0.5-3.5 s (300 points) were analyzed. The length of the frame was set as 100 (points), and the frame shift was 50 (points). The logarithmic power of each component on the band associated with motor imagery [39] was extracted by the PFCCSP method for an SVM classifier. The recognition results of the right hand and foot motor imagery on its test set are shown in Figure 9. From the figure, PFCCSP can indeed map the main difference between two classes of data to a few components (especially in the case of high-density EEG) and choose more discrinative features. However, compared with *al*, the accuracy of *aw* with small training data fluctuates along with the increase of components. This phenomenon is mainly due to the variability of EEG signals across time, and the test samples are selected from the final samples. It is difficult to estimate the band cross-power matrix of each kind of task when the training set is small. Compared with the two top winners (Yanjun Wang: *al*, 100%; *aw*, 100% Yuangqing Li: *al*, 98.2%, *aw*, 92.4%), the single feature of CCSP can reach a better recognition performance.

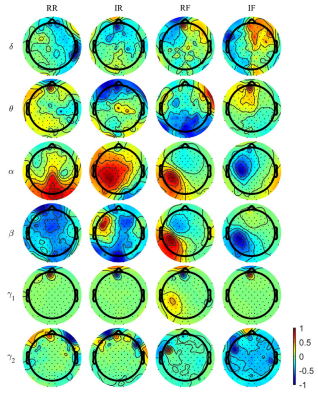


FIGURE 10. Spatial pattern of W^{-1} from subject aw, during the left-foot and right-hand motor imagery task. RR represents the real part of the maximum variance column of $W^{-1}W^{-1}$ for left-foot motor imagery and IF represents the imagery part respectively.δ, θ, α, β, γ represent corresponding EEG frequency bands.

C. SPATIAL PATTERN OF W^{-1}

Each column of $\boldsymbol{W^{-1}}$ can be seen as a time-invariant EEG source distribution vector and reflects the mapping weights of each CCSP source on all scalp electrodes. Figure 10 shows the spatial patterns of PFCCSP on the band when al was performing the right hand and foot motor imagery tasks. RR and IR are respectively the real part and the imaginary part of the column $\boldsymbol{W^{-1}}$ that corresponds to the maximum variance component for right-hand motor imagery. RF and IF are respectively the real and imaginary part of column $\boldsymbol{W^{-1}}$that corresponds to the maximum variance component for the right-foot motor imaginary. For the right-foot motor imagery task, the pattern is most obvious at $\boldsymbol{\alpha}$ band, mainly the left, middle and rear electrodes (RF and IF), which may be attributed to the sensorimotor cortex activation. At the high-frequency $\boldsymbol{\gamma}$band, the pattern is distinguishable as the lower-frequency bands. In addition, the real and imaginary parts of the transformation matrix are not completely independent and have a certain degree of similarity, which determines the similar recognition results of CCSP and PFCCSP in Figure 6. There is more difference between two distributions in complex space than in real space. Meanwhile, the real space is only a projection of the complex domain space, which reduces the distinction between the two classes. Therefore, by removing the real part, the recognition effect of PFCCSP must be better than that of CCSP.

IV. CONCLUSION

This paper proposes a CCSP-based feature extraction method for cognitive task classification. This method has several strengths. First, the method uses the cross-power matrix to measure the band correlation of signals. Then, the orthogonal transformation is constructed according to the simultaneous congruence diagonalization of the matrices. Finally, the logarithmic band power of the transformed components is extracted. In this paper we also described A variation of proposed CCSP method, Positive Frequency Common Cross-Spectral Patterns(PFCCSP). The PFCCSP method further improves the performance when the spectrum variability is more significant in EEG of mental tasks.

In the frequency domain, the filter banks can be flexibly chosen for extracting the cross-power. The features extracted with the method are more distinguishable. Its spatial pattern clearly reflects the difference of active brain regions in different mental tasks. The CCSP-based method avoids the calculation of the correlation function by using short time Fourier transformation, which effectively reduces the computation complexity. Experiments on public datasets show that considering various effect factors, this CCSP-based feature extraction method has better performance compared with method based on the channel band-power. The CCSP-based method performed well across frame shifts and cognitive tasks, which makes it more responsive in time and promises a broader field of application.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers who made constructive comments.

REFERENCES

[1] D. Galin and R. Ornstein, "Lateral specialization of cognitive mode: An EEG study," *Psychophysiology*, vol. 9, no. 4, pp. 412–418, Jul. 1972.
[2] J. C. Doyle, R. Ornstein, and D. Galin, "Lateral specialization of cognitive mode: II. EEG frequency analysis," *Psychophysiology*, vol. 11, no. 5, pp. 567–578, Sep. 1974.
[3] G. Pfurtscheller and C. Neuper, "Motor imagery activates primary sensorimotor area in humans," *Neurosci. Lett.*, vol. 239, nos. 2–3, pp. 65–68, Dec. 1997.
[4] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *Proc. IEEE*, vol. 89, no. 7, pp. 1123–1134, Jul. 2001.
[5] G. Pfurtscheller, G. Müller-Putz, R. Scherer, and C. Neuper, "Rehabilitation with brain-computer interface systems," *Computer*, vol. 41, no. 10, pp. 58–65, Oct. 2008.
[6] X. Lou, S. Xiao, Y. Qi, X. Hu, Y. Wang, and X. Zheng, "Corticomuscular coherence analysis on hand movement distinction for active rehabilitation," *Comput. Math. Methods Med.*, vol. 2013, Apr. 2013, Art. no. 908591.
[7] D. Huang, K. Qian, D. Fei, W. Jia, X. Chen, and O. Bai, "Electroencephalography (EEG)-based brain-computer interface (BCI): A 2-D virtual wheelchair control based on event-related desynchronization/synchronization and state control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 3, pp. 379–388, May 2012.
[8] S. C. Wriessnegger, C. Brunner, and G. R. Müller-Putz, "Frequency specific cortical dynamics during motor imagery are influenced by prior physical activity," *Frontiers Psychol.*, vol. 9, p. 1976, Oct. 2018.
[9] B. Blankertz, K.-R. Müller, G. Curio, T. Vaughan, G. Schalk, J. Wolpaw, A. Schlogl, C. Neuper, G. Pfurtscheller, T. Hinterberger, M. Schroder, 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. 1044–1051, Jun. 2004.
[10] B. Blankertz, K.-R. Müller, D. Krusienski, G. Schalk, J. Wolpaw, A. Schlogl, G. Pfurtscheller, J. Millan, M. Schroder, and N. Birbaumer, "The BCI competition III: Validating alternative approaches to actual BCI problems," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 153–159, Jun. 2006.
[11] M. Tangermann, K. R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. J. Miller, G. R. Müller-Putz, G. Nolte, G. Pfurtscheller, H. Preissl, G. Schalk, A. Schlögl, C. Vidaurre, S. Waldert, and B. Blankertz, "Review of the BCI competition IV," *Frontiers Neurosci.*, vol. 6, pp. 1–31, Jul. 2012.
[12] Z. Keirn and J. Aunon, "A new mode of communication between man and his surroundings," *IEEE Trans. Biomed. Eng.*, vol. 37, no. 12, pp. 1209–1214, Dec. 1990.

[13] M. Alansari, M. Kamel, B. Hakim, and Y. Kadah, "Study of wavelet-based performance enhancement for motor imagery brain-computer interface," in *Proc. 6th Int. Conf. Brain-Comput. Interface (BCI)*, Jan. 2018, pp. 1–4.
[14] H. U. Amin, W. Mumtaz, A. R. Subhani, M. N. M. Saad, and A. S. Malik, "Classification of EEG signals based on pattern recognition approach," *Frontiers Comput. Neurosci.*, vol. 11, p. 103, Nov. 2017.
[15] R. Palaniappan, "Utilizing gamma band to improve mental task based brain-computer interface design," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 3, pp. 299–303, Sep. 2006.
[16] L. Zhang, W. He, C. He, and P. Wang, "Improving mental task classification by adding high frequency band information," *J. Med. Syst.*, vol. 34, no. 1, pp. 51–60, Feb. 2010.
[17] Z. J. Koles, M. S. Lazar, and S. Z. Zhou, "Spatial patterns underlying population differences in the background EEG," *Brain Topogr.*, vol. 2, no. 4, pp. 275–284, 1990.
[18] G. F. Alpert, R. Manor, A. B. Spanier, L. Y. Deouell, and A. B. Geva, "Spatiotemporal representations of rapid visual target detection: A single-trial EEG classification algorithm," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 8, pp. 2290–2303, Aug. 2014.
[19] X. Yu, P. Chum, and K.-B. Sim, "Analysis the effect of PCA for feature reduction in non-stationary EEG based motor imagery of BCI system," *Optik*, vol. 125, no. 3, pp. 1498–1502, Feb. 2014.
[20] Y. Ma, X. Ding, Q. She, Z. Luo, T. Potter, and Y. Zhang, "Classification of motor imagery EEG signals with support vector machines and particle swarm optimization," *Comput. Math. Methods Med.*, vol. 2016, pp. 1–8, May 2016.
[21] R. Zhang, P. Xu, T. Liu, Y. Zhang, L. Guo, P. Li, and D. Yao, "Local temporal correlation common spatial patterns for single trial EEG classification during motor imagery," *Comput. Math. Methods Med.*, vol. 2013, pp. 1–7, Nov. 2013.
[22] Z. Koles, "The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG," *Electroencephalogr. Clin. Neurophysiol.*, vol. 79, no. 6, pp. 440–447, Dec. 1991.
[23] Z. Koles, J. Lind, and A. Soong, "Spatio-temporal decomposition of the EEG: A general approach to the isolation and localization of sources," *Electroencephalogr. Clin. Neurophysiol.*, vol. 95, no. 4, pp. 219–230, Oct. 1995.
[24] A. Soong and Z. Koles, "Principal-component localization of the sources of the background EEG," *IEEE Trans. Biomed. Eng.*, vol. 42, no. 1, pp. 59–67, Jan. 1995.
[25] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," *Clin. Neurophysiol.*, vol. 110, no. 5, pp. 787–798, May 1999.
[26] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Classification of movement-related EEG in a memorized delay task experiment," *Clin. Neurophysiol.*, vol. 111, no. 8, pp. 1353–1365, Aug. 2000.
[27] C. Guger, H. Ramoser, and G. Pfurtscheller, "Real-time EEG analysis with subject-specific spatial patterns for a brain-computer interface (BCI)," *IEEE Trans. Rehab. Eng.*, vol. 8, no. 4, pp. 447–456, Dec. 2000.
[28] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 441–446, Dec. 2000.
[29] Q. Novi, C. Guan, T. H. Dat, and P. Xue, "Sub-band common spatial pattern (SBCSP) for brain-computer interface," in *Proc. 3rd Int. IEEE EMBS Conf. Neural Eng.*, May 2007, pp. 204–207.
[30] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *Proc. Int. Joint Conf. Neural Netw.*, 2008, pp. 2390–2397.
[31] H. Zhang, Z. Y. Chin, K. K. Ang, C. Guan, and C. Wang, "Optimum spatio-spectral filtering network for brain-computer interface," *IEEE Trans. Neural Netw.*, vol. 22, no. 1, pp. 52–63, Jan. 2011.
[32] M. Villena-González, I. Palacios-García, E. Rodríguez, and V. López, "Beta oscillations distinguish between two forms of mental imagery while gamma and theta activity reflects auditory attention," *Frontiers Hum. Neurosci.*, vol. 12, p. 389, Sep. 2018.
[33] G. N. Dimitrakopoulos, I. Kakkos, Z. Dai, J. Lim, J. J. Desouza, A. Bezerianos, and Y. Sun, "Task-independent mental workload classification based upon common multiband EEG cortical connectivity," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 11, pp. 1940–1949, Nov. 2017.
[34] A. L. Gardony, M. D. Eddy, T. T. Brunyé, and H. A. Taylor, "Cognitive strategies in the mental rotation task revealed by EEG spectral power," *Brain Cognition*, vol. 118, pp. 1–18, Nov. 2017.
[35] D. Gheza, R. De Raedt, C. Baeken, and G. Pourtois, "Integration of reward with cost anticipation during performance monitoring revealed by ERPs and EEG spectral perturbations," *NeuroImage*, vol. 173, pp. 153–164, Jun. 2018.
[36] Y. He, M. Steines, J. Sommer, H. Gebhardt, A. Nagels, G. Sammer, T. Kircher, and B. Straube, "Spatial–temporal dynamics of gesture–speech integration: A simultaneous EEG-fMRI study," *Brain Struct. Function*, vol. 223, no. 7, pp. 3073–3089, Sep. 2018.
[37] S. L. Marple, *Digital Spectral Analysis: With Applications*, vol. 86. Upper Saddle River, NJ, USA: Prentice-Hall, 1987.
[38] Z. A. Keirn, "Alternative modes of communication between man and machine," M.S. thesis, Purdue Univ., West Lafayette, IN, USA, 1988.
[39] 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, Jun. 2004.

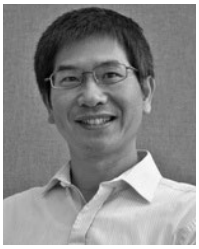
HAIFENG LI (Member, IEEE) received the Ph.D. degree in electro-magnetical measuring technique and instrumentation from the Harbin Institute of Technology (HIT), in 1997, and the Ph.D. degree in computer, communication, and electronic science from the University of Paris 6, France, in 2002. He started the teaching career with HIT, in 1994, where he was promoted as a Lecturer, in 1995, and a Professor, in 2003. From 1997 to 2002, he was involved in the Postdoctoral research with the University of Paris 6. He is currently a Professor with the School of Computer Science and Technology, HIT, and the Director of Cogniputing Lab (cognitive computing). His research interests include intelligent information processing, cognitive computing and artificial brain, audio information processing, and natural human–computer interaction. He is a member of the CCF and a Senior Member of the China Computer Federation.

SHANG FENG (Senior Member, IEEE) received the B.Sc. degree in communication engineering and the M.Sc. degree in software engineering from the Harbin Institute of Technology (HIT), China, and the M.Sc. degree from the UniversitÀ degli Studi di Pavia, Italy. He is currently pursuing the Ph.D. degree with the HIT. From 2015 to 2017, he was a Visiting Scholar with the Swartz Center of Computational Neuroscience, University of California San Diego. His main research interests include brain-computer interface, biomedical engineering, human cognition, and EEG signal processing.

LIN MA received the M.E. degree in computer science from the University of Paris 6, France, in 2001, and the Ph.D. degree in computer science from the Harbin Institute of Technology (HIT), Harbin, China, in 2008. She is currently an Associate Professor with the School of Computer Science and Technology, HIT, where she has been with the Biocomputing Research Center of Computer Science and Technology School, since 2002. Meanwhile, she was a Visiting Scholar with The Hong Kong Polytechnic University supported by Hong Kong Croucher Funding, in 2006 and 2011. Her current research interests include intelligent information processing, image processing, and cognitive science.

ZHONGLIANG XU received the B.E. degree in bioinformatics technology from the Harbin Institute of Technology, China, in 2012. He is currently pursuing the master's degree with Northeast Forestry University, China. His main research interests include EEG signal processing, sparse representation, compressed sensing, and brain-computer interface.

RUIFENG XU received the B.Eng. degree from the Harbin Institute of Technology (HIT), in 1995, and the M.Phil. and Ph.D. degrees from The Hong Kong Polytechnic University, in 2001 and 2006, respectively. He is currently a Professor and a Doctoral Supervisor with the Shenzhen Graduate School, HIT. His research interests include natural language processing, emotion computing, and text mining.



TZYY-PING JUNG (Fellow, IEEE) received the B.S. degree in electronics engineering from National Chiao Tung University, Hsinchu, Taiwan, in 1984, and the M.S. and Ph.D. degrees in electrical engineering from The Ohio State University, Columbus, OH, USA, in 1989 and 1993, respectively. He is an Associate Director of the Swartz Center for Computational Neuroscience, Institute for Neural Computation, University of California San Diego, where he is also an Adjunct Professor of bioengineering. His research interests include biomedical signal processing, cognitive neuroscience, machine learning, human EEG, functional neuroimaging, and brain-computer interfaces and interactions.