# Based on Stockwell transform algorithm and signal collection to efficient motion imaging of a brain–computer interface

Zeyu Zhao, Jimin Yang*

School of Physics and Electronics, Shandong Normal University, Jinan 250358, People's Republic of China

E-mail: jmyang@sdnu.edu.cn (J Yang)

## Abstract

A brain–computer interface (BCI) equips humans with the ability to control computers and technical devices mentally. The research based on the motor imagery BCI system was carried out in terms of the pretreatment, feature extraction, classification and identification methods of electroencephalography (EEG) signals. However, the enormous data and the existing irrelevant features of the electrocorticogram signal limit the performance of the classifier. In the signal pretreatment part, the Least Mean Square (LMS) adaptive filtering algorithm was employed for the filtering. Fast Independent Component Analysis (FastICA) were employed to reduce the dimensionality. The Stockwell transform (ST) in the feature extraction.The classification and identification was conducted by using Bayesian Linear Discriminant Analysis (BLDA) in order to achieve the further classification.The experimental results show that the best classification result (90%) can be obtained when the LMS adaptive filtering and FastICA dimensionality reduction were initially used for the pretreatment of the EEG signals, and then the ST was used to extract the EEG features. Finally, the EEG signals were reclassified by theBLDA identification method.

## Keywords

Brain Computer Interface, Motion Imaginary, Fast Independent Component Analysis, Stockwell transform, Bayesian Linear Discriminant Analysis.

## 1. Introduction

Brain Computer Interface (BCI) is an emerging human-computer interaction mode. A channel for direct information transmission between the brain and the computer or other external devices can be constructed without depending on the normal output path of the brain [1-3]. In 1973, the term BCI was firstly proposed in the article by Vidal *et al*. from the University of California in the United States, and the basic framework of the BCI system was also introduced in detail [4]. As shown in Figure 1, the typical BCI system is generally composed of four parts, including signal collection, feature extraction, classification and identification (i.e. algorithm translation) and output commands controlling. Detection of the corresponding signal changing in the brain is extracted and used as a characteristic signal for future actions. By means of classifying these characteristic signals, the human thinking activities can be directly converted into instructions for driving external devices [5].

The motor imagerybased BCI system (MI–BCI) system[6] mainly takes the electroencephalogram (EEG) which is induced by motor imagination stimulating the rhythmic variation of cerebral cortex, as input and judges the type of motor imagination through signal processing. Then, the computer translates the type of MI into control commands that ultimately enable communication and control functions between the human brain and external devices[7]. A crucial technique of MI–BCI is its efficient decoding around the motor cortex, that is, the accurate extraction of perceptionmovement rhythmic variation characteristics and the correct classification of MI tasks[8].

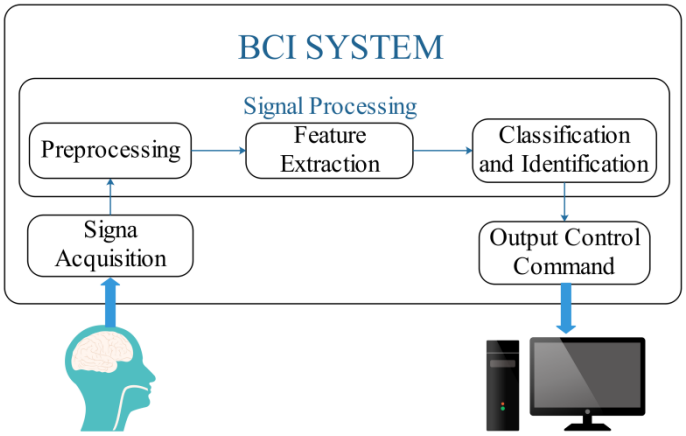



Fig.1 The basic composition of the BCI system

While there are many methods to represent EEG signals, the most common representation methods include time-point features, amplitude values of EEG signals[9] , power spectral density (PSD) values[10], and time-frequency features[11]. Time-point and band-power features are widely used in BCIs that utilize oscillatory activity, and they can be extracted from spatial filtering[12], principal component analysis (PCA), independent component analysis[13], common spatial patterns (CSP) [14]. It should be pointed out that Xu et al[10] introduced the Stockwell transform (ST) and modified ST (MST) in the feature extraction of ECoG signals to accurately locate the time-frequency information containing the changes in the perceptual motion rhythm and calculated the PSD of the transformed signal. Compared with other time-frequency methods, the ST has a good energy concentration, and can obtain better time-frequency distribution and a higher-resolution spectral density function [10].

Many of the characteristics of a BCI system depend critically on the employed machine-learning algorithm. Baysian LDA (BLDA) is fully automated, that is, no user intervention is needed to adjust the hyper parameters, and the computation of the classifier on a standard PC takes less than a minute[15]. BLDA can be regarded as a derivation of FLDA. Compared with FLDA, in BLDA, regularization is adopted to keep overfitting of high-dimensional and possibly noisy datasets. Owing to its merit of dicarding of time-consuming cross-validation, Bayesian analysis is used to automatically and quickly estimate the degree of regularization from the training data.

In this paper, a framework of MI–BCI system was constructed. We proposed the ST to extract features and calculate the PSD of the transformed signals. Then, we classified them using the BLDA algorithm. According to the referred system, collect EEG signals by ourselves was utilized to verify the validity of the algorithm. Through analysis and algorithm-based processing proposed in this paper, we achieved an accuracy of 90%, which proved that the algorithms were effective.The outline of the study is as follows. Section 2 describes preprocessing, feature extraction,classification. Sections 3 and 4 present and discuss the experiment and results. Finally, conclusions in section 5.

## 2. Method

### 2.1 Preprocessing

In this preprocessing part, the LMS adaptive algorithm (see 2.1.1) was firstly used for the filtering, in order to remove noise and artifact interference, such as EOG, EMG and etc., from the EEG signal. Then, FastICA (see 2.1.2) was used to perform the dimensionality reduction to reduce data storage and improve algorithm efficiency [16].

2.1.1 Least mean square (LMS) adaptive algorithm

The LMS algorithm is an algorithm theory that is widely used in adaptive filtering algorithms. It is developed on the basis of Wiener filtering and borrowed from the idea of the steepest descent method. It has been widely applied in information processing technology [17]. A notable feature of the LMS algorithm is its simplicity without the requirement of computing correlation functions or matrix

inversion. Its application principle is minimizing the mean-square error (MMSE) between the actual output value of the filter and the expected response [18].

When the input of the horizontal adaptive digital filter is set as $x(n)$, the ideal output is $d(n)$, and the actual output is $y(n)$, and the weighting coefficient of the filter is $w_i(n)(i=0,1,\cdots,M-1)$, then the LMS algorithm [18] can be conducted as follow:

$$y(n)=\sum_{i=0}^{M-1} w_i(n)x(n-i) \tag{1}$$

$$e(n)=d(n)-y(n) \tag{2}$$

$$w_i(n+1)=w_i(n)+2\mu e(n)x(n-i)\ ,\ i=0,1\cdots,M-1,$$

$$\mu \text{ is Convergencefactor} \tag{3}$$

2.1.2 Fast Independent Component Analysis (FastICA)

There are lots of categories of ICA algorithms available. FastlCA is widely applied in signal processing, due to its fast convergence speed and good separation performance. This algorithm is based on a fixed-point recursive algorithm and its existence enables the ICA analysis of data with high-dimensionality to become possible. When the number of iterations is set as $w_p$, fixed-point algorithm [19] is conducted as follow:

$$\text{Initial weight vector } w_p \leftarrow E[zg(w_p^T z)]-E[g'(w_p^T z)]w_p \tag{4}$$

Where, $E[\cdot]$ is the mean calculation, $g(\cdot)$ is a linear function $g(y)=-\exp(-y^2/2)$, and $z$ is a data whitening vector.

$$w_p=w_p-\sum_{j=1}^{p-1}(w_p^T w_j)w_j \tag{5}$$

Successive orthogonal

$$w_p \leftarrow w_p / \left\| w_p \right\| \tag{6}$$

FastICA has independent attributes of dimensionality reduction and extraction. Here, independence refers to no relationship, while irrelevant means non-linear relationship. FastICA is used to make each component independent and improve the efficiency of finding hidden factors [20].

**2.2 Feature extraction**

The ST is an extension of the ideas of the continuous wavelet transform (CWT), which is based on a moving and scalable localizing Gaussian window. It has some desirable characteristics that are absent in the CWT. The ST is unique as it provides

frequency-dependent resolution while maintaining a direct relationship with the Fourier spectrum. These advantages of the ST are owing to the fact that the modulating sinusoids are fixed with respect to the time axis, whereas the localizing scalable Gaussian window dilates and translates[21-22].

We can obtain the Fourier transform of a discrete time series $x[kT]$ as follows:

$$X[\frac{n}{NT}]=\frac{1}{N}\sum_{k=0}^{N-1} x[kT]e^{-\frac{i2\pi k}{N}} \tag{7}$$

where $k,n=0,1,\cdots,N-1$ and $T$ is the time sampling interval. It is obvious that ST is the projection of the vector defined by the time series $x[kT]$ onto a spanning set of vectors in the discrete case. For every basis vector of the Fourier transform, it is decomposed into N localized vectors, which are individually multiplied using the N-shifted Gaussians. Moreover, the sum of these N localized vectors

is identical to the original basis vector. The ST of a discrete time series $x[kT]$ is given by (let $f \to n/NT$ and $\tau \to jT$ ),

$$S[jT, \frac{n}{NT}] = \sum_{m=0}^{N-1} X[\frac{m+n}{NT}] e^{-\frac{2\pi^2 m^2}{n^2}} e^{\frac{i2\pi mj}{N}} \tag{8}$$

where $j, m = 0,1,\cdots, N-1$ and $n = 1,2,\cdots, N-1$. For $n = 0$, the ST is a constant equaled to

$$S[jT,0] = \frac{1}{N} \sum_{m=0}^{N-1} x(\frac{m}{NT}) \tag{9}$$

The PSD can describe how the power of a signal or time series is distributed over frequencies [48]. The PSD of the ST is defined as the mathematical expectation of the square of the ST absolute value of the time series $x[kT]$, representing a local power measure in the time-frequency domain, which is

$$S_{xx}[jT, \frac{n}{NT}] = E\{S_x[jT, \frac{n}{NT}] S_x^*[jT, \frac{n}{NT}]\} \tag{10}$$

where $S_x[jT, \frac{n}{NT}]$ and $S_x^*[jT, \frac{n}{NT}]$ denote the ST and its conjugate expressions of time series $x[kT]$ , respectively.

The frequency band is of the range 1–35 Hz and the interval is set at 1 Hz. Then, the signal is calculated based on the ST PSD feature according to (10). Each channel extracts 35 features, and an additional 2240 features were extracted through experiments; therefore, the training and test data after feature extraction are Tr2240×278 and Te2240×100, respectively. ST provides better energy concentration in the time-frequency domain. Therefore, the use of ST can obtain a spectral density function with higher time-frequency resolution. As ST has an independent Gaussian window, it can obtain higher frequency resolution at low frequencies and better time-domain positioning at high frequencies.

### 2.3 Classification

BLDA is used to perform regression in a Bayesian framework. To some extent, BLDA is relevant to the so-called evidence framework[23-24].

In Bayesian regression, a linear relation is assumed between targets t and feature vectors x, and it can be given by

$$t = w^T \mathrm{X} + n \tag{11}$$

where n is the additive white Gaussian noise.

The likelihood function for the weights $w$ under the aforementioned assumption is

$$p(D|\beta, w) = (\frac{\beta}{2\pi})^{\frac{N}{2}} \exp(-\frac{\beta}{2} \left\| \mathrm{X}^T w - t \right\|^2) \tag{12}$$

where $t$ is the target, X is the feature vector, $\beta$ denotes the inverse of the noise variance, $N$ is the number of training samples, and $D$ is denoted as $\{\mathrm{X}, t\}$.

The classifiers were trained on the data obtained from the extracted features and selected imagined movements. The training and validation datasets contained 278 and 100 trials, respectively. The imaginary samples of the left-hand small finger and tongue movement were labeled as +1 and −1, respectively. However, the output of the classifier is usually not exactly +1 or −1 but fluctuates around these values. Therefore, the threshold value was set and the threshold was determined according to the output variables.

## 3. Experiment

In this study, the EEG/ERP system from Neuroscan Company (United Sates) was used to collect EEG signals. The whole system includes 64-channels electrode cap, head box, Syn Amps 2 amplifier and Scan 4.5 analysis Software with frequency-domain adaptive notch filter. The sampling rate of 64-channel electrode cap can reach 20000 Hz per channel, and the electrodes of the electrode cap are arranged in accordance to the international standard 10-20 system [1], which can be seen in Figure 2.

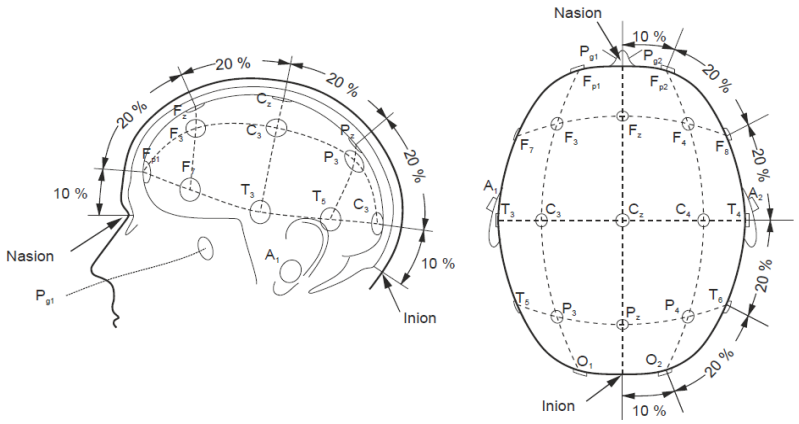



Fig.2 The diagram of the electrode position in the international standard 10-20 system

The collection of EEG signal is the process of obtaining the motor imagery EEG signal from the scalp surface of the subject through the application of a 64-channel dedicated electrode cap, which is firstly amplified by the Syn Amps2 amplifier and then transmitted to the signal processing computer. According to electrode positions in the 10-20 international standard system (see Figure 6), the EEG signals from 10 channels, including FC3, FC4, C5, C3, C1, C2, C4, C6, CP3 and CP4, are collected in a single pole way, with which the difference in the positions of the individual electrodes can be solved. The loss of useful EEG information caused by the small number of electrodes can be prevented. Meanwhile, the difficulty of signal processing will not increase due to the increase of the excessive number of electrodes.

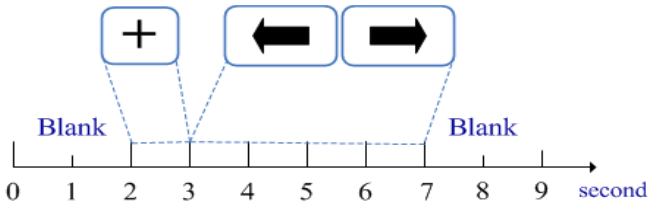



Fig.3 Experimental paradigm process

According to the experimental recording process of dataset I in BCI Competition Database III, a simple experimental paradigm based on E-Basic language and programmed on E-Prime software was designed. As shown in Figure 3, the execution time of a complete experiment costs 7 seconds in total. Firstly, within 2 seconds before the starting of the experiment, the screen of the stimulating computer was blank, allowing the subject to enter the preparation state. Then, the screen was presented for 1 second with a "+" cursor, reminding the subject to start the imaging task. At the 3rd second, a random arrow appeared on the screen with left (imagining the left thumb to bend) or to the right (imagining the right thumb). The duration of this imaging task was 4 seconds. When the arrow disappeared, the imaging stopped. It was stipulated that every 400 single experiments was set as a group, two types of sports imaging tasks were presented randomly for 200 times. The subjects had 2 minutes to have a rest between every 40 experiments. In total, the EEG signals from four subjects were recorded and saved.

During the collecting process of EEG signals, in order to display the different performance modes of ERD/ERS for the motor imagery of the left and right hands, the EEG signals of the C3 and C4 channels of each subject was dealt with the superimposed and averaged calculation. ERD/ERS can

be defined as the decreasing or increasing percentage of EEG power value at the target time point. In the comparison with EEG average power value in the reference power period[9], the formula is shown as follows:

$$ERD / ERS = \frac{EEG_A - EEG_R}{EEG_R} \tag{13}$$

Where, $EEG_A$ is EEG power value at the target time point; $EEG_R$ is EEG average power value during the reference power period. The ERD/ERS mode of the left and right hands is presented in Figure 4. The EEG signals of the C3 and C4 channels of the subject are illustrated by the relative percentage after superimposed and averaged calculation, while the ERD/ERS mode usually appears at 3.5s to the 6.5s of the single test. Therefore, only this piece of data of the collected signal was only processed (pretreatment, feature extraction, classification identification).

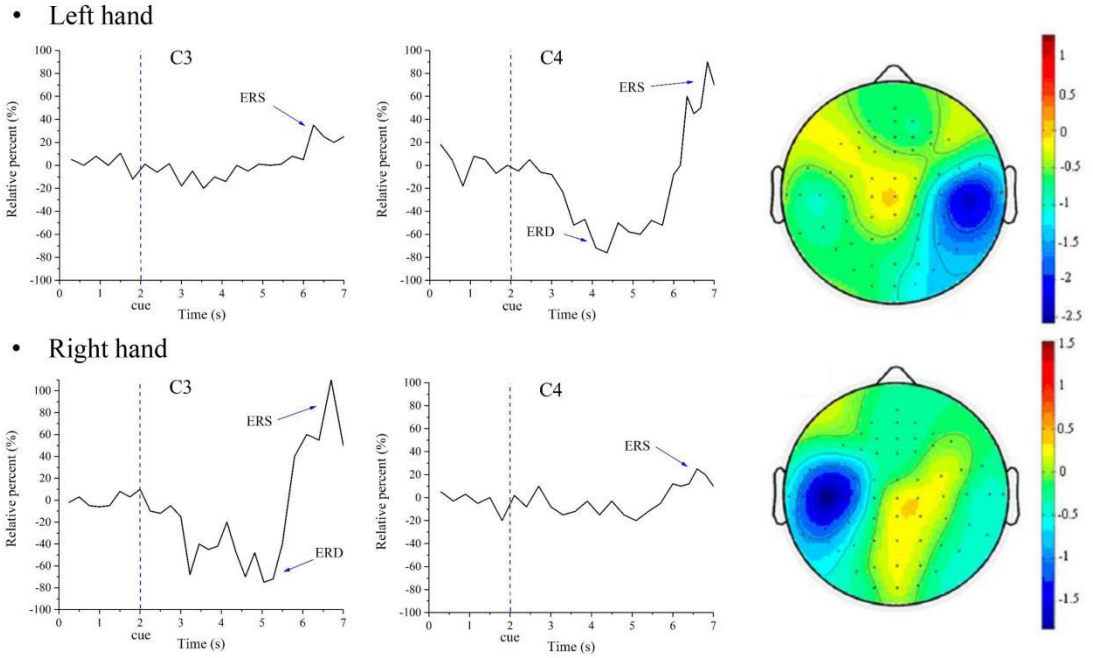



Fig. 4 is the timing sequence diagram of ERD/ERS and brain topographic map. The length of extraction is 0~7s during the single test. After the timing sequence diagram of ERD/ERD, the brain map of the single test from 3.5s to 6.5s can be provided. The blue color represents the ERD mode, while the yellow color stands for the ERS mode.

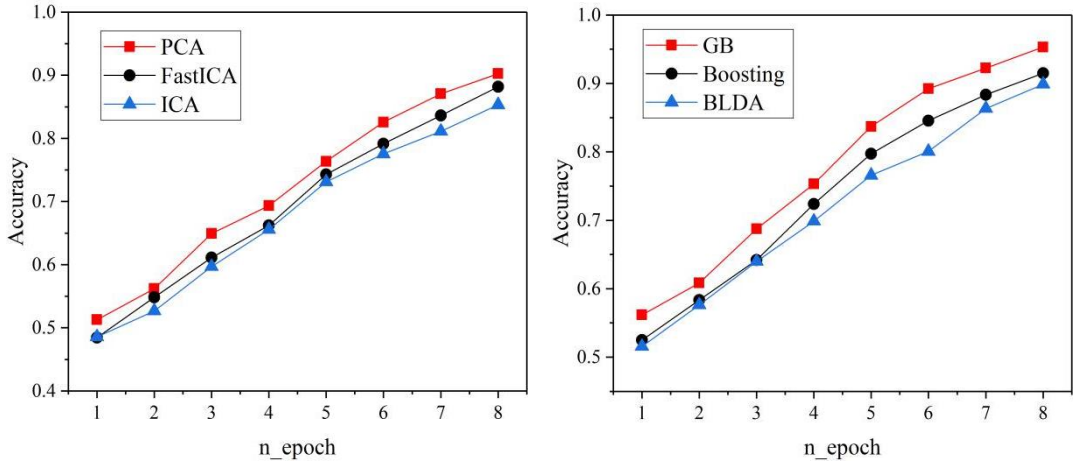
## 4. Results

In the experiment, the databasewas used to jointly train ST The preprocessing and classification identification algorithm with better combination effect with ST was selected. Firstly, by means of Python2.7, the better dimensionality reduction algorithm in PCA/FastICA was selected. Then, by using Python3.6, the better classification algorithm in GB algorithm/BLDA was selected. The training results are shown in Table 1. The preprocessing and classification identification algorithms are compared as presented in Figure 5:

Table 1 Accuracy of joint training with ST

| Epoch | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| Extraction | PCA | 51.28 | 56.20 | 64.92 | 69.33 | 76.35 | 82.56 | 87.06 | 88.25 |
| | FastICA | 49.43 | 54.84 | 61.12 | 66.22 | 74.29 | 79.13 | 83.64 | 88.19 |
| | ICA | 48.56 | 52.66 | 59.70 | 65.57 | 73.10 | 77.53 | 81.12 | 85.32 |
| | | | | | | | | | |
| Classification | GB | 56.17 | 60.85 | 68.75 | 75.32 | 83.71 | 89.23 | 92.25 | 88.32 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Boosting | 52.50 | 58.33 | 64.21 | 72.37 | 79.72 | 84.54 | 88.36 | 88.50 |
| | BLDA | 51.56 | 57.64 | 63.98 | 69.87 | 76.57 | 80.06 | 86.34 | 90.88 |




The joint training results of FastICA (b) The joint training results of BLDA

The training results (Figure 5 (a)) show that the accuracy of ST can only reach 85% after the dimensionality reduction by FastICA. That is to say, the dimensionality reduction performance is more efficient when FastICA combined with ST. The training results (Figure 5 (b)) show that the combined classification accuracy of ST and BLDA can reach 90%. This means that the classifying performs better when BLDA combined with ST.

## 5. Conclusion

In our research, the identification framework of the MI–BCI system is proposed. Especially, the sensitivity of ST in feature extraction is high, The proposed method can be used with any classifier to identify the feature subset that will lead to the best performance. The results show that our proposed algorithm can reduce the number of required features and improve the performance of the overall classifier.In the signal preprocessing part, LMS adaptive filtering algorithm was used for filtering the EEG signal. The comparison of dimensionality reduction methods FastIca, shows that the training results was better by FastICA dimensionality reduction, with the joint training accuracy rate with ST up to 88%. The classification identification was conducted by BLDA. The training results indicated that after ST, when the BLDA algorithm was used for the further classifying, the classification performance was better with the 90% accuracy of joint training. The results from the experimental process can be obtained in 2 minutes, in which the EEG data can be conducted in the program with final classification result of 90% or above. This also shows that after the classifying by employed ST model for two times and combined with the BLDA, the better results can be obtained, which demonstrates the feasibility and innovation of this method.

## References

[1] Nicolas-Alonso L F, Gomez-Gil J. Brain computer interfaces, a review[J]. Sensors, 2012, 12(2): 1211.

[2] Wolpaw J R, Birbaumer N, Mcfarland D J, et al. Brain-computer interfaces for communication and control[J]. Clinical Neurophysiology, 2013,113(6): 767-791.

[3] B.Z. Allison, E.W. Wolpaw, and J.R. Wolpaw, “Brain-Computer Interface Systems: Progress and Prospects,” Expert Rev. Medical Devices, vol. 4, no. 4, pp. 463-474, 2007.

[4] Vidal J J. Toward Direct Brain-Computer Communication[J]. Annual Review of Biophysics and Bioengineering, 1973, 2(1): 157-180.

[5] Wolpaw J R, Birbaumer N, Heetderks W J, et al. Brain-computer interface technology: a review of the first international meeting[J]. Rehabilitation Engineering IEEE Transactions on, 2000, 8(2): 164-173.
[6] Pfurtscheller G and Neuper C 2001 Motor imagery and direct brain–computer communication Proc. IEEE 89 1123–34.
[7] Jeannerod M 1995 Mental imagery in the motor context Neuropsychologia 33 1419.
[8] Babiloni F 2011 Brain computer interfaces for communication and control J. Craniofac. Surg. 22 1981–2.
[9] Kaper M, Meinicke P, Grossekathoefer U, Lingner T and Ritter H 2004 BCI competition 2003–data set IIb: support vector machines for the P300 speller paradigm IEEE Trans. Biomed. Eng. 51 1073–6.
[10] Xu F, Zhou W, Zhen Y and Yuan Q 2014 Classification of ECoG with modified S-transform for brain–computer interface J. Comput. Inform. Syst. 10 8029–41.
[11] Wang T, Deng J and He B 2004 Classifying EEG-based motor imagery tasks by means of time-frequency synthesized spatial patterns Clin. Neurophysiol. 115 2744.
[12] Benjamin B, Steven L, Matthias T, Stefan H and KlausRobert M 2011 Single-trial analysis and classification of ERP components—A tutorial NeuroImage 56 814–25.
[13] Kachenoura A, Albera L, Senhadji L and Comon P 2008 Ica: a potential tool for BCI systems IEEE Signal Process. Mag.25 57–68.
[14] Blankertz B, Tomioka R, Lemm S, Kawanabe M and Muller K R 2007 Optimizing spatial filters for robust EEG single-trial analysis IEEE Signal Process. Mag. 25 41–56.
[15] Hoffmann U, Vesin J M, Ebrahimi T and Diserens K 2008 An efficient P300-based brain–computer interface for disabled subjects J. Neurosci. Methods 167 115–25.
[16] Corralejo R, Nicolás-Alonso L F, lvarez D, et al. A P300-based brain–computer interface aimed at operating electronic devices at home for severely disabled people[J]. Medical & Biological Engineering & Computing, 2014, 52(10): 861-872.
[17] N.V. Thakor, Y.S. Zhu. Applications of adaptive filtering to ECG analysis: noise cancellation and arrhythmia detection[J]. IEEE Transactions on Biomedical Engineering, vol. 38, no. 8, pp. 785-794, 1991.
[18] N. SruthiSudha, D.V. RamaKoti Reddy. Detection and Removal of artefacts from EEG signal using sign based LMS Adaptive Filters[J]. International Journal of Scientific & Engineering Research, vol. 8, no. 2, pp. 950-954, Feb 2017.
[19] X. Lei, P. Yang, and D.Z. Yao. An empirical Bayesian framework for brain computer interfaces[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol 17, pp.521-529, 2009.
[20] MacKayD. Bayesian interpolationl[J]. Neural Computation,1992,4(3):415-447.
[21] Stockwell R G, Mansinha L and Lowe R P 1996 Localization of the complex spectrum: the S transform IEEE Trans. Signal Process. 44 998–1001.
[22] Stockwell R G 2007 A basis for efficient representation of the S-transform Digit. Signal Process. 17 371–93.
[23] Mackay D J C 1992 Bayesian interpolation Neural Comput.4 415–47.
[24] Bishop C M 2009 Pattern Recognition and Machine Learning(New York: Springer) p 049901.