Biomedical Signal Processing and Control 18 (2015) 127–137

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Multi-ganglion ANN based feature learning with application to P300-BCI signal classification

Wei Gao [a,b], Jin-an Guan [a,b,*], Junfeng Gao [a,b], Dao Zhou [a,b]

[a] College of Biomedical Engineering, South-Central University for Nationalities, Wuhan, People's Republic of China
[b] Key Laboratory of Cognitive Science, State Ethnic Affairs Commission, Wuhan, People's Republic of China

ARTICLE INFO

Article history:
Received 10 October 2014
Received in revised form 1 December 2014
Accepted 19 December 2014
Available online 24 January 2015

Keywords:
Imitating-Reading BCI
Multi-channel signal feature extraction
Multi-ganglion artificial neural network

ABSTRACT

The feature extraction of event-related potentials (ERPs) is a significant prerequisite for many types of P300-BCIs. In this paper, we proposed a multi-ganglion artificial neural network based feature learning (ANNFL) method to extract a deep feature structure of single-trial multi-channel ERP signals and improve classification accuracy. Five subjects took part in the Imitating-Reading ERP experiments. We recorded the target electroencephalography (EEG) samples (elicited by target stimuli) and non-target samples (elicited by non-target stimuli) for each subjects. Then we applied ANNFL method to extract the feature vectors and classified them by using support vector machine (SVM). The ANNFL method outperforms the principal component analysis (PCA) method and conventional three-layer auto-encoder, and then leads to higher classification accuracies of five subjects' BCI signals than using the single-channel temporal features. ANNFL is an unsupervised feature learning method, which can automatically learn feature vector from EEG data and provide more effective feature representation than PCA method and single-channel temporal feature extraction method.

1. Introduction

Brain-computer interfaces (BCIs) can utilize the electrical signals generated by the central nervous system (CNS), and then translate these signals into users' commands, which can control some external systems, such as personal computer, electrically propelled wheelchairs and prosthesis [1–5]. BCI system can provide a new bridge between the brain and the world that bypass the human body. Therefore, this device can be used as an alternative communication method to improve the life-quality of people with amyotrophic lateral sclerosis (ALS), stroke and some other neuromuscular diseases and may change the way we are communicating, entertaining and living remarkably in the future [6,7].

The BCI systems can be divided into two main categories, noninvasive [3,4] and invasive [5], according to the signal acquisition methods. Because of the utilization conveniences, the most of recent work focuses on the field of noninvasive BCIs. The three main noninvasive BCI methods base on event-related potentials (ERP) [8,9], the steady-state visual-evoked potentials (SSVEPs) [10,11] and the slow cortical potentials (SCPs) [12].

The Imitating-Reading BCI system is based on visual P300 signal, which is the most widely used ERP signal. P300 signals can be elicited by auditory, visual or somatosensory stimuli. The more unpredictable the odd stimulus is, the higher the amplitude of P300 signal is [13]. Farwell and Donchin first described the P300-BCI paradigm in their work [14]. Then the other researchers developed many efficient P300-BCI systems, some of the studies had test the performances of P300-BCI systems for people with ALS and stroke [15,16],

The EEG signals used by BCIs as the carriers of communication are always non-stationary and contain lots of undesired signals. The key in improving the performance of a P300-BCI system is enhancing the real-time detection of the ERP signals elicited by target stimuli [17]. Then, for the whole system, building a good feature representation is an opportunity to achieve that goal.

For that reason, many of the actual efforts in developing brain-computer systems go into the design of an appropriate representation for the EEG signal in order to provide promising classification performance. Researchers have proposed the applications of different advanced spatial, temporal and spectrum feature extraction methods to BCI systems. These methods include principal component analysis (PCA) [18], common spatial pattern (CSP) [19–22], wavelet transform method [23], blind source separation [24] etc. They can reduce the dimensionality of BCI signals' feature space for the classification procedure.

* Corresponding author. Tel.: +86 15171457728.
E-mail addresses: gw_scuec@126.com, guanja@mail.scuec.edu.cn (J.-a. Guan).

http://dx.doi.org/10.1016/j.bspc.2014.12.007
1746-8094/

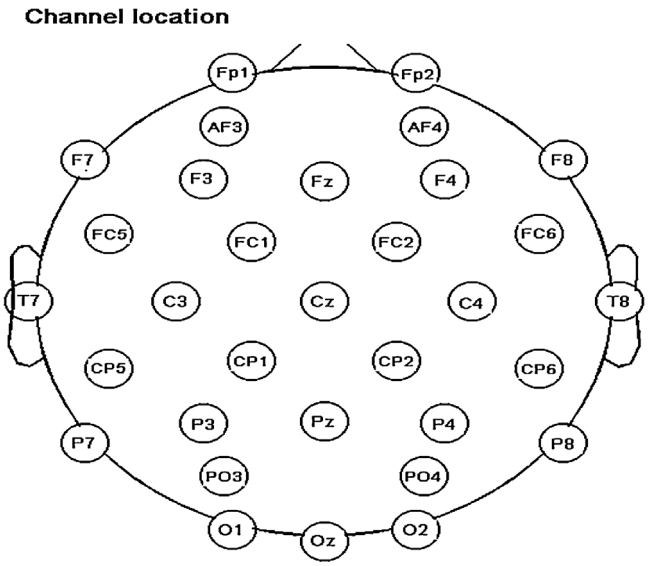

Fig. 1. Two-dimensional view of the electrodes setup configuration for the Imitating-reading BCI paradigm.

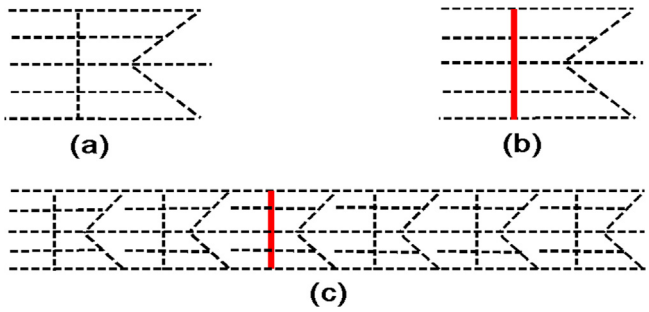

Fig. 2. The stimulus symbols. Fig. 2a shows the symbol of non-target stimulus, 2b shows the symbol of target stimulus, and 2c shows a stimulus-symbol line.

However, when training the feature extractors provided by these methods, like CSP methods, users need labeled EEG samples and to preserve the most distinctive information, the users should select part of these extracted features or assemble the features in

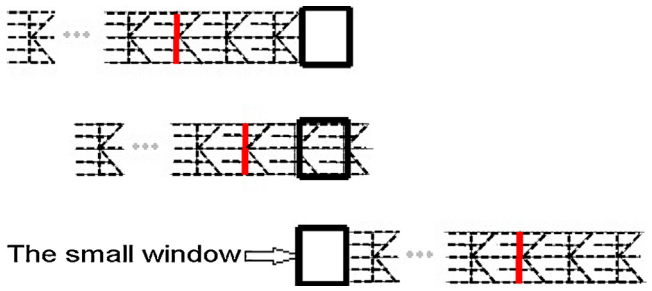

Fig. 3. The movement of a stimulus-symbol line. The stimuli move from left to right.

different spaces manually based on his or her experience. Though the PCA methods do not need label information and it can find the principal component of a multi-dimensional data matrix, it cannot suit for the situations when the signal has a low signal noise ratio (SNR) and may cost an information loss of EEG signals.

Meanwhile, Hinton and Salakhutdinov proposed an unsupervised feature learning method in their work [25]. They developed a multi-layer ANN composed by restricted Boltzmann machines to reduce the dimensionality of data and achieved a better reconstruction result than using PCA method. Researchers also proposed some other typical neural network for unsupervised learning, such as auto-encoder [26] and sparse coding model [27]. These methods have been applied to image processing and natural language processing [28,29]. Their works indicate that we can learn high-level representative features, components and patterns from data automatically by using unsupervised feature learning methods.

The unsupervised feature learning method always involves with a large scale ANN that has a considerable number of connection weights to be adjusted in the training stage [30]. The training time is the bottleneck of this kind of method. To introduce the unsupervised feature learning method to the field of P300-BCI research and improve the real-time performance of this method, in our study, we design the multi-ganglion artificial neural network feature learning (ANNFL) method. A multi-ganglion ANN has three-level network structure (neuron level, ganglion level and ganglion–network level), the connections occur within neurons in the same ganglion. Then we apply ANNFL method to the feature extracting procedure of Imitating-Reading BCI system, which uses

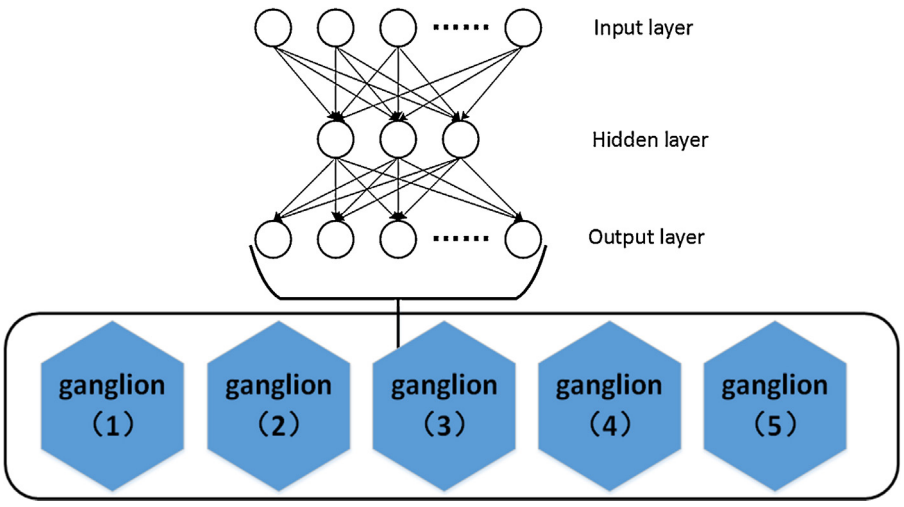

Fig. 4. The structures of a basic ganglion and the artificial neural network.

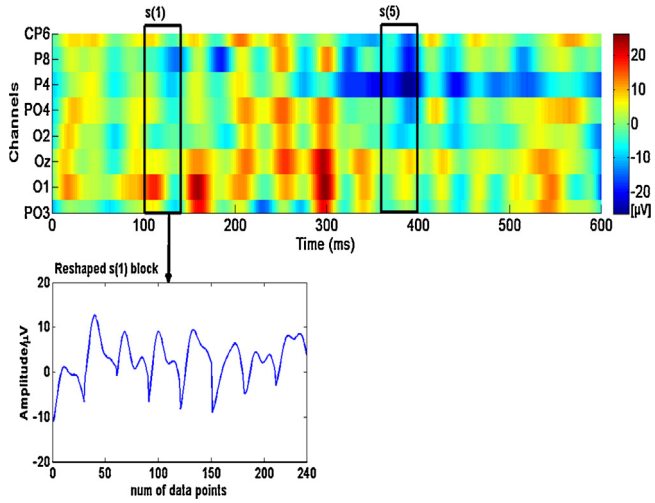

Fig. 5. The division of an EEG sample. The signals ranging from 100 ms to 400 ms after subjects receiving visual stimuli are selected as EEG samples. Channel labels denote the EEG channels used in the experiments. The magnitude map of an EEG sample is filled with colors according to its magnitudes. Maximum amplitude specifies the data value to the red end of colormap and the minimum to the blue end. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

event-related potential as the carrier of communication between brain and computer.

The remainder of this paper is organized as follows. In Section 2, we introduce the mental task of the P300-BCI study. We detail the ANNFL algorithm in Section 3 and give the experimental results in Section 4. Finally, we make a discussion in Section 5.

2. Imitating-Reading mental task and experimental settings

The paradigm of the Imitating-Reading BCI was designed and implemented in the College of Biomedical Engineering of South-Central University for Nationalities. The visual ERP evoked mode of it was designed to imitate the natural reading behavior, which successfully restrains the internal or external disturbing effects like exogenous brain noises and ocular artefact, and the communication speed of this kind paradigm can exceed 90 bit per minute in online experiments [31,32].

Five healthy subjects (Sub1, Sub2, Sub3, Sub4 and Sub5, three female, two male, aged 22–25) participated in this study, all participants reported normal or corrected-to-normal vision. Multi-channel EEG signals were recorded by BioSemi ActiveTwo system (BioSemi, Amsterdam, the Netherlands). The sampling rate is 2048 Hz and we rejected the trials in which the subjects reported they blinked eyes. Then we remove the baseline of each trial of EEG sample, filter the signals using a four-order IIR filter whose cut-off frequency is 30 Hz, and then decimate the sampling rate from 2048 Hz to 512 Hz.

According to the Imitating-Reading experimental settings [33], the subject who takes part in the experiment sits in an armchair at 40 cm distance in front of a computer screen and wears a cap with 32 scalp electrodes located at standard positions of the international 10–20 system (see Fig. 1, not including the reference electrodes at earlobe). A small window (size: 30 pix × 30 pix) is set on the center of a computer screen whose screen resolution is 1024 pix × 768 pix. In the experiments, the subject should watch on the small window. It displays two kinds of visual stimulus symbols (see Fig. 2a and b). The symbol shown in Fig. 2a is a non-target stimulus, and in Fig. 2b a target stimulus. The target stimulus symbol has a vertical red line across it. Then one target stimulus and nineteen non-target stimuli compose a stimulus-symbol line (see Fig. 2c).

In each trial of the experiments, the computer displays continuously the symbols of this line one by one in the small window, each stimulus stays in the window for 150 ms, it looks like the line is "walking" though the window (see Fig. 3). The interval between each of the two trials is 1.5 s, the participants can rest their eyes in the intervals.

In each trial of stimulation, we change the position of the scarce target stimulus symbol. Therefore, the subject cannot predict when the computer might display it. In this mental task, the target stimuli can elicit P300 event-related potentials, while the non-target stimuli cannot. Moreover, watching the stimulus-symbol line move across the window is like reading a book, while you do not need to move head or eyes and the book can move itself.

Consequentially, in the experiments, we can record two classes of EEG signals elicited by non-target and target stimuli. After signal preprocessing (remove the baseline, filter the signals and decimate the sampling rate from 2048 Hz to 512 Hz), in each trial, we use the EEG signals from 100 ms to 400 ms post a non-target stimulation and a target stimulation as non-target and target samples, so we can record one non-target sample and one target sample in each

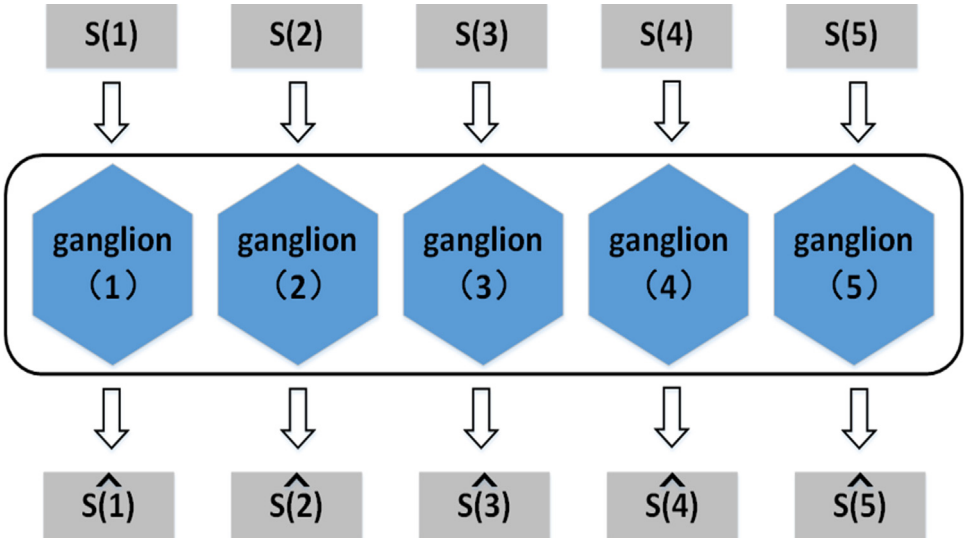

Fig. 6. The training stage of the multi-ganglion ANN. The input is the reshaped EEG samples of a training dataset and it reproduces its input at its output layer.

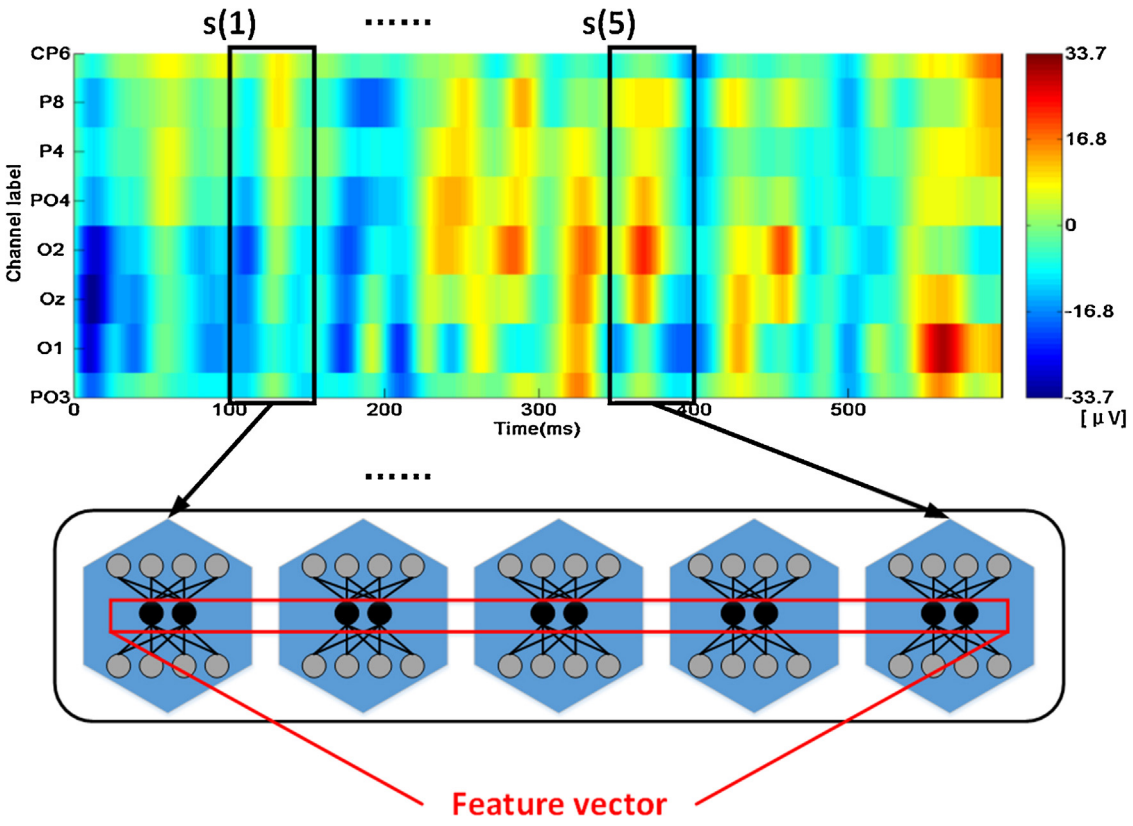

Fig. 7. The feature extracting procedure of an EEG sample. The feature vector consists of output values of all hidden units.

trial of Imitating-Reading ERP experiment. The performance of the Imitating-Reading BCI depends on how well the two classes of EEG signals can be classified. In this study, each dataset consists of 100 trials of target EEG signal samples and 100 trials of non-target EEG signal samples.

3. Methods

3.1. Multi-ganglion artificial neural network feature learning (ANNFL) method

For the Imitating-Reading BCI paradigm, we denote non-target and target samples as matrices E_U $\left(E_U \in R^{N\times t}\right)$ and E_T $\left(E_T \in R^{N\times t}\right)$. N is the total number of EEG channels used for feature extracting procedure (we use the EEG signals detected in parts of the channels) and t is the number of data points of each channel. In this study, channel PO3, O1, Oz, O2, PO4, P4, P8 and CP6 are chosen and the number of data points of each channel is 150, then the size of a sample matrix (E_T or E_U) is 8×150.

The structure of the ANN designed to extract feature vectors from E_U and E_T is shown in Fig. 4. The ANN consists of five ganglions and each ganglion is an auto-encoder (three-layer neural network, number of input units: 240, number of hidden units: 30, number of output units: 240).

Before putting into the ANN, each trial of the multi-channel EEG samples is cut into five same large parts (size: 8×30). Each part, denoted from s(1) to s(5), consists of the multi-channel signals in a certain time range (see Fig. 5), we then reshape them respectively into vectors (cascade each column of a part of the EEG sample, size of each vector: 1×240).

In the training stage, the ANN is ganglion-wise initialized and tuned. The training set contains equal number of non-target and target samples. Then the dataset is divided into five blocks, denoted from S(1) to S(5). Each block is composed of the parts of EEG samples with the same serial number as the block and used as the input and training goal of corresponding ganglion (e.g. training block S(1) of the first ganglion contains $s_1(1)$, $s_2(1)$, ... $s_L(1)$, L is the trial number). Then each ganglion is trained with Back Propagation (BP) algorithm [34,35]. After that we can put a BCI signal sample into the multi-ganglion ANN and extract the output values of the hidden-layer units of all ganglion as an EEG feature vector (size: 1×150). The procedures are shown in Figs. 6 and 7.

3.2. Magnitude squared of coherences and correlation distances between the EEG channels

The ANNFL method is a multi-channel feature learning technique. When a subject receive a target visual stimulus, the specific components of ERP can be detected by different electrodes. To find the target-specific components out, we estimate the Magnitude Squared of Coherence (MSC, see formula (1)) [36] of signals detected in different EEG channels.

$$\mathrm{MSC}_{XY}(f) = \left| \frac{\sum_i^N \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)^*}{\sqrt{\sum_i^N \left(X_i - \overline{X}\right)\left(X_i - \overline{X}\right)^* \sum_i^N \left(Y_i - \overline{Y}\right)\left(Y_i - \overline{Y}\right)^*}} \right| \quad (1)$$

In formula (1), X and Y are the Fourier transforms of the EEG samples detected in two different channels, i is the trial number of an EEG sample, N is the total number of the EEG samples detected in one channel, $\overline{X}$ and $\overline{Y}$ are the average of the Fourier-transform results of all trials of EEG signals for the EEG signals of two different channel respectively.

According to this BCI paradigm, target-specific components are the components have higher coherence-value differences between non-target and target samples. In the filtering stage (before using

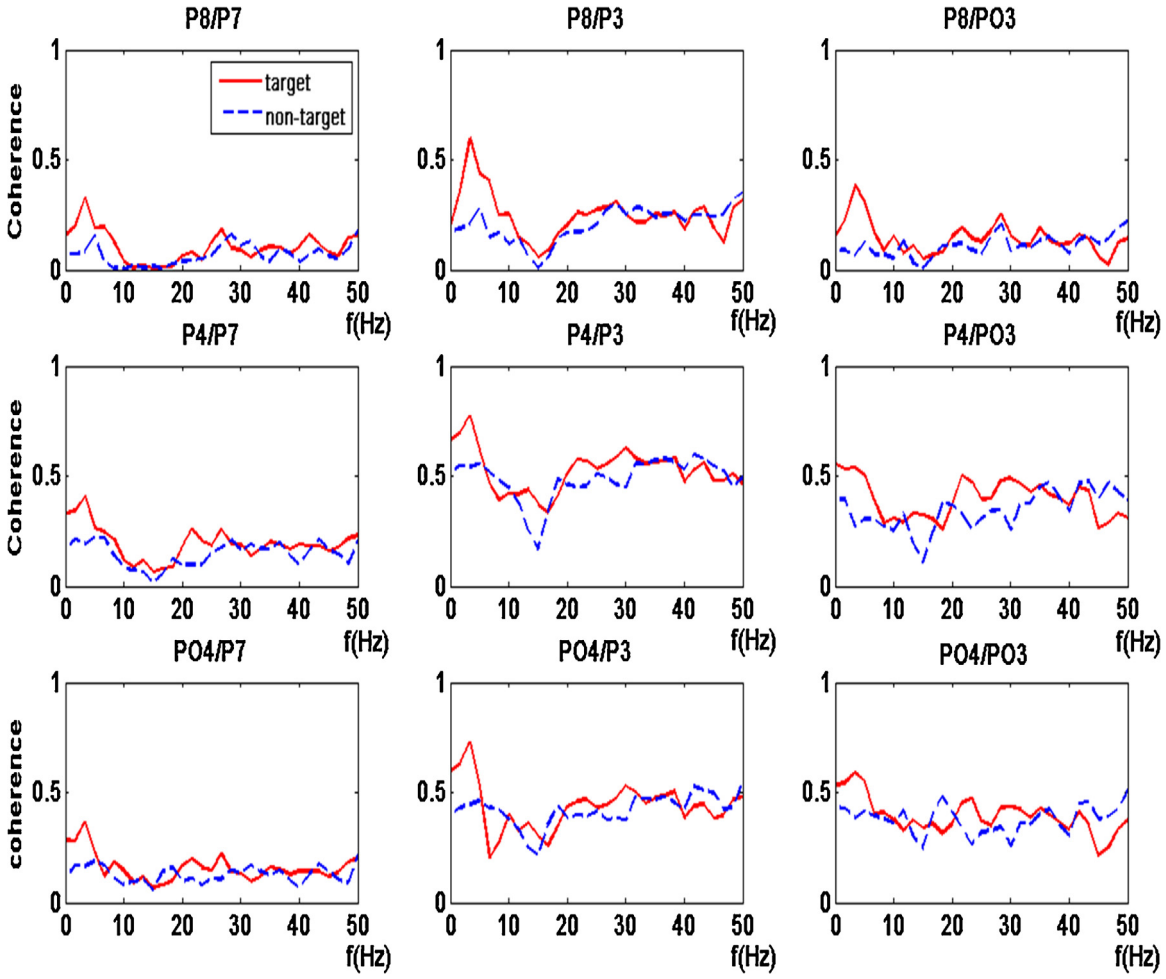

Fig. 8. The Magnitude Squared of Coherences (MSC) of the EEG samples in training subset of set1 (between P8/P7, P8/P3, P8/PO3, etc). The red solid lines are corresponding to target signals and blue dash lines to non-target signals. The coherence values of target signals are larger than that of non-target signals in the frequency band ranging from 1 Hz to 30 Hz. The low-frequency components of the EEG signals used in this BCI paradigm are closely associated with the target visual stimuli. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

ANNFL method), we preserve the EEG signal in the frequency band where the target-specific components exist.

For channel selection, in this paper, we define the correlation distances between EEG channels (see formula (2)).

$$d(x,y) = \left| \frac{(x(t)-\bar{x})(y(t)-\bar{y})}{\left\|x(t)-\bar{x}\right\| \left\|y(t)-\bar{y}\right\|} - 1 \right| \tag{2}$$

In formula (2), $x(t)$ and $y(t)$ are the average waves of the target samples detected in two EEG channels. $\bar{x}$ and $\bar{y}$ are the mean values of the average waves. Small correlation distance indicates that the brain electrical activities detected in these two channels are similar.

We can calculate the correlation distances between all channels used in the Imitating-Reading BCI paradigm and sort them in a matrix (see formula (3))

$$\left(d_{ij}\right)_{N\times N} = \begin{pmatrix} d_{11} & \cdots & d_{1N} \\ \vdots & \ddots & \vdots \\ d_{N1} & \cdots & d_{NN} \end{pmatrix} \tag{3}$$

In formula (3), the lower cases i, j are the channel-serial numbers, N is the total number of the channels used in the experiments. The diagonal elements $(d_{11}, d_{22}, \ldots, d_{NN})$ of the distance matrix are the distances between the channels with themselves, according to the definition these values are zeroes. With the correlation distance matrix, we apply hierarchical clustering method to all channels and cluster them in two groups. The EEG signals detected in the channels of one group are more relevant with target visual stimuli and the proof is given in Section 4.2.

3.3. The feature vector classification methods

The classifier is another important part of a practical P300-BCI system, it should be accurate, reliable and easy to use. Several studies have proposed some effective methods to enhance the classification performances of P300-BCI systems, such as fisher criterion (FC) based method [37], stepwise linear discriminant analysis (SWLDA) [16], Shrinkage linear discriminant analysis (SKLDA)[38] and spatial-temporal discriminant analysis (STDA)[39]. The design of a better classification method for P300-BCI systems is ongoing. These proposed methods cannot suit for every situation and, in the pattern recognition field, support vector machine (SVM) is another widely used classifier, the aim of SVM is to find optimal classification hyper plane that can separate the two classes of samples and maximize the distance between samples and the plane [40]. With the kernel methods, SVM can deal with the non-linear classification problem efficiently. Therefore, we choose SVM as the classifier in the study and compare the classification performance of it with FC method.

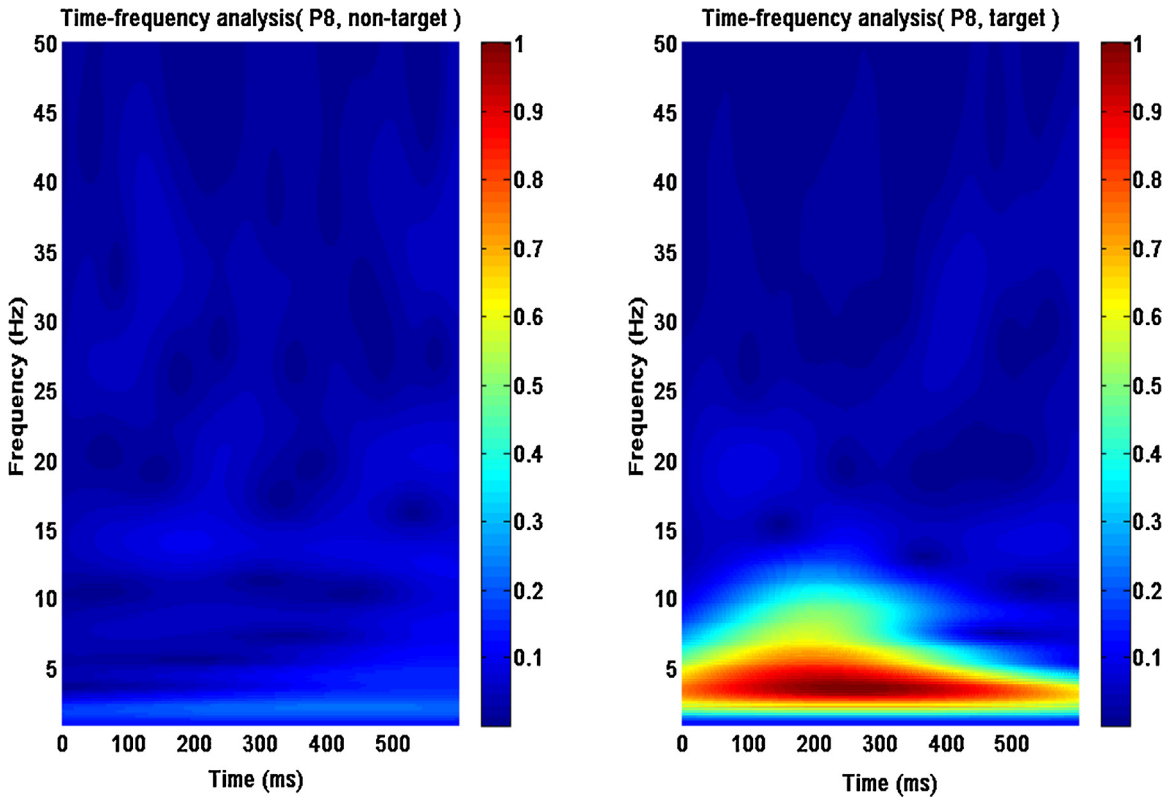

Fig. 9. The time-frequency (TF) maps of the mean non-target and target waves, zero time point is when visual stimulus received. We average the non-target and target samples in training subset of set1 respectively to form the mean waves. This figure shows the normalized results of Continuous Wavelet Transform (CWT) analysis of the mean waves of the samples detected in channel P8 (chosen as an example), the left for non-target samples and the right for target. TF maps are filled with colors according to the normalized magnitude of the components in EEG signals. Maximum magnitude specifies the data value to the red end of colormap and the minimum to the blue end. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

SVM solves the C-SVC optimization problem [41] in formula (4), where x_i is a training vector, y_i is the corresponding label and $C > 0$ is the regularization parameter.

$$\begin{aligned} &\min_{\omega,b,\xi_i} \frac{1}{2}\omega^T\omega + C\sum_{i=1}^{l}\xi_i \\ &sty_i\left(\omega^T\Phi(x_i) + b\right) \geq 1 - \xi_i \end{aligned} \tag{4}$$

We executed the classification procedure by using the software LIBSVM [42]. The SVM in use has a radial basis function (RBF) kernel (see formula (5)).

$$K(y, v) = \exp\left(-\gamma\|u - v\|^2\right) \tag{5}$$

For each subject, the ANNFL-extracted feature vectors are used as the training and testing samples. After that, in the training stage of SVM, we use 3-fold cross-validation technique and try the exponentially growing sequences of C and γ ($C = 2^{-5}, 2^{-4}, 2^{-3}, \ldots, 2^3, 2^4, 2^5$; $\gamma = 2^{-5}, 2^{-4}, 2^{-3}, \ldots, 2^3, 2^4, 2^5$) to find the best parameter C and γ.

4. Results

We recorded two datasets of EEG signals for each subject in two different weeks respectively and denoted these datasets from set1 to set10 (odd number datasets for week one, even number for week two, each set has 100 target samples and 100 non-target samples). Each dataset consists of a training subset and a testing subset. The training subset has 50 randomly-selected target samples and 50 randomly-selected non-target samples, the testing set has the rest EEG samples.

4.1. The suggestions of coherence analysis and time-frequency analysis

We calculate the coherences of the EEG signals in set1's training subset, which were detected in the left-sphere channels P7, P3, PO3 and right-sphere channels P8, P4, PO4. Fig. 8 illustrates the coherence difference between the target and non-target trials of EEG signals. The differences of low-frequency (under 30 Hz) EEG signals were significant. And the time-frequency maps of the mean waves of non-target and target EEG signals corresponding to channel P8 in training subset of set1 show that (see Fig. 9), in this BCI paradigm, the target-specific components (under 30 Hz) are explicit in the time range from 100 ms to 400 ms after the subject receives target-visual stimulus. To preserve the main difference between two kinds of EEG signal, we filter the EEG signals for each dataset, using a four-order IIR filter whose cut-off frequency is 30 Hz. After that, we select the signals in time range from 100 ms to 400 ms after the subjects received visual stimulus as new EEG samples.

4.2. Results of channel selection

After the filtering step, we average each subject's training EEG signals detected by different electrodes respectively. The average target and non-target EEG signals of Sub1's P8 channel are shown in Fig. 10. The ERP signal elicited by target stimuli is explicit. In our experiments, the N2 peak emerges at around 190 ms after the subject perceives target stimulus.

The EEG signals of channels (FP1, FP2, AF3, AF4, F7, F3, Fz, F4, F8) on frontal area are vulnerable to the effects of EOG and the amplitudes of the P300 signals of frontal channels are lower than that of other channels. Consequently we reject the frontal channels,

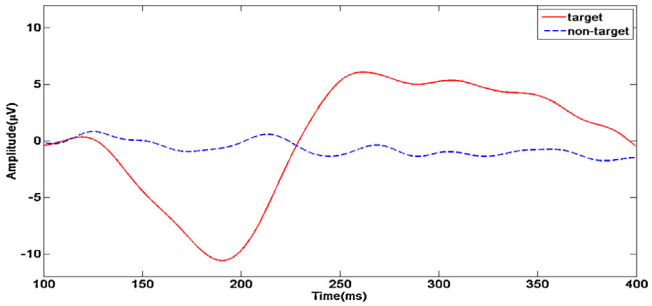

Fig. 10. The average target and non-target EEG signals of Sub1's P8 channel. The red solid line corresponds to target signals and blue dash line to non-target signals. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

then apply the channel-selection method proposed in Section 3.2 to the rest channels (see Fig. 11-a.). We calculate the correlation distances between the rest channels for each subject. According to the distances matrix of set1 (see Fig. 11b.), we create hierarchical cluster trees (see Fig. 12).

After clustering the channels into two groups, for set1 we apply the ANNFL method described in Section 3.3 to extract the features of the EEG samples detected in the channels of group A and group B respectively. Then the classification accuracies are listed in Table 1.

The channels were divided into two categories. Taking the results listed in Table 1 and the fact that, in this BCI paradigm, the power of target-specific components in the EEG signals detected in the channels on occipital area are higher than in other areas of brain [33] into consideration, we choose the group that contains most of the channels on occipital area. The results of channel selection for all

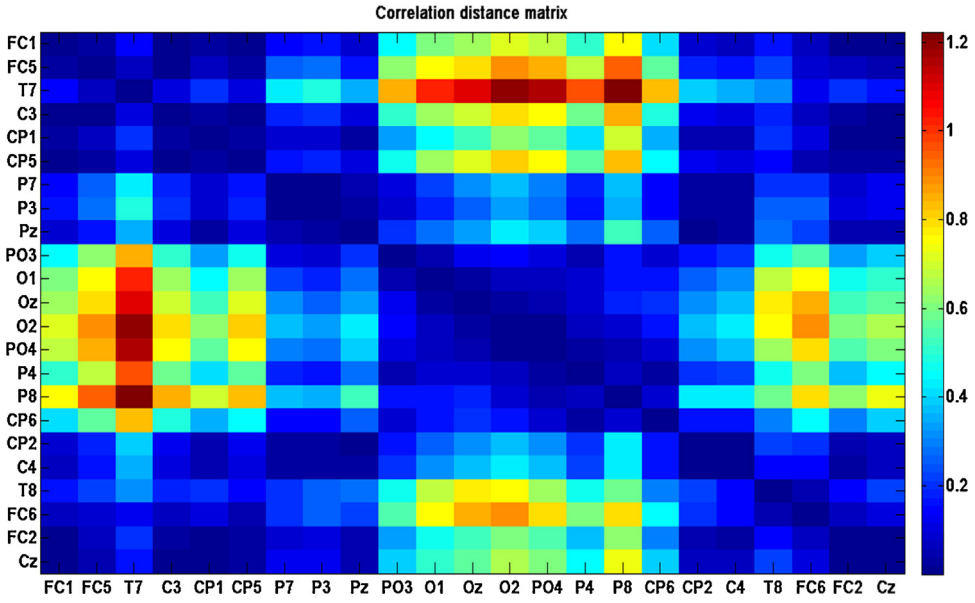

Fig. 11. The positions and correlation distances of channels in use for set1. Fig. 11 shows the distance matrix, which is filled with colors according to the distance. Maximum distance specifies the data value to the red end of colormap and the minimum to the blue end. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

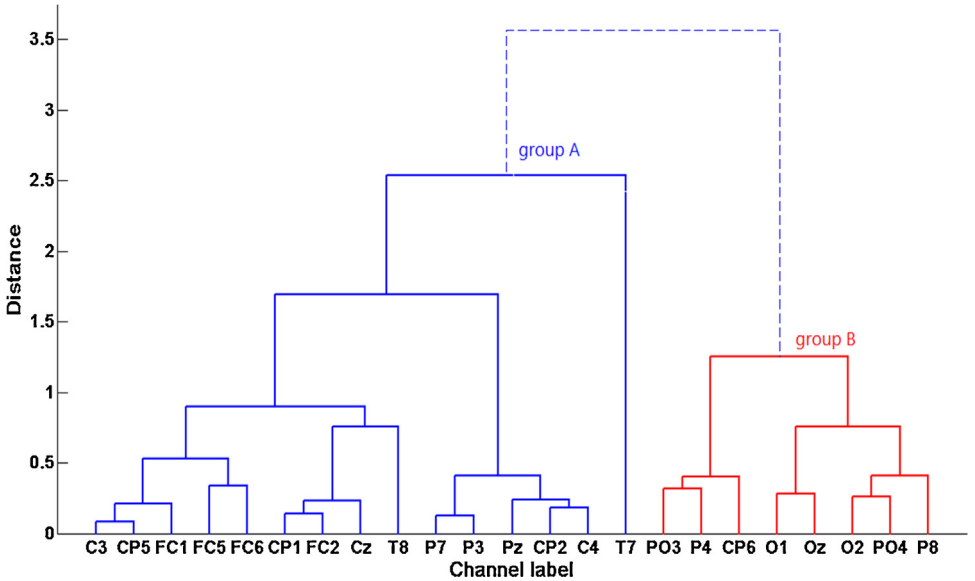

Fig. 12. The hierarchical cluster trees of channels in use for set1. These channels are divided into two groups. In this strategy, the correlations between the EEG signals within one group is stronger than the signals among the groups.

Table 1
The classification accuracies according to different groups of channels. We train two ANNs using the training samples detected in the channels of group A and group B, then extract the feature vectors from the samples in set1 and apply a SVM as the classifier used in classification experiments.

Group	The classification accuracies (%) and experiment number						
Label	1	2	3	4	5	m	sd
A	69	78	82	80	88	79.4	6.9
B	86	89	88	88	92	88.6	2.2

Table 2
The selected channels of all datasets in use.

Subjects	Datasets	Labels of chosen channels
Sub1	set1	PO3 O1 Oz O2 PO4 P4 P8 CP6
	set2	PO3 O1 O2 PO4 P4 P8 CP6 CP2
Sub2	set3	P7 PO3 O1 Oz O2 PO4 P4 P8 CP6
	set4	CP5 P7 P3 Pz PO3 O1 Oz O2 PO4 P4 P8 CP6 T8
Sub3	set5	P7 P3 PO3 O1 Oz O2 PO4 P4 P8 CP6
	set6	P7 P3 PO3 O1 Oz O2 PO4 P4 P8
Sub4	set7	CP5 P3 PO3 O1 Oz O2 PO4 P4 P8 CP6
	set8	CP5 P3 O1 Oz O2 PO4 P4 P8 CP6
Sub5	set9	O1 Oz O2 PO4 P4 P8 CP6
	set10	CP5 P7 P3 PO3 O1 Oz O2 PO4 P4 P8 CP6 T8

datasets were listed in Table 2. Because the datasets were recorded in two different weeks for each subject, the mental states cannot be the same even for one subject, the results of channel selection are different for each dataset. To quickly check the selection frequency of each channel, we plot a bar-chat of the selection frequencies of all channels (see Fig. 13). Finally, we choose the channels whose selection frequency is above the threshold (mean selection frequency of all channels) denoted in Fig. 13. Fig. 14 shows the results of channel selection experiment, signals detected in channel PO3, O1, Oz, O2, PO4, P4, P8 and CP6 are selected for the ANN based feature learning procedure.

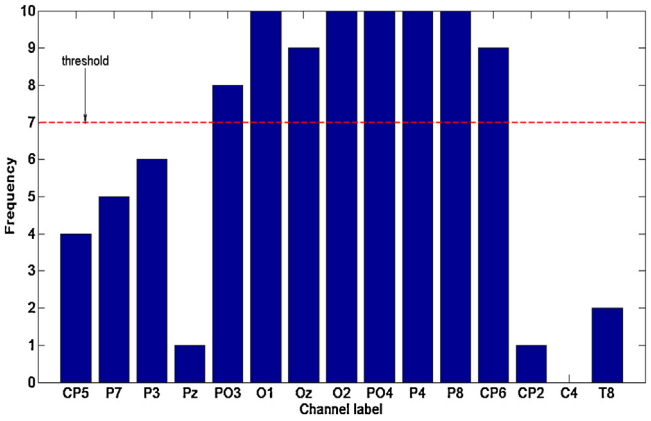

Fig. 13. The bar-chart of channnel selection for all datasets in use. The red line marks the threshold (mean selection frequency) for channel selection. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.3. Results of classification

We train a five-ganglion artificial feature extracting neural network (each ganglion has 240 input units, 30 hidden units and 240 output units) and a SVM classifier for each dataset, using the samples in its training subset (the training subset consists of 50 non-target samples and 50 target samples are chosen randomly from the dataset, the training stage is evaluated by 3-fold cross-validation), then apply the ANN to extract feature vectors of the samples in testing subset, which consists of the rest samples of the corresponding dataset, and do classification experiments. To validate the performance of whole feature extraction and classification procedure, we repeat the whole procedure (sample choosing, feature learning and classification) ten times. After that we redo the experiments but extract multi-channel EEG features using PCA method, because when reducing the dimensionality of an EEG signal matrix using this method, just like the multi-ganglion feature learning method, we do not need the label information. The principal component of each trial of EEG signal matrix is used as the feature vector. At last, we compare the classification accuracies in

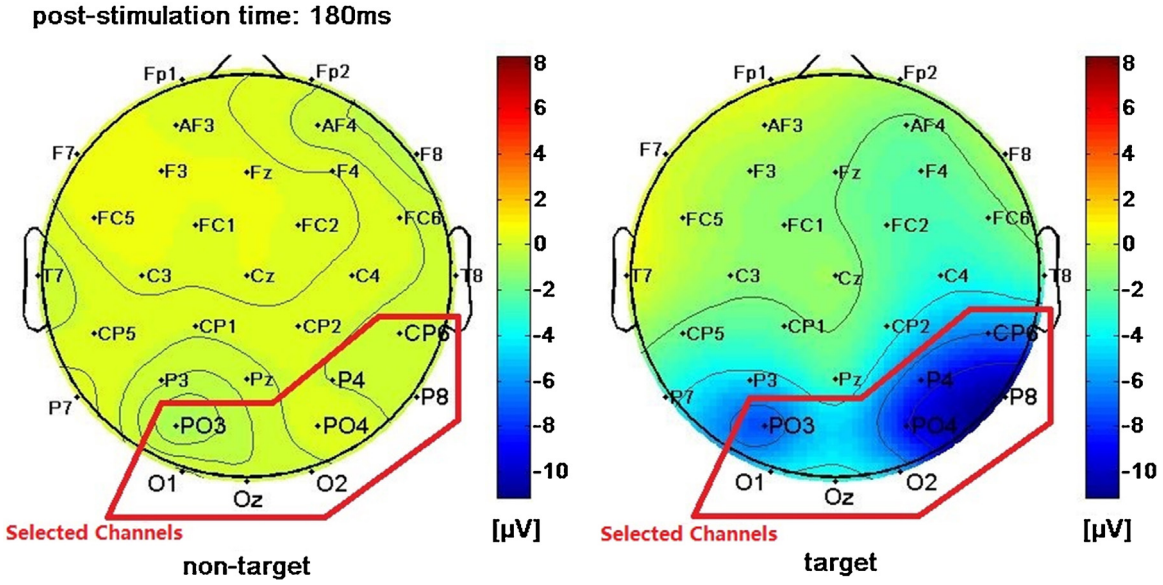

Fig. 14. The results of channel selection. The channels in red polygon are the selected channels. The brain electrical activity mapping (180 ms after subject1 receiving visual stimuli) shows the selected channels can cover the area where the main difference between non-target and target EEG signals exists for Imitating-Reading BCI paradigm. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

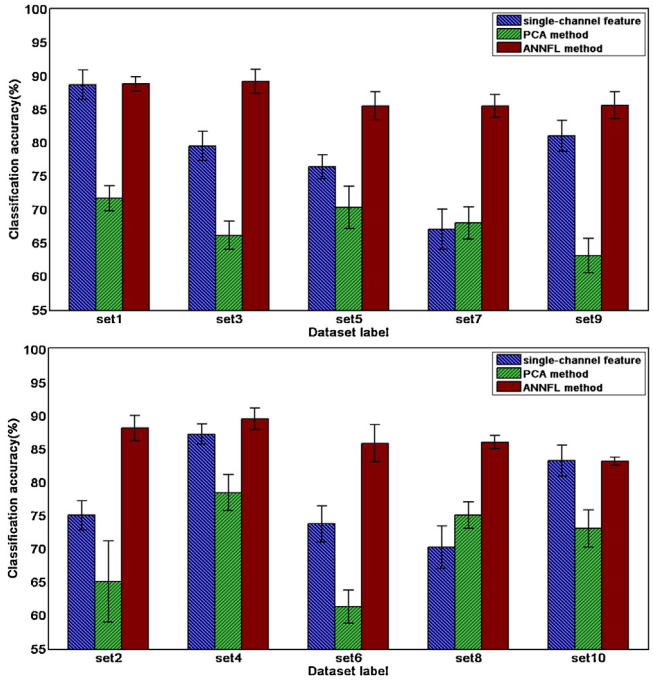

Fig. 15. The classification-accuracy comparison. The blue bars show the highest mean classification accuracies we got when successively using the single-channel temporal features of the channels we selected (PO3, O1, Oz, O2, PO4, P4, P8 and CP6) for all datasets, then the red bars show the results corresponding to ANNFL method and green bars to PCA method (upper Fig. for the dataset detected in week one, lower for week two). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

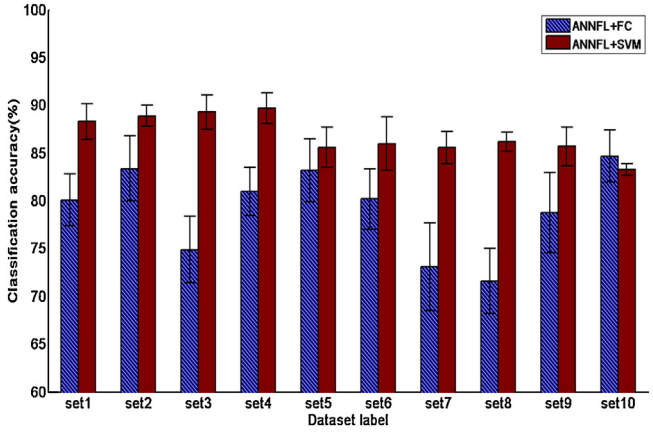

Fig. 16. The classification-accuracy comparison. The blue bars show the mean classification accuracies we got when apply the fisher criterion method to classify the feature vectors extracted by ANNFL method for all datasets, then the red bars show the results corresponding to ANNFL method with SVM. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

these two schemes with the results we got when using single-channel temporal EEG feature vectors as the samples for SVM classifier (see Fig. 15).

After that, we apply the fisher criterion method to classify the feature vectors extracted by ANNFL method. We repeat the whole procedure (sample choosing, feature learning and classification) ten times and compare the classification accuracy of fisher criterion linear discriminant analysis with the performance of SVM for each dataset (see Fig. 16). In this context, SVM outperforms the fisher criterion method (p-value of paired t-test < 0.025).

Table 3
The comparison of training time (2000 iterations) and classification accuracies between ANNFL method and conventional auto-encoder (using computer with 4-core 2.5 GHz Inter i5 with 6GB RAM).

Test	ANNFL (consists of five ganglion)		Auto-encoder (three-layer)	
Time	Time(s)	Accuracy (%)	Time(s)	Accuracy (%)
1	60.7	84	487.8	84
2	59.4	86	483.6	81
3	59.3	89	491.8	86
4	58.9	88	502.1	90
5	59.1	88	516.1	80
6	59.4	92	481.5	81
7	58.8	89	472.8	88
8	58.7	89	486.1	88
9	59.4	91	485.6	82
10	59.2	85	521.4	87
mean	59.3	88.1	492.9	84.7
sd	0.6	2.5	15.6	3.6

For set1, we also extract the feature vectors by using a three-layer conventional auto-encoder, which has 1200 input, 150 hidden and 1200 output units. The three-layer conventional auto-encoder has the same number of units as the five-ganglion artificial feature learning neural network. We perform this test 10 times, the comparison of training time and classification accuracies (obtained by using SVM as the classifier) is shown in Table 3.

5. Discussion

In this study, we propose the ANNFL method and apply it to the feature extracting procedure of Imitating-Reading BCI system. The result shows that the classification accuracies that corresponding to ANNFL method of the ten datasets are higher than that to PCA method (p-value of paired t-test < 0.025) or using temporal feature directly (p-value < 0.025).

Principal component analysis (PCA) is a typical widely used dimension reduction method, which can find the direction of greatest variance (principal component) in a dataset [18,25]. With this method, we can reduce the dimension of a dataset and preserve the most important information of a dataset.

However, for the Imitating-Reading BCI signals that contain many subtle event-related components, when PCA method applied to the signal matrices, some significant information lose (see Fig. 17). Then the inevitable information loss can lead to the lower classification accuracies.

In the classification procedure, when using the temporal feature directly, we do not always achieve the highest classification accuracies in the same EEG channel for all subjects. For practical use, it is a costly way to find the suitable EEG channel for each subject.

To solve this problem, in this study, we use hierarchical cluster method to divide the channels into two groups. The channel groups can cover a larger area than a single channel. When using the group containing most of the channels on occipital area, we can achieve higher classification accuracies. This channel selection method is more convenient and reliable than traditional single-channel selection method. It provides effective channel choice for ANNFL method and ushers in the higher classification accuracies on most of the datasets.

In the context of ANNFL method, the settings of the basic ganglions in an ANN are also significant to whole network. The number of the ganglions can affect the ANN training time and the feature learning performance considerably. Each ANN has 1200 input units, 150 hidden units and 1200 output units, if we train it as a traditional three-layer auto-encoder, there are $2 \times 1200 \times 150$ (from input layer to hidden layer and hidden layer to output layer: $2 \times$ input units number $\times$ hidden units number) connection weights to be adjusted. To make the whole network more efficient, we should

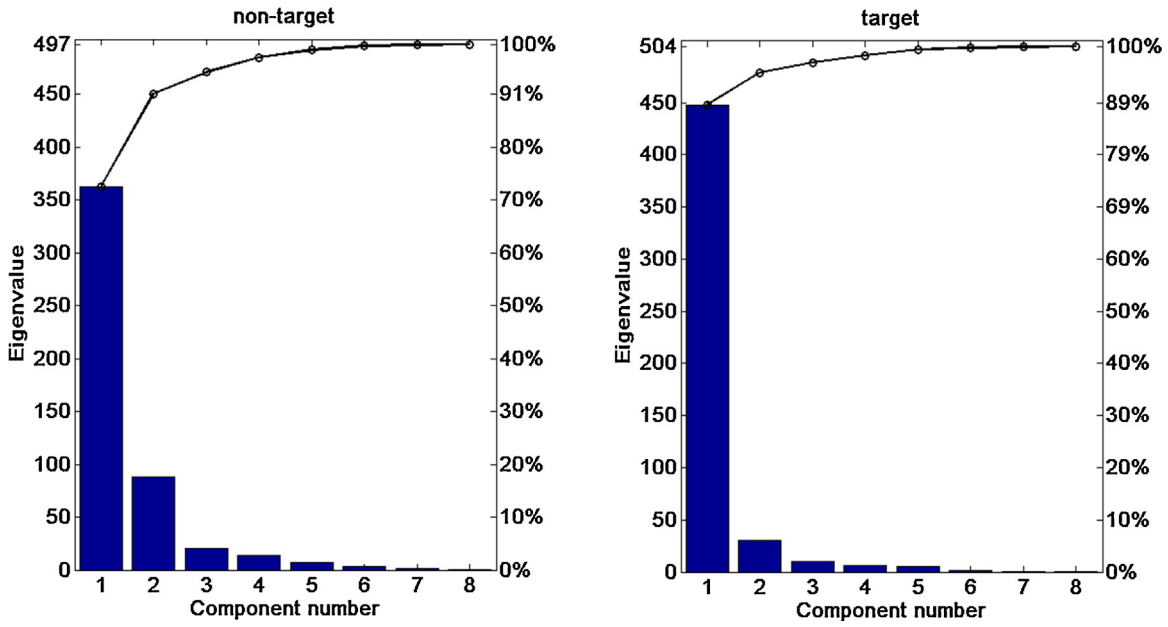

Fig. 17. The Pareto charts of the components in Imitating-Reading BCI signals. The left is the Pareto chart for a non-target sample and the right for a target sample. The first one component in each chart is the principal one, the eigenvalue corresponding to it of non-target sample takes 72% of the sum of all eigenvalues and 88% for target sample.

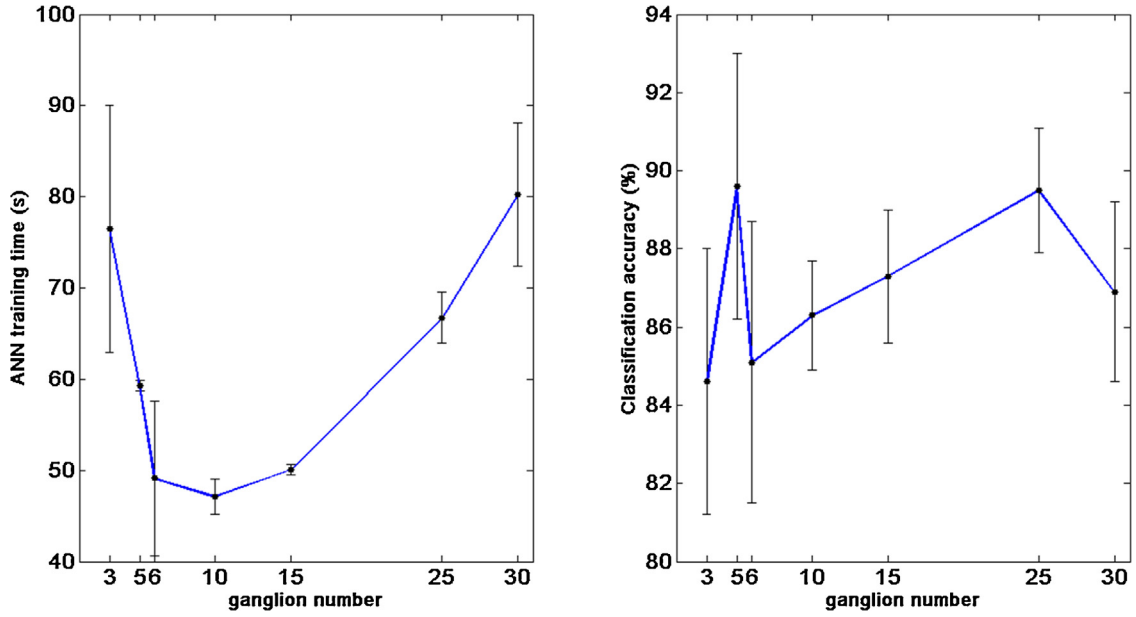

Fig. 18. The effects of the ganglion number on ANN training and classification. The left shows the relationship between the ganglion number and ANN training time (using computer with 4-core 2.5 GHz Inter i5 with 6 GB RAM). The minimum training time achieved when the ganglion number is set as 10. The right shows the relationship between the ganglion number and classification accuracy, highest classification accuracy obtained when five ganglions are used.

make the trade-off between the complexity of network structure and the feature learning performance. Each ganglion is trained 2000 iterations, the less ganglion we use, the more complicated a ganglion structure is. The structure parameters of ANNs with different number ganglion are listed in Table 4. Fig. 18 shows the effects of the number of ganglion on ANN training and classification procedure.

When we set five ganglions in the ANN, we can achieve the highest classification accuracy and the training time is acceptable. With this multi-ganglion ANN, we reduce the number of connection weights to $2 \times 240 \times 30 \times 5$ (from input layer to hidden layer and hidden layer to output layer: 2 × input units number × hidden units number × ganglion number), each ganglion just extracts EEG feature from a center part of a signal matrix, then we reduce the

Table 4
The structure parameters of all ANNs. Each ANN has 1200 input units, 150 hidden units and 1200 output units in all but we divide each ANN into different number of ganglions.

ANN Label	Number of ganglion	Number of units (each ganglion)		
		Input	Hidden	Output
A	3	400	50	400
B	5	240	30	240
C	6	200	25	200
D	10	120	15	120
E	15	80	10	80
F	25	48	6	48
G	30	40	5	40

training time of a large ANN remarkably and obtain more representative feature vectors for Imitating-Reading BCI signals than PCA method and conventional three-layer auto-encoder can provide.

Acknowledgment

The work is supported by National Nature Science Foundation of China (#91120017, #81271659) and Fundamental Research Funds for the Central Universities (#CZY13031). The authors would like to thank all the subjects in the experiment for their hard work.

References

[1] J.R. Wolpaw, N. Birbaumer, W.J. Heetderrks, D.J. McFarland, P.H. Peckham, G. Schalk, E. Donchin, L.A. Quatrano, C.J. Robinson, T.M. Vaughan, Brain–computer interface technology: a review of the first international meeting, IEEE Trans. Rehabil. Eng. 8 (2000) 164–173.
[2] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, T.M. Vaughan, Brain-computer interfaces for communication and control, Clin. Neurophysiol. 113 (2002) 767–791.
[3] J.N. Mak, D.J. McFarland, T.M. Vaughan, L.M. McCane, P.Z. Tsui, D.J. Zeitlin, E.W. Sellers, J.R. Wolpaw, EEG correlates of P300-based brain–computer interface (BCI) performance in people with amyotrophic lateral sclerosis, J. Neural Eng. 9 (2012) 026014.
[4] X. Chen, Z. Chen, S. Gao, X. Gao, Brain–computer interface based on inter modulation frequency, J. Neural Eng. 10 (2013) 066009.
[5] K. Xu, Y. Wang, Y. Wang, F. Wang, Q. Zhang, S. Zhang, W. Chen, X. Zheng, Local-learning-based neuron selection for grasping gesture prediction in motor brain machine interfaces, J. Neural Eng. 10 (2013) 026008.
[6] K. LaFleur, K. Cassady, A. Doud, K. Shades, E. Rogin, B. He, Quadcopter control in three-dimensional space using a noninvasive motor imagery-based brain–computer interface, J. Neural Eng. 10 (2013) 046003.
[7] B. Choi, S. Jo, A low-cost EEG system-based hybrid brain-computer interface for humanoid robot navigation and recognition, PLoS ONE 8 (9) (2013) e74583.
[8] J. Shen, J. Liang, J. Shi, Y. Wang, A dynamic submatrix-based P300 online brain–computer interface, Biomed. Signal Process. Control 15 (2015) 27–32.
[9] A.C. Atencio, T.F.B. Filho, A. Ferreira, A.B. Benevides, Evaluation of ERD/ERS caused by unpleasant sounds to be applied in BCIs, Proceedings of the Biosignals and Biorobotics Conference (BRC), (2013), http://dx.doi.org/10.1109/BRC.2013.6487533
[10] Z. İşcan, Z. Dokur, A novel steady-state visually evoked potential-based brain-computer interface design: character plotter, Biomed. Signal Process. Control 10 (2014) 145–152.
[11] P. Poryzala, A. Materka, Cluster analysis of CCA coefficients for robust detection of the asynchronous SSVEPs in brain–computer interfaces, Biomed. Signal Process. Control 10 (2014) 201–208.
[12] T. Jiang, X. Wu, S. Zhang, Functional localization of the cortical motor area in the brain based on wavelet analysis of slow cortical potential, Proceedings of the IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), (2012) 317–320.
[13] J.K. Steven, An Introduction to the Event-Related Potential Technique, The MIT press, Cambridge, Massachusetts, USA, 2005, pp. 40–58.
[14] L.A. Farwell, E. Donchin, Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials, Clin. Neurophysiol. 70 (1988) 510–523.
[15] U. Hoffmann, J.M. Vesin, T. Ebrahimi, K. Diserens, An efficient P300-based brain–computer interface for disabled subjects, J. Neurosci. Methods 167 (2008) 115–125.
[16] E.W. Sellers, E. Donchin, A P300-based brain–computer interface: initial tests by ALS patients, Clin. Neurophysiol. 117 (2006) 538–548.
[17] J.N. Mak, Y. Arbel, J.W. Minett, L.M. McCane, B. Yuksel, D. Ryan, D. Thompson, L. Bianchi, D. Erdogmus, Optimizing the P300-based brain–computer interface: current status, limitations and future directions, J. Neural Eng. 8 (2011) 025003.
[18] X. Yu, P. Chum, K.-B. Sim, Analysis the effect of PCA for feature reduction in non-stationary EEG based motor imagery of BCI system, Int. J. Light Electron Opt. 125 (2014) 1498–1502.
[19] Z.J. Koles, The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG, Electroencephalogr. Clin. Neurophysiol. 79 (1991) 440–447.
[20] Z.J. Koles, J.C. Lind, P. Flor-Henry, Spatial patterns in the background EEG underlying mental disease in man, Electroencephalogr. Clin. Neurophysiol. 91 (1994) 319–328.
[21] J. Müller-Gerking, G. Pfurtscheller, H. Flyvbjierg, Designing optimal spatial filter for single trial EEG classification in a movement task, Clin. Neurophysiol. 110 (1999) 787–798.
[22] H. Ramoser, J. Müller-Gerking, G. Pfurtscheller, Optimal spatial filtering of single EEG during imagined hand movement, IEEE Trans. Rehabil. Eng. 8 (2000) 441–446.
[23] B.H. Yang, G.Z. Yan, R.G. Yan, T. Wu, Feature extraction for EEG-based brain–computer interfaces by wavelet packet best basis decomposition, J. Neural Eng. 3 (2006) 251.
[24] A.C. Tang, T. Sutherland, Y. Wang, Contrasting single trial ERPs between experimental manipulations: improving differentiability by blind source separation, NeuroImage 29 (2006) 335–346.
[25] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (2006) 504–507.
[26] P. Vincent, H. Larochelle, Y. Bengio, P.-A. Manzagol, Extracting and composing robust features with denoising auto-encoders, Proceedings of the International Conference on Machine Learning, 592 (2008) 1096–1103.
[27] J. Ngiam, P. Koh, Z. Chen, S. Bhaskar, A.Y. Ng, Sparse Filtering, Proceedings of the Neural Information Processing Systems Conference, 4334 (2011)1–9.
[28] B. Mitchell, J. Sheppard, I. Arel, D. Rose, Deep spatiotemporal feature learning with application to image classification, Proceedings of the Ninth International Conference on Machine Learning and Applications, (2010) 883–888.
[29] D. Yu, G. Hinton, N. Morgan, J.-T. Chien, S. Sagayama, Introduction to the special section on deep learning for speech and language processing, IEEE Trans. Audio, Speech, Lang. Process. 20 (2012) 4–6.
[30] H. Larochelle, Y. Bengio, Jérôme Louradour et Pascal Lamblin, exploring strategies for training deep neural networks, J. Mach. Learn. Res. 1 (2009) 1–40.
[31] C. Yaguang, Y. Zhongle, L. Haihua, C. Junbo, Single trial ERP extraction for brain-computer interface, Proceedings of the 2nd International Symposium on Instrumentation Science and Technology, (2002) 3712–3716.
[32] G. Jin-an, C. Yaguang, Design and performance virtual keyboard, Chin. J. Clin. Rehabil. 10 (2006) 124–126.
[33] G. Jin-an, W. Feng-yan, C. Yaguang, Selection of features in single-trial BCI signal estimation, J. Cent. China Norm. Univ. (Nat. Sci.) 40 (2006) 193–196.
[34] N. Prasad, R.Singh, S.P. Lal, Comparison of back propagation and resilient propagation algorithm for spam classification, Proceedings of the Fifth International Conference on Computational Intelligence, Modelling and Simulation, (2013) 29–34.
[35] S.S. Shai, B.D. Shai, Understanding Machine Learning: From Theory to Algorithms, Cambridge University Press, Cambridge, United Kingdom, 2014, pp. 228–236.
[36] S.A.S. Filho, C.J. Tierra-Criollo, A.P. Souza, M.A.S. Pinto, M.L.C. Lima, G.M. Manzano, Magnitude squared of coherence to detect imaginary movement, EURASIP J. Adv. Signal Process. 2009 (2009) 1–12, http://dx.doi.org/10.1155/2009/534536.
[37] G. Pires, U. Nunes, M. Castelo-Branco, Statistical spatial filtering for a P300-based BCI: test in able-bodied, and patients with cerebral palsy and amyotrophic lateral sclerosis, J. Neurosci. Methods 195 (2) (2011) 270–281.
[38] B. Blankertz, S. Lemm, M. Treder, S. Haufe, K. Müller, Single-trial analysis and classification of ERP components—a tutorial, NeuroImage 56 (2) (2011) 814–825.
[39] Y. Zhang, G. Zhou, Q. Zhao, J. Jin, X. Wang, A. Cichocki, Spatial-temporal discriminant analysis for ERP-based brain-computer interface, IEEE Trans. Neural Syst. Rehabil. Eng. 21 (2) (2013) 233–243.
[40] T.N. Lal, M. Schroder, T. Hinterberger, J. Weston, M. Bogdan, N. Birbaumer, Scholkopf Bernhard, support vector channel selection in BCI, IEEE Trans. Biomed. Eng. 51 (2004) 1003–1010.
[41] C. Cortes, V. Vapnik, Support-vector network, Machine Learning 20 (1995) 273–297.
[42] C-C. Chang, C-J. Lin, LIBSVM: a library for support vector machines, ACM Trans. Intell. Syst. Technol. 2 (2011) 1–27.