Research on Feature Extraction and Classification of P300 EEG Signals

Ye Ma[1,*], Guangping Jiang[2], Tanqing Chang[1] and Libin Guo[1]

[1]Academy of Armored Force Engineering, Beijing

[2]Hospital of Traditional Chinese Medicine, Urumqi

[*]Corresponding author

Abstract—**For the feature of P300's low signal-to-noise ratio (SNR) and difficult classificion, in this paper, we use an EEG signal processing method based on independent component analysis(ICA) and support vector machine. First, make the P300 EEG singal to the superposition averaging, according to the requirements of the ICA algorithm, the superimposed average signal is de-averaged and whitened. Then, the fast fixed-point algorithm which called FastICA is used to extract the feature vector of P300 EEG signal, in the end, put the feature vector into the support vector machine for classification. Using the DataSet II datasets in the International BCI Contest III to verify, the highest classification accuracy of the algorithm is 90.12%. The principle of this algorithm is simple, can successfully extract the feature of P300 EEG signal, and provide reference method for P300 EEG feature extraction and classification.**

Keywords-P300 EEG signal; feature extraction; independent component analysis; support vector machines

I. INTRODUCTION

In recent years, as the world's population continues to increase and aging intensified, muscle atrophic lateral sclerosis, paralysis, Alzheimer's disease and other patients also increased the base, to the community and the families of patients with a heavy burden. And in past several decades, brain-computer interface(BCI) came into being, there have been involved in multidisciplinary, multi-field ,such as Neuroscience, cognitive science, computer science, control engineering, medicine, and so on [1]. BCI is a technology to establish a brain associated with a computer or other device, this association not use the usual brain output pathway (the peripheral nerve and muscle tissue of the brain) [2]. This form of human-computer interaction can replace human language and limb movements, Making the restore and enhance the human body, psychological function, thinking and mind control to possible. So there are applications in the military target search [3], Flight simulator control [4], Car driving [5], new type game entertainment [6] and help people with disabilities who are experiencing problems with exercise or sensation to resume information and communication functions [7] are all have great potential.

The performance of the brain-computer interface system is mainly determined by the EEG signal processing module. The core of EEG signal processing module is composed of feature extraction and classification. There are many Common EEG signal extraction methods, for different EEG signals have different methods. For instance, Power spectrum analysis and fast Fourier transform (FFT) are time domain analysis methods, these methods are suitable for P300, N400 and other which latent period and waveform is constant, stimulate the relationship between the strict locking of the Evoked EEG signal. There are autoregressive models and digital filters in the frequency domain analysis method, it applies to the frequency feature of motor image the EEG signal. Time-frequency domain analysis method has wavelet transform, it is suitable for EEG signals with constant time-frequency feature over time. The above method is better in real time, more widely used, but can not directly express the relationship between the EEG lead. Spatial domain feature extraction methods are common space mode (CSP), independent component analysis, etc, this method utilize the spatial distribution and correlation information between the EEG signals, and is generally used for multi-channel EEG feature extraction[8-10].

In terms of classification, the classification methods used in BCI include artificial neural network, support vector machine (SVM), K means clustering, genetic algorithm, etc. Support vector machine is widely used in the classification of EEG signals. It has strong adaptability and high classification accuracy in the classification and recognition problems with strong randomicity and nonlinearity.

II. INTRODUCTION TO EXPERIMENTAL DATA

A. Event-Related Potential P300

Event-related potential (ERP) is that people experience some kind of stimulus event, the potential changes of a series of brain electrical activity evoked by information processing in the scalp, is a psychological or linguistic factors involved in the special evoked potential, 1965 by Sutton first reported. From the scalp record to ERP has two feature, first of all, there is a strict locking relationship between latent period and stimulation, second, it has a specific waveform and potential distribution. ERP is composed of two parts: source and endogenous, Exogenous ingredients include P100, N100, P200, The latent period is short, which is affected by the physical feature of stimulation. Endogenous components include N200, P300, influenced by psychological factors, related to people's attention, memory and other cognitive processes.

P300 is the most widely used endogenous event related potential, because of the latent period of more than 300ms and it is a positive wave, hence the name, it is also known as P300. The current research results show that, P300 is the result of

combined cortical activity, associated with complex multi-level mental activity (cognitive process), which is the feeling, perception, memory, and other psychological changes in the process of reflection, it is the reaction of objective things on the process. therefore, P300 is an objective index that does not need to rely on external behavior to judge the cognitive process of the subject, or it can be said to be an objective indicator of the advanced function of the brain.

B. Experimental Data Selection

The experimental data is taken from the DataSetII data in the International BCI Contest III. This dataset represents a complete record of P300 evoked potentials recorded with BCI2000 using a paradigm described by Donchin et al., 2000, and originally by Farwell and Donchin, 1988[11]. In these experiments, a user focused on one out of 36 different characters. The objective in this contest is to predict the correct character in each of the provided character selection epochs.

The user was presented with a 6 by 6 matrix of characters, see Figure 1. The user's task was to focus attention on characters in a word that was prescribed by the investigator, one character at a time. All rows and columns of this matrix were successively and randomly intensified at a rate of 5.7Hz. Two out of 12 intensifications of rows or columns contained the desired character, one particular row and one particular column. The responses evoked by these infrequent stimuliare different from those evoked by the stimuli that did not contain the desired character and they are similar to the P300 responses previously reported. The 2 out of 12 stimuli that did contain the desired character.

FIGURE I. THE CHARACTER MATRIX IN FRONT OF THE SUBJECTS

The datasets collected signals (bandpass filtered from 0.1-60Hz and digitized at 240Hz) from two subjects in five sessions each. Each session consisted of a number of runs. In each run, the subject focused attention on a series of characters. For each character epoch in the run, user display was as follows: the matrix was displayed for a 2.5 s period, and during this time each character had the same intensity, the matrix was blank. Subsequently, each row and column in the matrix was randomly intensified for 100ms, resulting in 12 different stimuli-6 rows and 6 columns). After intensification of arow/column, the matrix was blank for 75ms. Row/column intensifications were block randomized in blocks of 12. The sets of 12 intensifications were repeated 15 times for each character epoch, any specific row/column was intensified 15

times and thus there were 180 total intensifications for each character epoch. Each character epoch was followed by a 2.5 s period, and during this time the matrix was blank. This period informed the user that this character was completed and to focus on the next character in the word that was displayed on the top of the screen, The current character was shown in parentheses.
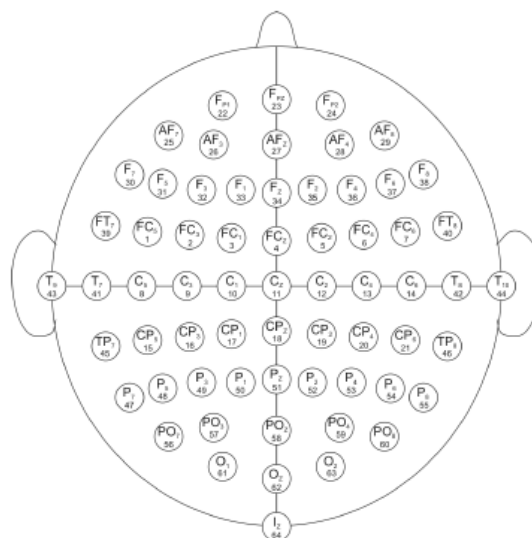

FIGURE II. SCHEMATIC DIAGRAM OF ELECTRODE

FIGURE III. P300 CHARACTER SPELLING PARADIGM

III. P300 EEG FEATURE EXTRACTION

A. Principle of Independent Component Analysis

The basic idea of ICA is to decompose the multi-channel observation signals into several independent components according to the principle of statistical independence, to achieve signal enhancement and analysis. Let $\{S_i; i = 1, 2, \cdots, N\}$ be an independent signal source of N, $\{x_i; i = 1, 2, \cdots, M\}$ is M observation signal $(M \geq N)$, Which,

$$x_i = \begin{bmatrix} x_i(0) & x_i(1), & \cdots & x_i(L-1) \end{bmatrix},$$

$$S_i = \begin{bmatrix} S_i(0) & S_i(1), & \cdots & S_i(L-1) \end{bmatrix}.$$ They are expressed in the form of vectors, $S = \begin{bmatrix} S_1 & S_2, \cdots S_N \end{bmatrix}^T$, $x = \begin{bmatrix} x_1 & x_2, \cdots x_M \end{bmatrix}^T$. The

components in X are stacked by the independent sources in S, which is

$$x_j = a_{j1}S_1 + a_{j2}S_2 + \cdots a_{jN}S_N = \sum_{k=1}^{N} a_{jK}S_K \qquad (1)$$
$$(j = 1, 2, \cdots, M)$$

The above formula can be expressed as a matrix:

$$X = AS \qquad (2)$$

where X is a regular coefficient mixed matrix, A is an unknown coefficient. A and S under the premise of unknown, looking for a solution of mixed matrix W, making

$$Y = WX \qquad (3)$$

In the formula $Y = \begin{bmatrix} y_1 & y_2 & \cdots y_N \end{bmatrix}^T$, and the components of Y are independent of each other, That is, Y is a good estimate of the source signal S, and it can be said that S is separated in Y.

The ICA algorithm can be divided into two types: iterative estimation method based on information theory and algebraic method based on statistics. In principle, they use the independence and non Gauss property of the source signal. Based on the research of information theory, Scholars from various countries have proposed a series of estimation algorithms from the perspective of maximum entropy, minimum mutual information, maximum likelihood and negative entropy. Such as FastICA algorithm and Infomax algorithm, maximum likelihood estimation algorithm, etc. Statistical methods based on second-order cumulant, fourth-order cumulant and other high-order cumulant method. This article uses the FastICA algorithm based on the largest negative entropy. It is the largest negative as a search direction, can be achieved in order to extract the independent source. The negative entropy is defined as

$$N_g(Y) = H(Y_{Gauss}) - H(Y) \qquad (4)$$

Y_{Gauss} is a Gauss random variable with the same covariance as Y,H () as a derivative of random variables.

$$H(Y) = -\int p_Y(\xi) \lg p_Y(\xi) d\xi \qquad (5)$$

Fast ICA learning rule is to find a direction so that $W^T X (Y = W^T X)$ has the largest non Gauss property. So the iterative formula of FastICA algorithm is

$$W^* = E\{X_g(W^T X)\} - E\{g(W^T X)\}W \qquad (6)$$
$$W = W^* / \|W^*\|$$

where W^* is the new value of W.

B. Feature Extraction of EEG Signals Based on FastICA

Because P300 EEG signal amplitude is only 2~10 μV, while the spontaneous EEG signal amplitude is about 10~100 μV, It is submerged in the spontaneous EEG signal, the two constitute the relationship between small signal and large noise, Therefore, the P300 signal should first be superimposed average. In this paper, the EEG signal data of 1s after each stimulation were selected for 15 times, Effect shown in Figure 4. To facilitate the display effect, only show the results of Cz lead, followed by Cz lead. It can be seen that the EEG signal containing P300 has a positive peak near 300 ms.
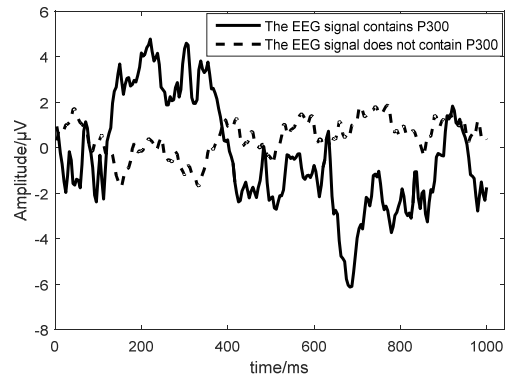

FIGURE IV. SIGNAL SUPERIMPOSED AVERAGE EFFECT DIAGRAM

In the vast majority of ICA algorithms, it is assumed that the mean of the source signal s is zero, Therefore, in order to make the actual blind source separation problem can meet the ICA mathematical model, Before applying the ICA method to the observed signal x, the following de-averaging preprocessing is usually performed first.

$$Y = x - E[x] \qquad (7)$$

The de-averaged process is only to simplify the ICA algorithm, After estimating the separation matrix w from the vector Y after the de-averaged, the mean vector WE [x] needs to be added back to the source estimate. The effect of the signal which has been superimposed averag is shown in figure 5.
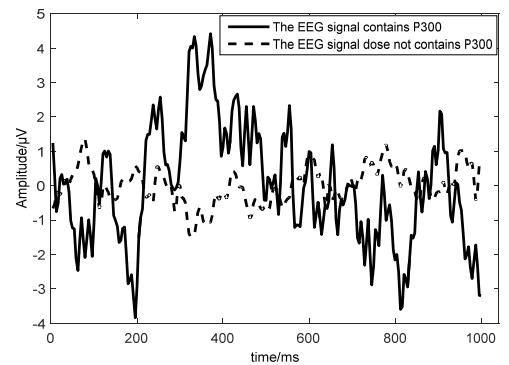

FIGURE V. SIGNAL TO DE-AVERAGED EFFECT DIAGRAM

To further simplify the ICA algorithm, the signal is whitened. The so-called whitening imposes a linear transformation on the observed signal vector x such that the individual components of the new vector Z are uncorrelated, while the covariance matrix of Z is a unit matrix $E\left(ZZ^T\right)=I$,

At this time that the new vector z for the space white or referred to as white. The effect of whitening the signal on the de-averaged is shown in figure 6.
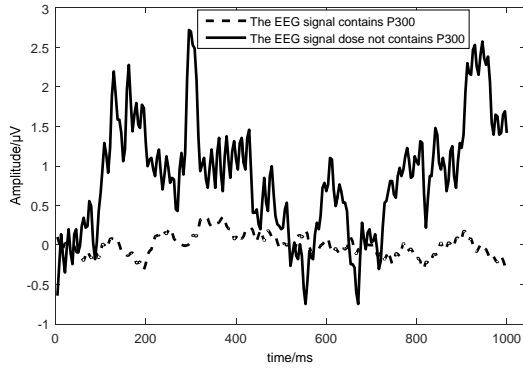

FIGURE VI. SIGNAL WHITENING EFFECT DIAGRAM

The expectations in the FastICA algorithm must be replaced by their estimates value, the best estimate is the corresponding sample average. Ideally, all valid data should be involved in the calculation, but this will reduce the computational speed. So the average number of samples is usually estimated, the number of samples on the final estimate of the accuracy of a great impact. The sample points in the iteration should be selected separately, and if the convergence is not ideal, the number of samples can be increased. So the experimental sample selected 15 characters per character after the superimposed average EEG signal, the FastICA is processed to extract the eigenvalues, and then sent to the support vector machine for classification.

IV. EEG SIGNAL CLASSIFICATION

A. Principle of Support Vector Machine

Define $m/\|\omega\|$ for the margin，Where m is the distance from the decision face along the ω direction to the nearest training sample, Each class has at least one m. Distance from the decision-making surface of the recent training examples for the support vector. The support vector machine is used to calculate the classification decision surface which maximizes the interval $m/\|\omega\|$, so that the data of the original space linearly indivisible can be linearly separable in the transformed high space. This 6-lead P300 signal is more prominent, FZ, CZ, PZ, OZ，C3, C4. The corresponding EEG signals can be selected as input to the FastICA algorithm as needed. 100 sets of EEG signals were randomly selected as training data, and the remaining 80 groups of signals were used as test data to check whether the classification effect was effective. In this paper, the support vector machine kernel function is Radial Basis Function:

$$H(\mathrm{x},\mathrm{x}^{'})=\exp(-\gamma\left\|x-x^{'}\right\|^2) \qquad (8)$$

If the kernel function $H(\mathrm{x},\mathrm{x}')$ satisfies the Mercer condition, the kernel function may correspond to the dot product in the corresponding transformation space. Which γ is a positive number of control radius. Expand (8):

$$H(\mathrm{x},\mathrm{x}^{'})=\exp(-\gamma\left\|x\right\|^2)*\exp(-\gamma\left\|x^{'}\right\|^2)* \\ \exp(2\gamma\,\mathrm{x}^T\mathrm{x}^{'}) \qquad (9)$$

All of the $\exp(-\gamma\left\|x\right\|^2)$, $\exp(-\gamma\left\|x^{'}\right\|^2)$, $\exp(2\gamma\,\mathrm{x}^T\mathrm{x}^{'})$ meet the Mercer condition.

B. Classification Results

In order to consider the experimental results, the influence of the number of different leads on the accuracy of classification is tested. Each test is done 10 times in order to find the average classification accuracy. The classification results are shown in Table I.

TABLE I. CLASSIFICATION RESULTS

the number of leads	Average classification accuracy rate
1	60.37%
2	73.49%
3	80.52%
4	84.33%
5	87.23%
6	89.65%
7	90.12%
8	86.44%
9	77.38%

It can be seen that the accuracy of the average classification increases as the number of leads increases. When increasing to 7 leads, the average classification rate is the highest, up to 90.12%. But more than seven lead, the average classification accuracy and a downward trend. It can be seen that the increase in the amount of data can improve the average classification accuracy, but the blind increase will make FastICA convergence effect is not ideal in the iteration. The method used in this paper can effectively extract and classify the P300 EEG signal, and the classification rate is ideal for the study of P300 EEG signal processing.

V. CONCLUSION

In this paper, according to the feature of low SNR of P300 ERP signal and strong randomness, the EEG signal is processed by superposition averaging, de-averaged and whitened, Secondly, the FastICA algorithm is used to extract the feature vector of EEG signal. Finally, the feature vectors of

the EEG signals are classified by SVM, and the highest average classification accuracy is up to 90.12%. Therefore, the feature extraction and classification of EEG signals can be effectively carried out by P300. The algorithm used in this paper is simple and the higher accuracy of EEG classification can enhance the stability of brain-computer interface. It provides a research idea for real-time online brain-computer interface system, so that the brain-computer interface can be used more widely in the military, games, medical and other fields.

REFERENCES

[1] Rajesh P. N. Rao. Brain-Computer Interfacing:an introduction[M].China Machine Press, 2016.

[2] Schalk G, Dennis J, Thilo H, et al. BCI2000: A general-purpose brain-computer interface(BCI) system [J]. IEEE Transactions on Biomedical Engineering, 2004, 51 (6):1034-1043.

[3] Robbin A. Miranda, William D. Casebeerb, Amy M. Heinz, et al. DARPA-funded efforts in the development of novel brain-computer interface technologies[J]. Journal of Neuroscience Methods, 2015, 244: 52-57.

[4] Xing Li-chao. Brain-machine interface research status [J]. Science and Technology Innovation and Application, 2015 (6): 24-24.

[5] Guan Wei, Yang Liu, Jiang Shixiong, Zhang Wenyi. Overview of the application of EEG in traffic driving behavior [J]. Transportation System Engineering and Information, 2016, 16 (): 35-44.

[6] Wang Fei, Zhang Yuzhong, Ning Ting Hui, Wen Shiguang. Research progress of brain computer interface [J]. Journal of intelligent systems, 2011, 6 (): 189-199.D

[7] Li Bo, Liu Yang, Zhou Zongtan. Computer simulation and implementation of [J]. Computer Simulation, 2007, 24 (7): 292-295

[8] Jia Huaping, Zhao Junlong. EEG signal analysis method and brain machine interface technology [M]. Beijing: Science Press, 2016

[9] Duan Suolin, Shang Yunkun, Pan Lizheng. Multi class motor imagery EEG feature extraction and classification of [J]. Computer Measurement and Control, 2016, 24 (2): 283-287.

[10] Han Min, Sun Zhuoran. Based on wavelet transform and classification of EEG in epilepsy Carbuncle AdaBoost extreme learning machine [J]. Computer Measurement and Control, 2015, 35 (9): 2701-2709.

[11] onchin E, Spencer K M, Wijesinghe R. The mental prosthesis: Assessing the speed of a P300-based brain-computer interface[J]. IEEE Transactions on Rehabilitation Engineering, 2000,8(2):174-179.