Frontiers of Information Technology & Electronic Engineering
www.zju.edu.cn/jzus; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Fast removal of ocular artifacts from electroencephalogram signals using spatial constraint independent component analysis based recursive least squares in brain-computer interface*

Bang-hua YANG[†‡], Liang-fei HE[†], Lin LIN[†], Qian WANG

(*Shanghai Key Laboratory of Power Station Automation Technology, Department of Automation, School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200072, China*)

[†]E-mail: yangbanghua@126.com; heliangfei126@126.com; fxlinlin@shu.edu.cn

Received Aug. 18, 2014; Revision accepted Jan. 13, 2015; Crosschecked May 7, 2015

Abstract: Ocular artifacts cause the main interfering signals within electroencephalogram (EEG) signal measurements. An adaptive filter based on reference signals from an electrooculogram (EOG) can reduce ocular interference, but collecting EOG signals during a long-term EEG recording is inconvenient and uncomfortable for the subject. To remove ocular artifacts from EEG in brain-computer interfaces (BCIs), a method named spatial constraint independent component analysis based recursive least squares (SCICA-RLS) is proposed. The method consists of two stages. In the first stage, independent component analysis (ICA) is used to decompose multiple EEG channels into an equal number of independent components (ICs). Ocular ICs are identified by an automatic artifact detection method based on kurtosis. Then empirical mode decomposition (EMD) is employed to remove any cerebral activity from the identified ocular ICs to obtain exact artifact ICs. In the second stage, first, SCICA applies exact artifact ICs obtained in the first stage as a constraint to extract artifact ICs from the given EEG signal. These extracted ICs are called spatial constraint ICs (SC-ICs). Then the RLS based adaptive filter uses SC-ICs as reference signals to reduce interference, which avoids the need for parallel EOG recordings. In addition, the proposed method has the ability of fast computation as it is not necessary for SCICA to identify all ICs like ICA. Based on the EEG data recorded from seven subjects, the new approach can lead to average classification accuracies of 3.3% and 12.6% higher than those of the standard ICA and raw EEG, respectively. In addition, the proposed method has 83.5% and 83.8% reduction in time-consumption compared with the standard ICA and ICA-RLS, respectively, which demonstrates a better and faster OA reduction.

Key words: Ocular artifacts, Electroencephalogram (EEG), Electrooculogram (EOG), Brain-computer interface (BCI), Spatial constraint independent component analysis based recursive least squares (SCICA-RLS)

doi:10.1631/FITEE.1400299 **Document code:** A **CLC number:** TP399; R318.18

1 Introduction

A brain-computer interface (BCI) is an alternative communication and control channel that does not depend on the brain's normal output pathway of peripheral nerves and muscles (Yang *et al*., 2007). A BCI system is intended to help severely disabled people to communicate with computers or to control electronic devices through their thoughts, and it has the prospect of wide applications. An electroencephalogram (EEG) is the neurophysiologic measurement of the electrical activity of the brain recorded through electrodes placed on the scalp. It is the most widely used brain imaging modality for BCIs (Mammone *et al*., 2007). However, it is known that EEG has low spatial resolution and high noise level, which makes it challenging to extract useful information from EEG signals for BCI applications. Ocular artifacts (OAs)

‡ Corresponding author

* Project supported by the National Natural Science Foundation of China (Nos. 31100709 and 60975079) and the Shanghai Pujiang Program, China (No. 14PJ1431300)

ORCID: Bang-hua YANG, http://orcid.org/0000-0003-4261-9875

including eye movements and eye blinks, which often produce large and distracting artifacts in EEG, are the most important forms of interference in EEG signals (Jung *et al*., 2000). Those OAs can be detected by electrodes pasted around eyes (Wang, 2010).

At present, OA removal methods for EEG include mainly things such as linear regression analysis, adaptive filtering, and blind source separation (Delorme *et al*., 2007). (1) Linear regression analysis: it is the most common EEG artifact removal method, which has been generally used to remove OAs. This method makes use of transferred coefficients which are derived from the time or frequency domain to show the relationship between the OAs and the EEG. Clear EEG can be obtained by subtracting electrooculogram (EOG) with corrupted EEG segments. (2) Adaptive filtering: it is a more effective method compared with the time-domain regression method (Chan *et al.*, 2010). He *et al.* (2004) used adaptive filtering to eliminate OAs in EEG. The method employs the raw EEG signal as input, and the horizontal and vertical EOGs as reference signals. Clear EEG can be obtained by subtracting the filtered outputs. During the experiment, to guarantee that the reference signals have the most proper characteristics relative to the OAs, the filter coefficients need to be adjusted by the mean squared value of the obtained clear EEG. (3) Blind source separation method: it has been shown by many researchers that the blind source separation method can be used to efficiently separate the distinct artifactual components from EEG data. Blind source separation assumes that the EEG signals are a linear mixture of neural and ocular activities. Independent component analysis (ICA) is one of popular blind source separation methods (Hazra *et al*., 2010). In the last decade, ICA has had a crucial role in neuroscience research with a lot of attention paid to it for artifact rejection (Klados *et al*., 2010). ICA can decompose multiple EEG channels into an equal number of independent components (ICs). ICs that contain OAs can be identified through visual inspection. After discarding OA-related ICs, the uncorrupted EEG signals can be reconstructed by the remaining ICs. Akhtar *et al.* (2012) summarized that ICA yields superior performance compared with the regression method.

Linear regression analysis and adaptive filtering are effective methods that are used to reduce ocular activity in EEG signals. However, they share a limitation: they need an additional EOG recording. Unfortunately, collecting EOG signals during a long-term EEG recording is inconvenient and uncomfortable for the subject. ICA can separate OA-related ICs efficiently, but some useful EEG information content can be found in the component accounting for OAs. Thus, discarding these OAs would cause information loss. At the same time, ICs corresponding to artifacts are manually selected using visual inspection in the ICA method, which is subject to human bias.

This paper describes a novel method of removing OAs from EEGs using spatial constraint independent component analysis based recursive least squares (SCICA-RLS). Our objective is to remove artifacts quickly from EEG data and have the following properties at the same time: (1) fast and automatic identification of OA components, and (2) reducing the loss of desired information from original artifact-free segments. The SCICA-RLS method can be divided into two stages (Fig. 1). In the first stage, ICA decomposes a randomly single trial selected from raw EEG data into a set of ICs. Kurtosis is employed to quickly and automatically identify the OA-ICs. Then we apply empirical mode decomposition (EMD) to OA-ICs to remove leaked cerebral activity and obtain exact artifact signals. The second stage is based mainly on SCICA and the adaptive filter. SCICA, which incorporates references or constraint topographies (the artifact-only ICs obtained from the first stage) in the ICA algorithms, allows us to extract only desired ICs effectively without the necessity of extracting all ICs. These desired ICs,

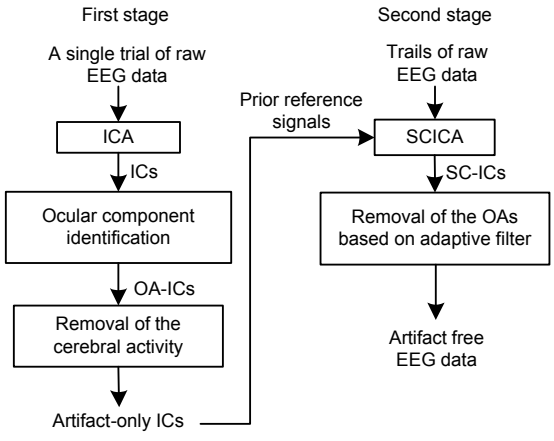

Fig. 1 Simple flowchart of the proposed method

called SC-ICs, serving as references of the adaptive filter method, might help to overcome the limitations of an RLS-based adaptive filter of parallel EOG recordings. Finally, the adaptive filter, whose parameters are optimized by the RLS method, is used to remove OAs from the EEG signal. The second stage is fast in computation as not all of the ICs are extracted and processed. Moreover, it does not eliminate the OA-ICs directly compared with the standard ICA, which will retain the desired information to a maximum. The first stage, running to obtain exact artifact ICs, can be seen as the training stage. The second stage is used mainly for testing and online application.

2 Method

2.1 First stage of the SCICA-RLS method

In this stage, our objective is to realize fast and automatic identification of artifactual components and to obtain exact artifact components in which cerebral information has been removed from the ocular component. To achieve the goal, ICA is employed to decompose multiple EEG channels into an equal number of ICs. The OAs significantly contribute to some ICs but not others. Then OA-ICs are identified by an automatic artifact detection method based on kurtosis. Finally, EMD is employed to remove any cerebral activity from the identified ocular sources, and thus exact artifact components are obtained.

2.1.1 Separation of ICs based on ICA

ICA is a blind source separation technique for transforming an observed multidimensional random vector into components that are statistically as independent of each other as possible. Classical ICA can work well without any information about the mixing matrix (Winkler *et al*., 2014). More precisely, let $y_l(n)$ denote the EEG signal from channel l (l=1, 2, …, L), where L is the total number of channels and n is the sample number. Let $\boldsymbol{Y}(n)=[y_1(n), y_2(n), \ldots, y_L(n)]^{\mathrm{T}}$ be the input signal. ICA separates $\boldsymbol{Y}(n)$ into ICs which are denoted by $\boldsymbol{S}(n)=[s_1(n), s_2(n), \ldots, s_m(n)]^{\mathrm{T}}$ and therefore

$$\boldsymbol{Y}(n)=\boldsymbol{A}\boldsymbol{S}(n), \tag{1}$$

where $\boldsymbol{A}$ is the mixing matrix. Then there exists a de-mixing matrix $\boldsymbol{W}$ such that

$$\boldsymbol{S}(n)=\boldsymbol{W}\boldsymbol{Y}(n). \tag{2}$$

The objective of ICA is to find $\boldsymbol{W}=[\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_n]$. ICA is actually an optimization problem, depending on an objective function and an optimization algorithm (Huang and Wang, 2013). In this study, according to the maximum negentropy criterion, the objective function is defined as

$$C(s)=\sum_{i=1}^{m} J(s_i), \tag{3}$$

where $s_i=\boldsymbol{w}_i^{\mathrm{T}}\boldsymbol{Y}(n)$, $J(s_i)\approx\rho(E\{G_i(s_i)\}-E\{G_i(v)\})^2$ with ρ a positive constant, $G_i(\cdot)$ a non-quadratic function, $E\{\cdot\}$ a mean function, and v a Gaussian variable having zero mean and unit variance. We next identify the OA-ICs from $\boldsymbol{S}(n)$ using an automatic identification method based on kurtosis.

2.1.2 Identification of OA-ICs based on kurtosis

Kurtosis, as a measure of 'peakedness', is used to detect 'abnormal' peaked distributions. So, we can identify the OA-ICs based on kurtosis as they have higher kurtosis values than others. Each IC's kurtosis value k_l can be computed by

$$k_l=m_4-3m_2^2, \tag{4}$$

where m_j (j=2, 4) is the jth-order central moment with the following form:

$$m_j=E\{[s_l(n)-\overline{s}_l(n)]^j\}, \tag{5}$$

where $\overline{s}_l(n)$ is the mathematical expectation.

Fig. 2 shows 15 ICs ($\boldsymbol{S}(n)=[s_1(n), s_2(n), \ldots, s_{15}(n)]^{\mathrm{T}}$) separated by ICA and their kurtosis values. From Fig. 2 we can see that the kurtosis value k_3 of s_3 is 47.1576, which is much larger than those of the remaining ICs. So, s_3 is identified as the OA-IC, denoted by $S_{\mathrm{OA}}(n)$. The computation of kurtosis is fast and efficient. It took only 0.012335 s to calculate the kurtosis value of an epoch of EEG signals with 4000 samples. In Barbati *et al*. (2004) and Güçlü *et al*.

(2011), fraction dimension and sample entropy were employed to identify OA-ICs, respectively. However, it took 0.124346 s and 33.554516 s to calculate the fraction dimension and sample entropy, respectively, of the same epoch of EEG signals with 4000 samples. The time for calculating the kurtosis is much less than that for calculating fraction dimension and sample entropy. Thus, kurtosis is more suitable for use in de-noising. After each IC's kurtosis value is calculated, $\boldsymbol{K}=[k_1, k_2, \ldots, k_m]$ is obtained. Then OA-ICs, which are denoted as $\boldsymbol{S}_{\mathrm{OA}}=[s_{\mathrm{OA}_1}, s_{\mathrm{OA}_2}, \ldots, s_{\mathrm{OA}_p}]$, where p is the number of the OA-ICs, are automatically identified by a threshold K_v. However, there is some cerebral information leakage into the OA-ICs. So, EMD will be applied to remove any cerebral activity further from the OA-ICs.

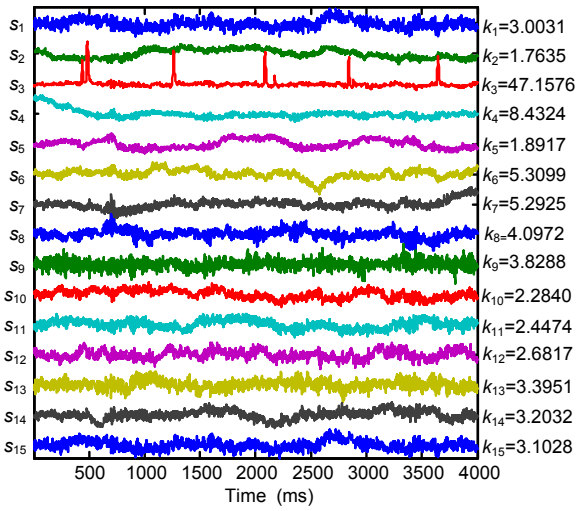

Fig. 2 Fifteen ICs separated by ICA and their kurtosis values

2.1.3 Removal of the cerebral activity from OA-ICs based on EMD

ICA can transform an observed multidimensional random vector into components that are statistically as independent of each other as possible (Winkler *et al*., 2014). It has been widely accepted that the artifactual signals are independent of the ongoing cerebral activity (Klados *et al*., 2010). So, ICA is usually able to concentrate on the artifactual information in a single component or a few ICs. However, at most times those ICs also carry non-artifactual information. Using the components as references of the adaptive filter may cause cerebral activity loss. Here, we apply EMD to remove any cerebral activity that is leaked to these artifact ICs, and obtain only artifact ICs.

EMD is aimed at decomposing the original signal into a series of intrinsic mode functions (IMFs) with time scale feature. The iterative extraction of these components is based on the local representation of the signal. With EMD, the signal is decomposed into a set of IMFs, which is a kind of complete, adaptive, and almost orthogonal representation of the analyzed signals (Arasteh *et al*., 2010).

More precisely, if $S_{\mathrm{OA}}(n)$ is the input signal, then EMD decomposes $S_{\mathrm{OA}}(n)$ into intrinsic mode functions denoted by $\{d_i(n)\}_{i=1}^{N}$ such that

$$S_{\mathrm{OA}}(n)=\sum_{i=1}^{N} d_i(n)+r(n), \tag{6}$$

where $r(n)$ denotes the residual monotonic function. To obtain a meaningful estimate of the instantaneous frequency, IMF should be designed as a close symmetry around the local mean and its number of extrema and zero-crossing must be equal or different at most by one.

EMD can be employed to separate neural activity from the strong artifacts. The basic principle is that the decomposition of an OA signal based on EMD has the property of 'concentrating on' the information in a few IMFs. The OA component is usually concentrated on in a few low-frequency IMFs. Consider an example from a random trial shown in Fig. 3a. It shows the results of EMD for a segment of an OA of real EEG data, which comprises artifacts and neural activity. EMD decomposes an OA-IC into 10 IMFs (IMF1–IMF10). Fig. 3b shows the synthesized artifact signal by selecting different numbers of IMFs, in which the first few IMFs with high frequency are discarded to obtain the desired signals. Fig. 3c is the remaining information of Fig. 3b. Thus, the final objective is to model the artifact as exactly as possible. IMF (1, 2) and IMF (1–3) in Fig. 3c indicate that IMF1, IMF2, and IMF3 might be neural activity, and the OA contributes to the remaining IMFs (4–10). Artifacts can be seen in the last three sub-graphs of Fig. 3c (IMF (1–4), IMF (1–5), and IMF (1–6)), which are marked by the circle. Thus, IMF1–IMF3

are regarded as neural activity and IMF4–IMF10 are chosen to synthesize an exact artifact signal.

In addition to visual selection, kurtosis can be used to select artifact IMFs as mentioned above. The kurtosis values of IMF (1, 2), IMF (1–3), IMF (1–4), and IMF (1–5) are calculated, which are 3.0026, 3.1602, 3.3805, and 3.5240, respectively. From the third combination of IMF (1–4), the kurtosis value is obviously larger than those of IMF (1, 2) and IMF (1–3). So, a threshold value can be set to select IMFs.

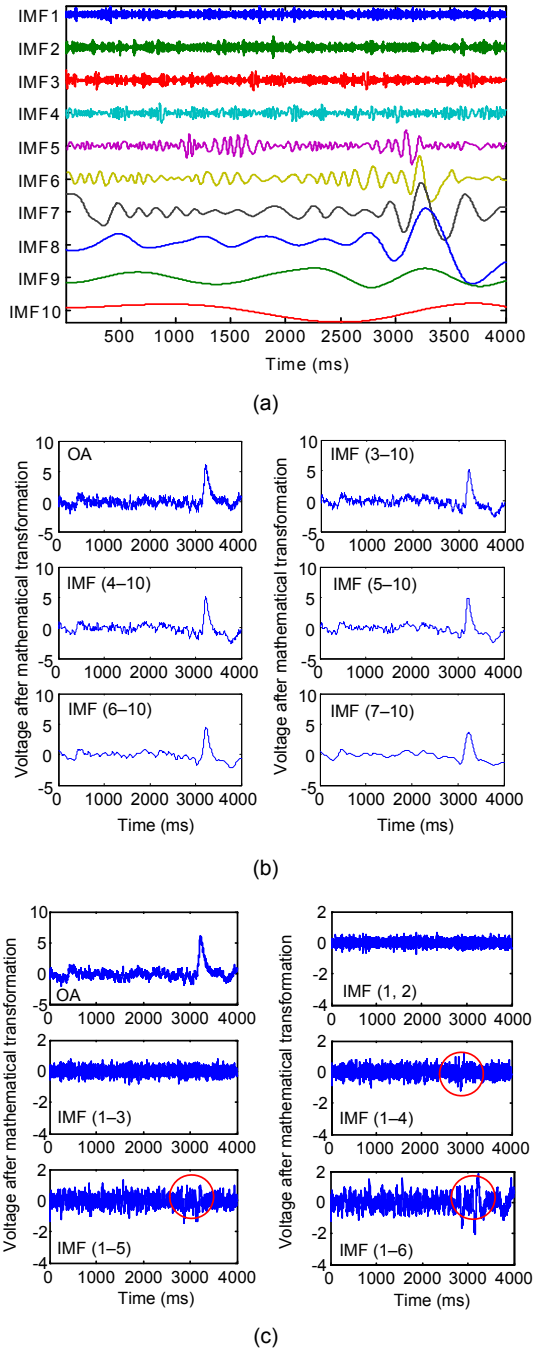

Fig. 3 EMD results for removing any cerebral activity leaked to OA-ICs

(a) Ten IMFs decomposed by EMD from an OA of real EEG data; (b) Synthesized artifact signal by selecting different numbers of IMFs; (c) Remaining information of Fig. 3b

The remaining trials for the same subject are also processed, and similar results are obtained. That is to say, the first N IMFs are regarded as neural activity and the other IMFs are chosen to synthesize an exact artifact signal. We use N=3 in this study, and it is different from subject to subject. The neural activity information leaked in subsequent IMFs can be reduced as much as possible by optimizing N according to different subjects.

Wavelet de-noising (WD) is used to remove any cerebral activity leaked to the OA related ICs using wavelet thresholding (Nguyen *et al*., 2012). The most competitive advantage of EMD over WD is that the base function and the number of decomposition levels in EMD do not need to be given and the algorithm itself can obtain these parameters according to the characteristics of signals. So, the EMD approach is quite simple as compared with the WD method. So far, EMD has been widely used in the processing of EEG signals (Li and Wang, 2013).

2.1.4 Implementation of the first stage

After the first stage described above, we obtain exact artifact ICs, which will be used as a spatial constraint for the second stage. Fig. 4 provides the block diagram of the first stage. The procedure for separation and identification of the OA is summarized as follows:

1. ICA is employed to separate raw EEG $\boldsymbol{Y}(t)=[y_1(t), y_2(t), \ldots, y_L(t)]^{\mathrm{T}}$ into a set of statistics ICs, denoted by $\boldsymbol{S}(t)=[s_1(t), s_2(t), \ldots, s_q(t)]^{\mathrm{T}}$, where q is the number of ICs.

2. Calculate the kurtosis value of each IC from $\boldsymbol{S}(t)$, which is denoted by $\boldsymbol{K}=[k_1, k_2, \ldots, k_m]$. Then sort them in descending order. According to the threshold value K_v, the first l ICs are identified as OA related ICs, denoted as $\boldsymbol{S}_{\mathrm{OA}}(t)=[s_{\mathrm{OA}_1}(t), s_{\mathrm{OA}_2}(t), \ldots, s_{\mathrm{OA}_n}(t)]$.

3. EMD is applied to decompose the identified OA ($\boldsymbol{S}_{\mathrm{OA}}(t)$) into a set of IMFs (**IMF**=[imf_1, imf_2, …, imf_x]). A few of IMFs with high frequency are discarded and the remaining IMFs are composed to

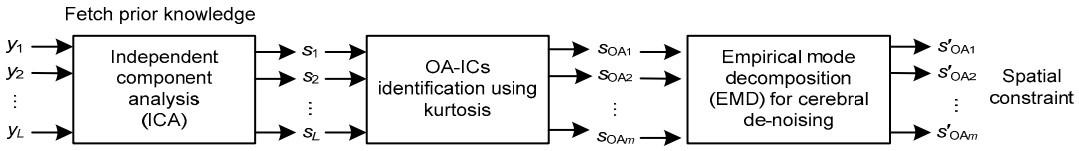

Fig. 4 Block diagram of the first stage

obtain an exact artifact signal, which is denoted by $\boldsymbol{S}'_{\mathrm{OA}}(t)=[s'_{\mathrm{OA}_1}(t), s'_{\mathrm{OA}_2}(t), \ldots, s'_{\mathrm{OA}_l}(t)]$.

Then $\boldsymbol{S}'_{\mathrm{OA}}(t)$ serves as a spatial constraint for the second stage.

2.2 Second stage of the SCICA-RLS method

In this stage, the objective is to carry out an efficient procedure for artifact removal from EEG data. An RLS-based adaptive filter with reference signals from an EOG can remove the ocular interference efficiently. However, there is a problem: collecting EOG signals during a long-term EEG recording is inconvenient and uncomfortable for the subject. SCICA can extract the 'only desired' ICs, that is, SC-ICs. The SC-ICs that are extracted by SCICA can be approximately regarded as a reference for the RLS adaptive filter, which avoids the need for parallel EOG recordings. At the same time, it is not necessary for SCICA to run a complete ICA to extract all sources, which will reduce processing time. Thus, the SCICA method enjoys an important practical advantage. The procedure of the second stage is summarized as follows.

2.2.1 Fast extraction of SC-ICs based on SCICA

SCICA is a blind source separation approach which incorporates reference or constraint topographies in the ICA algorithms. The detailed information about the SCICA method can be found in Hesse and James (2005). Compared with the conventional ICA methods, the SCICA method defines a spatial constraint which is denoted by $\boldsymbol{A}_{\mathrm{c}}$ on the mixing matrix $\boldsymbol{A}$ to represent specific prior knowledge. Thus, the original mixing matrix $\boldsymbol{A}$ can be divided into two parts. One part comprises the spatial constraint columns and the other part comprises the unconstrained columns, denoted by $\boldsymbol{A}_{\mathrm{u}}$:

$$\boldsymbol{A}=[\hat{\boldsymbol{A}}_{\mathrm{c}}, \boldsymbol{A}_{\mathrm{u}}], \tag{7}$$

where $\hat{\boldsymbol{A}}_{\mathrm{c}} \approx \boldsymbol{A}_{\mathrm{c}}$, the columns of which are subjecting to the constraint. For experimental purpose, the specific prior knowledge which is used as the spatial constraint can be extracted from a previous data segment by conventional ICA methods.

Here we consider the FastICA algorithm for SCICA. Using the Newtonian iteration method and choosing an initial weight vector $\boldsymbol{W}$, the basic form of the FastICA iteration algorithm is as follows:

$$\boldsymbol{W}=E\{\boldsymbol{Y}G(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{Y})\}-E\{G'(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{Y})\}\boldsymbol{W}, \tag{8}$$

$$\boldsymbol{W}=\boldsymbol{W}/\|\boldsymbol{W}\|, \tag{9}$$

where G is a nonlinear function and G' is the derivative of G. The main purpose of the SCICA algorithm is to extract the ICs which are most relative to the spatial constraint. The specific algorithm mechanisms can be found in Hesse and James (2005). That is to say, we extract only the ICs which we are interested in. Thus, the outputs of the SCICA algorithm are SC-ICs, which are denoted by $\boldsymbol{S}_{\text{sc-ic}}(n)=[S_{\text{sc-ic1}}(n), S_{\text{sc-ic2}}(n), \ldots, S_{\text{sc-ic}m}(n)]^{\mathrm{T}}$. $\boldsymbol{S}_{\text{sc-ic}}(n)$ serves as a reference for the adaptive filter.

2.2.2 Removal of ocular artifacts based on RLS using SC-ICs

Each of the EEG signals $y_l(n)$ is assumed to contain both neural activity $x(n)$ and ocular activity $v(n)$. These aforementioned SC-ICs ($\boldsymbol{S}_{\text{sc-ic}}(n)$) serve as reference signals $r_1(n)$, $r_2(n)$, …, $r_p(n)$, where p is the number of reference signals for an adaptive filter, written as

$$v_j(n)=\sum_{k=1}^{H} h_j(k)r_j(n-k+1), \tag{10}$$

where $h_j(k)$ is the kth coefficient of a finite impulse response filter with length H. The corrupted EEG ($e(n)$) is corrected by subtracting the sum of the filtered reference signals from the primary signal $y_l(n)$:

$$e(n)=y_l(n)-\sum_{j=1}^{p} v_j(n). \tag{11}$$

To obtain the best cancellation of OAs, the filter coefficients $h_j(k)$ are adjusted using an RLS method by minimizing the sum $e(n)$ of the weighted squared errors:

$$\varepsilon_n = e^2(n) + \lambda e^2(n-1) + \cdots + \lambda^{n-L} e^2(L), \qquad (12)$$

where λ ($0<\lambda<1$) is a 'forgetting factor' that gradually reduces the effects of previous errors.

2.2.3 Implementation of the second stage

Fig. 5 provides the block diagram of the second stage. The procedure for removing the OA is summarized as follows:

1. SCICA with a spatial constraint is employed to separate raw EEG $\boldsymbol{Y}(t)=[y_1(t), y_2(t), \ldots, y_L(t)]^{\mathrm{T}}$ to obtain SC-ICs, i.e., $\boldsymbol{S}_{\text{sc-ic}}(n)=[S_{\text{sc-ic1}}(n), S_{\text{sc-ic2}}(n), \ldots, S_{\text{sc-ic}m}(n)]^{\mathrm{T}}$.

2. $\boldsymbol{S}_{\text{sc-ic}}(n)=[S_{\text{sc-ic1}}(n), S_{\text{sc-ic2}}(n), \ldots, S_{\text{sc-ic}m}(n)]^{\mathrm{T}}$ serves as reference signals $r_1(n), r_2(n), \ldots, r_p(n)$ for an adaptive filter.

3. Adaptive filters based on RLS are used to estimate $\tilde{\boldsymbol{v}}(n)=[v_1, v_2, \ldots, v_m]$ using Eq. (10).

4. Correct EEG $e(n)$ is obtained by subtracting $\tilde{\boldsymbol{v}}(n)=[v_1, v_2, \ldots, v_m]$ from the primary input $\boldsymbol{Y}(t)=[y_1(t), y_2(t), \ldots, y_L(t)]^{\mathrm{T}}$.

After the second stage is carried out, OAs have been significantly removed.

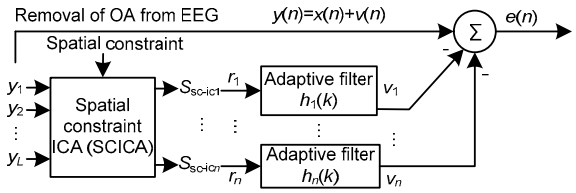

Fig. 5 Block diagram of the second stage

3 Data description

Dataset used in the study is the publicly available dataset of BCI Competition IV. The dataset was recorded from seven healthy subjects. Each subject selected two classes of motor imagery from the three classes (left hand, right hand, and foot imagery). During the first two runs, arrows pointing left, right, or down were presented on a computer screen as a cue to instruct a subject to perform a corresponding imagination task. The cues and imagination task lasted 4 s. The cues used for different tasks were interleaved with 2 s of blank screen and 2 s with a fixation cross shown in the center of the screen. The fixation cross was superimposed on the cues. These datasets were provided with complete marker information. EEG signals were recorded from 59 channels, which were densely positioned over sensorimotor areas. The signals were band-pass filtered between 0.05 and 200 Hz and then sampled at 100 Hz. More details were described in Blankertz *et al*. (2007).

4 Results and discussion of SCICA-RLS

In this study we chose 15 channels (F3, F1, FZ, F2, F4, C5, C3, C1, CZ, C2, C4, C6, P1, PZ, P2) associated with the motor imagery. EEG in this study is from motor imagery-based BCI. This kind of EEG analysis focuses mainly on the cerebral cortex which is mostly related to motion tasks. At the same time, considering the practical requirement on the number of channels for EEG analysis and the application of BCI, 15 EEG channels which are mostly relevant to motor imaginary were chosen. Specific electrode distribution is shown in Fig. 6. To evaluate the OA removal technique, there are two main criteria: (1) how well the artifacts were removed and (2) how well the neural data were preserved. The classification accuracy can reflect the amount of residual information and the distortion degree of the EEG to some degree. Here, the proposed method was evaluated

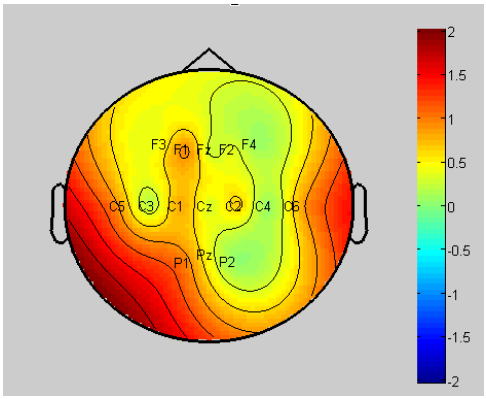

Fig. 6 Distribution diagram of the 15 motor imagery electrodes

from observation of the signal after removal of OAs, computation of integrity of the signal, classification result, and time consumption.

To check the performance of the proposed algorithm, the following two methods were also used in the experiments:

1. ICA: ICA is used to decompose multiple EEG channels into an equal number of ICs. The OAs significantly contribute to some ICs but not others. ICs that include OAs can be identified using the method of kurtosis. Then the artifact ICs are set to zero. The remaining ICs are reconstructed to obtain the correct EEG.

2. ICA-RLS: On the basis of ICA, after identifying the artifact ICs with kurtosis, an adaptive filter based on RLS uses reference signals from the identified artifact ICs to reduce OAs, which avoids setting the artifact ICs to zero.

Classical ICA algorithms include mainly the JADE (joint approximative diagonalization of the eigen matrix) algorithm, extended maximum entropy algorithm, and FastICA algorithm. FastICA is one of the popular and referenced ICA techniques, based on its own unique fast fixed-point iterative algorithm (Qiao *et al*., 2009). It has the advantage of fast convergence and good separation, which makes it widely used in artifact separation. In this study, the ICA method, ICA-RLS method, and SCICA-RLS method all use FastICA.

4.1 Observation of the signal after removal of OAs

A trial from the raw dataset was randomly selected, and then the trial was used as the input for the SCICA-RLS method. Fig. 7 shows the processing results of the trial. Fig. 7a shows the raw EEG data. It is obvious that the raw data contain many OAs (this can be seen in the red box). Fig. 7b shows 15 ICs separated by ICA. The kurtosis values of these ICs are calculated and shown in Table 1. It can be seen that the kurtosis value of IC2, which is identified as

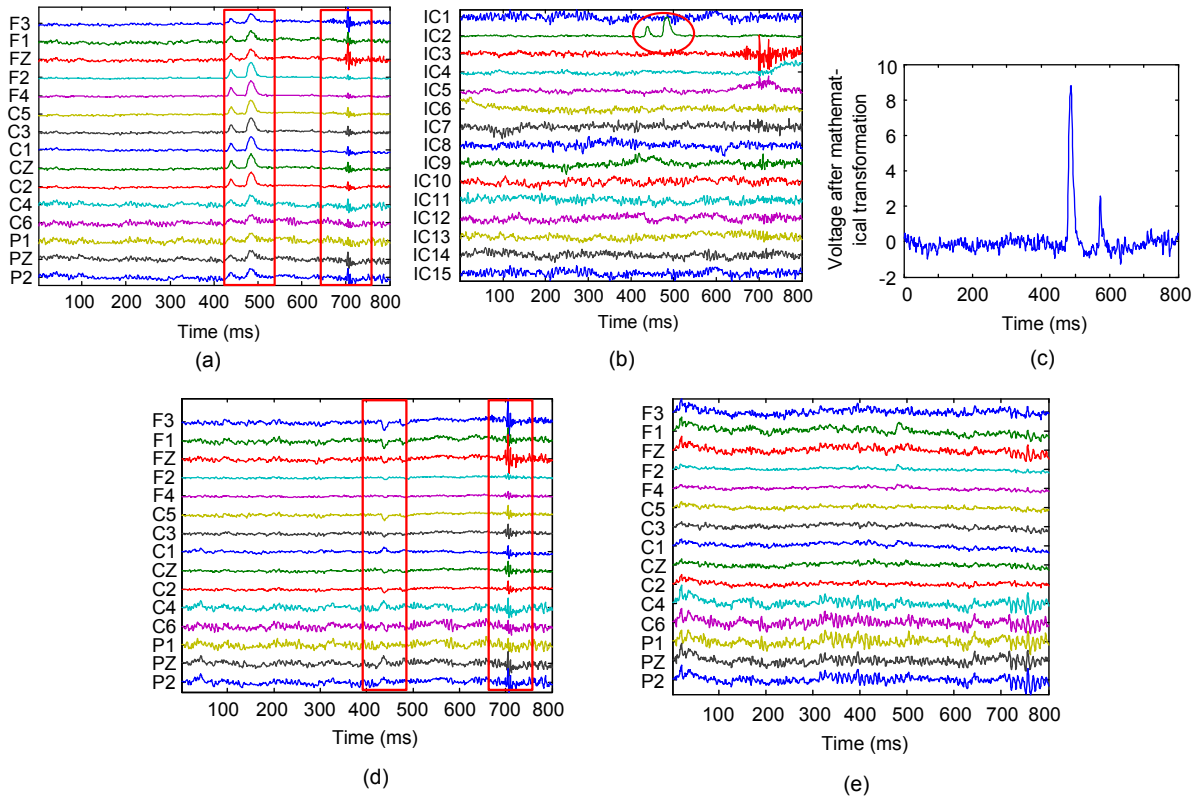

Fig. 7 Processing results by the proposed SCICA-RLS method and ICA method

(a) Raw EEG data; (b) Fifteen ICs separated by ICA; (c) SC-IC separated by SCICA; (d) Results of the standard ICA method (not all the artifacts were removed); (e) Results of the SCICA-RLS method (almost all the artifacts were removed). References to color refer to the online version of this figure

Table 1 Kurtosis value of each IC

IC	*K*-value	IC	*K*-value	IC	*K*-value
IC1	11.3459	IC6	6.4653	IC11	2.5893
IC2	**33.2809**	IC7	4.8158	IC12	2.6165
IC3	26.8648	IC8	6.1406	IC13	2.6125
IC4	8.4005	IC9	4.7230	IC14	3.1821
IC5	7.1427	IC10	2.4104	IC15	2.9986

K-value: kurtosis value. The bold value represents the largest kurtosis value

OA-IC, is much larger than those of other ICs. Fig. 7c shows the SC-IC separated by SCICA. Fig. 7d shows the processing result of the standard ICA method. The result was obtained by reconstructing non-artifactual EEG components after zeroing IC2 which was identified as the OA-IC. Though almost all the OA components were removed from the raw EEG signals, there were still some burrs and OA components (this can be clearly seen in the red box). Fig. 7e shows the processing results of the SCICA-RLS method, i.e., the method using SC-IC as reference to filter the raw EEG. It is obvious that almost all the OAs and burrs have been significantly removed. We also processed the remaining trials from the dataset. Similar results were obtained. Comparison of Figs. 7d and 7e indicates that the proposed method has a better performance on artifact removal than ICA.

4.2 Computation of integrity of the signal after removing OA

The root mean square error (RMSE) and Pearson's correlation coefficient were used to verify the effectiveness of the proposed approach. The two metrics can measure how well the neural data were preserved in artifact-free EEG after the OA removal procedure has been carried out. The RMSE is defined as

$$\mathrm{RMSE}=\sqrt{\sum_{i=1}^{N}(\mathrm{srcEEG}_i-\mathrm{crtEEG}_i)^2}, \tag{13}$$

and Pearson's correlation coefficient is written as

$$\mathrm{Corrcoef}=\frac{\sum_{i=1}^{N}(\mathrm{srcEEG}_i-\overline{\mathrm{srcEEG}})(\mathrm{crtEEG}_i-\overline{\mathrm{crtEEG}})}{\sqrt{\sum_{i=1}^{N}(\mathrm{srcEEG}_i-\overline{\mathrm{srcEEG}})^2(\mathrm{crtEEG}_i-\overline{\mathrm{crtEEG}})^2}}, \tag{14}$$

where srcEEG and crtEEG denote the input and reconstructed signals, respectively, and $\overline{\mathrm{srcEEG}}$ and $\overline{\mathrm{crtEEG}}$ represent the corresponding mathematical expectations, respectively. The srcEEG was unknown because we collected only the scalp EEG as observed signals. Clear EEG that does not contain OAs was used as srcEEG. It is an epoch of clear EEG which came from the same subject as crtEEG. The RMSE and Pearson's correlation coefficient mean value across the channels were computed after normalizing the EEG signals from the dataset. The smaller the RMSE was, the more the preserved information of neural data was. The larger the value of Corrcoef was, the more the preserved information of neural data was. The mean and variance of RMSE and Corrcoef for the ICA method, the ICA-RLS method, and the SCICA-RLS method are shown in Table 2.

Table 2 Comparison of RMSE and Corrcoef based on 200 trials from the dataset

Method	RMSE ($\times10^{-4}$)	Corrcoef
ICA	7.3±0.26	0.94±0.028
ICA-RLS	7.1±0.24	0.97±0.021
SCICA-RLS	**7.0±0.18**	**0.97±0.018**

RMSE: root mean square error; Corrcoef: Pearson's correlation coefficient. Values are in the form of 'mean±variance'. Bold values represent the best results

Table 2 clearly shows that the mean values of RMSE for ICA, ICA-RLS, and SCICA-RLS are small. The RMSE for SCICA-RLS is the smallest, which means that the EEG signal in which OAs have been removed by the SCICA-RLS method preserved more neural information. Moreover, the mean Pearson correlation coefficient for SCICA-RLS is larger, while the variance for SCICA-RLS is smaller. It means that the proposed method preserves more neural data than the other two methods.

4.3 Classification results after removing OAs

The EEG polluted by OAs can reduce the classification accuracy. After the OAs were removed, the classification accuracy would be improved. Classification accuracies of seven healthy subjects from the dataset were calculated, in which each subject includes 200 trials.

The common spatial pattern (CSP) method is a widely used spatial filtering technique that can extract discriminative features for EEG-based BCI classification tasks (Li and Koike, 2012). The support vector machine (SVM) is one of the best-known techniques for its good theoretical foundations and high classification accuracy. It has been used extensively in biomedical signal analysis, speech recognition, and face recognition (Xu *et al*., 2010). After obtaining a non-artifact EEG dataset by ICA, ICA-RLS, and SCICA-RLS methods, respectively, CSP was used to extract 6D features of the non-artifact EEG dataset and the SVM classifier was used to classify the features extracted. The classification accuracies in the case of 3-fold cross-validation for the three methods are shown in Fig. 8.

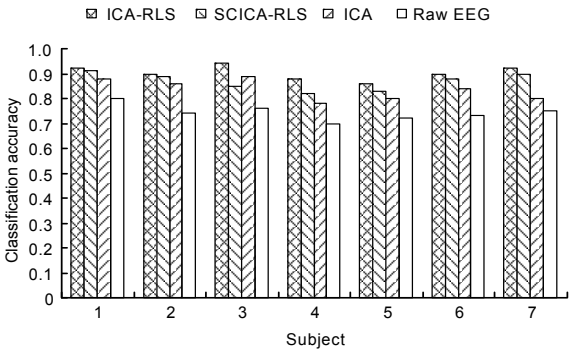

Fig. 8 Comparison of classification accuracies for different methods and subjects

It can be seen from Fig. 8 that the three methods (ICA, ICA-RLS, and SCICA-RLS) can improve the classification accuracy to different degrees compared with the raw EEG. Compared with the standard ICA, ICA-RLS and the proposed SCICA-RLS method achieved average classification accuracies of 6.7% and 3.3% higher, respectively. The average classification accuracy of ICA-RLS was 3.4% higher than that of the proposed method. ICA-RLS separates and identifies OA-IC in every trial, which improves the classification accuracy but is more time-consuming.

A high classification accuracy can reflect the amount of residual information and the distortion degree of the EEG to some degree. In other words, the higher the classification accuracy, the more cerebral information remains and the lower the distortion degree of the EEG. From the comparison of classification accuracy, we can see that ICA-RLS and the proposed method achieve better performance than ICA in removing OAs.

4.4 Time consumption

Time consumption is a very important issue in an online BCI system. It is also a metric to measure the effectiveness of the method. To quantify the computational complexity of the methods discussed in this study, the processing time under different methods is given in Fig. 9. The data were obtained using tic and toc commands of MATLAB (release R2013a) running on Windows 7 (Intel Core i3-4130 CPU 3.4 GHz processor, 3 GB RAM). Fig. 9 clearly shows that the time consumptions for 200 trials (each trial contains 800 samples) for the three methods were 62.044, 63.388, and 10.233 s, respectively. All these three methods used kurtosis values to identify OAs. As expected, the proposed approach was quite a lot faster compared with the other approaches discussed in this study. It achieved a reduction of 83.5% and 83.9% in time consumption compared with ICA and ICA-RLS, respectively.

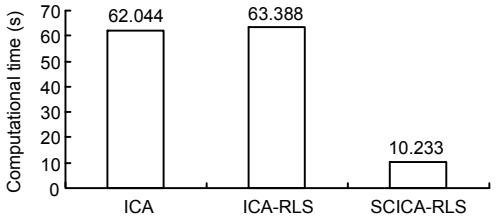

Fig. 9 Comparison of computational time (200 trials) for the three methods

5 Conclusions

In this paper, a new method, SCICA-RLS, has been presented for automatically removing OAs in EEG data. This method has combined SCICA with RLS, which eliminates the difficulty of collecting EOG data during a long-term EEG recording. The proposed method can remove artifacts clearly and achieve high classification accuracy. In addition, it might greatly reduce processing time compared with the ICA and ICA-RLS methods. From the experimental results, OAs among the raw EEG have been removed effectively by the proposed method.

SCICA-RLS can lead to an average classification accuracy of 3.3% higher than that of the standard ICA. Thus, the proposed method has better performance than the standard ICA. In addition, the proposed method is much faster than the ICA and ICA-RLS methods, which can provide a foundation for online and practical applications. The practical online performance of the proposed SCICA-RLS method will be investigated further.

References

Akhtar, M.T., Mitsuhashi, W., James, C.J., 2012. Employing spatially constrained ICA and wavelet denoising, for automatic removal of artifacts from multichannel EEG data. *Signal Process.*, **92**(2):401-416. [doi:10.1016/j.sigpro.2011.08.005]

Arasteh, A., Janghorbani, A., Moradi, M.H., 2010. A survey on EMD sensitivity to SNR for EEG feature extraction in BCI application. Proc. 5th Cairo Int. Biomedical Engineering Conf., p.175-179. [doi:10.1109/CIBEC.2010.5716061]

Barbati, G., Porcaro, C., Zappasodi, P., *et al*., 2004. Optimization of an independent component analysis approach for artifact identification and removal in magnetoencephalographic signals. *Clin. Neurophysiol.*, **115**(5):1220-1232. [doi:10.1016/j.clinph.2003.12.015]

Blankertz, B., Dornhege, G., Krauledat, M., *et al*., 2007. The non-invasive Berlin brain-computer interface: fast acquisition of effective performance in untrained subjects. *NeuroImage*, **37**(2):539-550.

Chan, H.L., Tsai, Y.T., Meng, L.F., *et al*., 2010. The removal of ocular artifacts from EEG signals using adaptive filters based on ocular source components. *Ann. Biomed. Eng.*, **38**(11):3489-3499. [doi:10.1007/s10439-010-0087-2]

Delorme, A., Sejnowski, T., Makeig, S., 2007. Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *NeuroImage*, **34**(4):1443-1449. [doi:10.1016/j.neuroimage.2006.11.004]

Güçlü, U., Güçlütürk, Y., Loo, C.K., 2011. Evaluation of fractal dimension estimation methods for feature extraction in motor imagery based brain computer interface. *Procedia Comput. Sci.*, **3**:589-594. [doi:10.1016/j.procs.2010.12.098]

Hazra, B., Roffel, A., Narasimhan, S., *et al.*, 2010. Modified cross-correlation method for the blind identification of structures. *J. Eng. Mech.*, **136**(7):889-897. [doi:10.1061/(ASCE)EM.1943-7889.0000133]

He, P., Wilson, G., Russell, C., 2004. Removal of ocular artifacts from electroencephalogram by adaptive filtering. *Med. Biol. Eng. Comput*, **42**(3):407-412. [doi:10.1007/BF02344717]

Hesse, C.W., James, C.J., 2005. The FastICA algorithm with spatial constraints. *IEEE Signal Process. Lett.*, **12**(11):792-795. [doi:10.1109/LSP.2005.856867]

Huang, L., Wang, H., 2013. Reducing the computation time for BCI using improved ICA algorithms. Proc. 10th Int. Symp. on Neural Networks, p.299-304. [doi:10.1007/978-3-642-39068-5_37]

Jung, T.P., Makeig, S., Humphries, C., *et al.*, 2000. Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, **37**(2):163-178. [doi:10.1007/978-3-540-74819-9_84]

Klados, M.A., Bratsas, C., Frantzidis, C., *et al*., 2010. A kurtosis-based automatic system using naïve Bayesian classifier to identity ICA components contaminated by EOG or ECG artifacts. Proc. XII Mediterranean Conf. on Medical and Biological Engineering and Computing, p.49-52. [doi:10.1007/978-3-642-13039-7_13]

Li, X., Wang, W.B., 2013. Studying on denoising of chaotic signal using ICA and EMD. Proc. Int. Symp. on Geo-Informatics in Resource Management and Sustainable Ecosystem, p.564-572. [doi:10.1007/978-3-642-45025-9_55]

Li, Y., Koike, Y.H., 2012. A real-time BCI with a small number of channels based on CSP. *Neur. Comput. Appl.*, **20**(8):1187-1192. [doi:10.1007/s00521-010-0481-6]

Mammone, N., Inuso, G., La Foresta, F., *et al*., 2007. Multi-resolution ICA for artifact identification from electroencephalographic recordings. Proc. 11th Int. Conf. on Neural Networks, p.680-687. [doi:10.1007/978-3-540-74819-9_84]

Nguyen, H.A.T., Musson, J., Li, F., *et al*., 2012. EOG artifact removal using a wavelet neural network. *Neurocomputing*, **97**:374-389. [doi:10.1016/j.neucom.2012.04.016]

Qiao, X., Li, D., Dong, Y., 2009. P300 feature extraction based on parametric model and FastICA algorithm. Proc. 5th Int. Conf. on Natural Computation, p.585-589. [doi:10.1109/ICNC.2009.160]

Wang, B., 2010. Comprehensive Study on Removal of Artifacts from EEG Data. MS Thesis, Zhejiang University, Hangzhou, China (in Chinese).

Winkler, I., Brandl, S., Horn, F., *et al*., 2014. Robust artifactual independent component classification for BCI practitioners. *J. Neur. Eng.*, **11**(3):035013.1-035013.10. [doi:10.1088/1741-2560/11/3/035013]

Xu, H., Song, W., Hu, Z.P., *et al*., 2010. A speedup SVM decision method for online EEG processing in motor imagery BCI. Proc. 10th Int. Conf. on Intelligent Systems Design and Applications, p.149-154. [doi:10.1109/ISDA.2010.5687274]

Yang, B.H., Yan, G.Z., Wu, T., *et al*., 2007. Subject-based feature extraction using fuzzy wavelet packet in brain-computer interfaces. *Signal Process.*, **87**(7):1569-1574. [doi:10.1016/j.sigpro.2006.12.018]