A Novel Combination Method To Detect P300 Peak In Raw EEG Signal

Malihe Nasr Azadani (1) – Rasoul Amir Fattahi(2) – Hossein Pourghassem(3)

(1,3) Department of Electrical Engineering, Najafabad Branch I, slamic Azad University, Iran
(2) Department of Electrical Engineering, Esfahan University Of Technology, Iran

Abstract_In this paper different feature extraction methods are compared to detect p300 peaks in raw EEG signals, so as to determine the classification accuracy. the feature extraction procedure is based on averaged downsampling signal epochs, ICA and single trial PCA technique. for classification we used linear support vector machine. Results show that besides single trial is more faster but P300 peak is more clear while using ICA combined with downsampling. Applying this method to the IIB of BCI Competition 2003,we could achieve an accuracy of 100% in P300 detection.

Key words: Brain-computer interface, ICA, principal components, P300, single trial

I. Introduction

A Brain Computer Interface or Brain Machine Interface is a direct Communication path between a human brain and an external world. it allows users to send their messages or commands without any muscle movement but thinking. The most fields of BCI is toward neuroprosthetics applications in order to restore damaged hearing, sight and movement. scientists use electroencephalography(EEG)to record human scalp potentials. There are different types of brain activity that can be used in a BCI system. These activities are visual evoked potential (VEP)[1,2], slow cortical potential(SCP)[3,5,8], and brain rhythms such as alpha, beta, mu[4]. Here just we focus on P300, an event related potential (ERP) that elicited by infrequent, task relevant stimuli. It is a positive peak occurring approximately 300ms after stimulus onset. History of P300 goes as follows: in 1988 Donchin and Farwell first introduced some P300 detection methods like stepwise discriminant analysis (SWDA), peak picking area and covariance [6], also Donchin used discrete wavelet transform (DWT) in addition to the SWDA [7]. Detecting P300 in single trial is difficult as it has very little amplitude in comparison with raw EEG signals. Then ensemble averaging can make this problem but it is at the expense of longer transfer time.

Therefore there should be a trade-off between transfer time and accuracy. This article compares different methods to test their accuracy:1)averaged downsampled signal, 2)ICA decomposition, 3)single trial PCA method. The classifier used is a linear support vector machine.

II. THE DATA

In this study we used the data acquired from the dataset IIb available in BCI Competition 2003 website(http://bci2000.org). The paradigm is a 6x6 matrix (see figure.1) of alphanumeric
characters [8]. The user’s task is to focus attention on target character when corresponding row and column are intensified and count the number of flashing of the desired character silently. This was used only to keep subjects attention.

Each row and column flashes randomly and successively for 100ms and at the rate of 5.7HZ. It means the time delay between every flash is 75ms with the frequency sampling of 240HZ. As we have 6 rows and 6 columns, it has to be 12 intensifications in every trial. 2 out of these 12 are target that are related to the desired character. There were 15 trials for every character that ends to a total of 180 intensifications for a single character. After finishing 180 intensifications, there was a 2.5s time interval so that the user can prepare for the next character. The EEG recording was performed based on 10-20 system with sixty-four electrode channels (figure 2).

III.METHODS

Here we used three types of features to extract P300 signal. for the first we used 10 channels from locations FZ,CZ,PZ,OZ,C3,C4,P3,P4,PO7,PO8. These channels were found to have larger amplitude responses in the P300 latency window [8], then raw signals were bandpass filtered by a 8th order chebyshev type I with cut off frequency between 2-8HZ. It has been proved through spectrum analysis that principle energy of P300 is centered in this frequency band[10]. The next step was ensemble averaging. We present a new averaging method; instead of using data with the same stimulus code just target and non-target responses were used for ensemble averaging. This will reduce the time of procedure. Then we chose a window of 667ms after stimulus onset. As we mentioned earlier the sampling frequency was 240HZ, then every epoch contained 160 data points. In order to reduce these points data were downsampled based on filter high cut off frequency, now we have 14 samples for every channel that selected as features for classification. Figure 3 shows a signal from channel Cz extracted by this method.

![Figure 1: Matrix of Spelling Paradigm](image1.png)

![Figure 2: position of electrodes based on 10-20 system](image2.png)

![Figure 3: Signal from channel Cz extracted by this method](image3.png)
The second was using PCA on single trial EEG signal. Principal component analysis has been called one of the most valuable results from applied linear algebra [11, 12]. It is used abundantly in all forms of analysis-from neuroscience to computer graphics-because it is a simple, non-parametric method of extracting relevant information from confusing datasets. This linear transform tries to uncorrelated the input data in a way that the new data variances are equal to the eigenvalues of the covariance matrix. It can also provide a roadmap for how to reduce a complex dataset to a lower dimension.

As in previous part we have data from 10 channels and length of 667ms after the stimulus onset. To apply PCA almost always data have to be centered so that the mean is equal to zero. If we consider that C is the covariance matrix of X (160×10), C would be a matrix of size 10×10 that can be computed using

\[ C = E(X^T X) \]  

(1)

Now the goal of PCA is to extract the most important information from the data. To this end after calculating covariance matrix the next step is to find eigenvalues of C, called F and D diagonal matrix of its eigenvalues. As C was a 10×10 matrix we had 10 eigenvalues that sorted from maximum to minimum. The principal component could be obtained by

\[ PC = F^T X^T \]  

(2)

The problem here is to reconstruct PCs that contain only ERP i.e., only the important information needs to be extracted from the covariance matrix. One guideline is to plot the eigenvalues according to their size so called scree plot. The answer is to see if there is a point in this graph that the slope of the graph falls from “steep” to “flat” and to keep only the components before the elbow. This plot is shown in figure 4.

Another standard method in PCA is to maintain the components whose eigenvalues is larger than the eigenvalue[13].

\[ \lambda_i > \frac{\sum_{i=1}^{L} \lambda_i}{L} \]  

(3)

Where L is the rank of X.

Then these selected PCs were used to reconstruct the signal that contains only ERP using this formula:

\[ Y = SP^T V^T \]  

(4)

Which sp denotes selected PCs and V is corresponding eigenvectors.

Another approach that has been used in this paper was combining three methods, including PCA, ICA and downsampling. In this work the point was decompose signals into ICs with the help of ICA and then select the ICs based on prior knowledge of P300 response (i.e., P300 response has large amplitude between 290-360ms after beginning the stimulus).

ICA is a general statistical technique related to the method called blind source separation [14]. It has been widely used in biomedical signal and image processing like EEG [15, 16], ECG [17], EMG [18, 19] and so on. The general concept of ICA is as follows:

Suppose that we have n sources \( s_i \) (i=1,2,…,n) that they are mutually independent. Now let us denote X, the mixture of these sources that can be write like X=AS where A is mixing matrix that is unknown. The problem is to find the inverse matrix of A, \( A^{-1} = W \) named demixing matrix so that we can estimate independent components \( u_i \) (i=1,2,…,n); that are latent variables; from their mixtures.

\[ u = W^T X \]  

(5)

Of course there are many different contrast functions for measuring the independence of u, and as a result there are various ways to find W. The algorithm we used here was an efficient method of maximizing the contrast function called FastICA. The FastICA learning rule finds a direction, i.e., a unit vector W such
that the projection $W^T X$ maximizes non-gaussianity. Non-gaussianity here was measured by the approximation on negentropy[20]. In this method the entire procedure was as following:

a) At first like two previous methods we mentioned before, data were bandpass filtered and then averaged.

b) In this step all 64 channels were considered. PCA was exploited for data to reduce the dimension. As it is proved numbers of ICs in VEP signals is around 15-35[21], we reduced the dimension from 64 to 22 based on the eigenvalues size [10], so the eigenvector matrix got $22 \times 64$, and reconstructed data was like

$$\hat{X} = V^T X^T$$ (6)

c) After PCA, FastICA was used to the dimensionally reduced data matrix. Because $V$ contained only 22 rows, A $22 \times 22$ demixing matrix $W$ was obtained after ICA training. We multiplied this matrix by the mixed data and received 22 ICs. As we know the features of ICs are clearer in time domain between 250 to 350 ms we kept just ICs that had larger amplitude in this window.

d) The final step was to downsample the preprocessed data from c and applies these data to the classifier, now the P300 peak was easier to detect.

![Figure 3: averaged wave on Cz using downsampling](image)

![Figure 5: signal extracted by single trial PCA on Cz](image)

![Figure 4: scree plot of eigenvalues](image)

![Figure 6: averaged wave on Cz](image)

Using ICA, the dashed line shows target signal
IV. DISCUSSION and CONCLUSION

In this work three methods were discussed for extracting P30 wave. In first and third, signals were averaged because response averaging removes background EEG activity (considered to be noise). Ensemble averaging we used here was different with the standard averaging method. Instead of averaging trials with the same stimulus code we just did the procedure on target and non-target trials. This task was helpful for time reduction. We also used in second method a single trial in order to reduce time and computationally cost. Another one we did was investigating PCA efficiency in 2 ways. First we tested some PC selection methods to separate VERP signals from background EEG and second we Used PCA just as a method to reduce the data dimension. Figure shows the P300 signal that has been extracted by the third method. We can see in this figure that the target signal peak has happened 310-320ms after stimulus onset that corresponds to the physiological feature of P300. By comparing figure.5 and figure.6 we find that the ICA was more robust to detect the signal of interest. It is because FastICA algorithm was used here tries to make the Components independent while PCA makes them uncorrelated. Another reason is averaging that didn't use in the second method. Of course the time of these three algorithm is comparable, then a trade-off needs to be happen but as this work is an off-line one, the accuracy is more important. Signals were applied to a linear support vector machine to classify and word prediction. Table 1 shows the results. Applying these proposed algorithms to P300 speller data of BCI competition 2003, we achieved accuracy of 90%, 98.5%, and 100% on the test data. Due to the good accuracy we can claim to be one of the winners in the competition. Without missing the accuracy we reduced the numbers of trial from 15 to 10 and 7 in averaging to increase the transfer time.

REFERENCES


<table>
<thead>
<tr>
<th></th>
<th>ACCURACY OF P300 DETECTION</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FIRST METHOD</td>
<td>SECOND METHOD</td>
<td>THIRD METHOD</td>
</tr>
<tr>
<td>TESTING DATA</td>
<td>90%</td>
<td>98.5%</td>
<td>100%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>ACCURACY OF WORD PREDICTION</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FIRST METHOD</td>
<td>SECOND METHOD</td>
<td>THIRD METHOD</td>
</tr>
<tr>
<td>TESTING DATA</td>
<td>81%</td>
<td>94%</td>
<td>97.5%</td>
</tr>
</tbody>
</table>

Table 1. Accuracy of P300 detection and Word prediction