Biocybernetics and Biomedical Engineering
2013, Volume 33, Number 1, pp. 3–20

Comparison of Performance of Different Feature Extraction Methods in Detection of P300

ZAHRA AMINI*, VAHID ABOOTALEBI, MOHAMMAD T. SADEGHI

Electrical and Computer Engineering Department, Yazd University, Safaeie, Yazd, Iran

The aim of this paper is to design a pattern recognition based system to detect the P300 component in the EEG trials. This system has two main blocks, feature extraction and classification. In the feature extraction block, in addition to morphological features, some new features including intelligent segmentation, common spatial pattern (CSP) and combined features (CSP + Segmentation) have also been used. Two criteria were used for the feature evaluation. Firstly, a t-test has been applied. Secondly, each of these four groups of features was evaluated by a Linear Discriminant Analysis (LDA) classifier. Afterwards, the best set of features was selected by using Stepwise Linear Discriminant Analysis (SWLDA). In the classification phase, the LDA was used as a linear classifier. The algorithm described here was tested with dataset II from the BCI competition 2005. In this research, the best result for the P300 detection was 97.4% .This result has proven to be more accurate than the results of previous works carried out in this filed.

K e y w o r d s: P300, brain computer interface (BCI), pattern recognition, feature extraction, classification

1. Introduction

Brain computer interface (BCI) is a system that creates a direct channel between computer and the brain. Among various BCI systems, electroencephalography (EEG) is still the most common method because of its non-invasive nature. By analyzing the electroencephalographic activities recorded from the scalp, a computer can recognize the brain's intention and translate it to commands for output devices such as a computer application or a neuroprothesis.

---

* Correspondence to: Zahra Amini, Electrical and Computer Engineering Department, Yazd University, Safaeie, Yazd, Iran, e-mail: zamini@stu.yazduni.ac.ir
Received 04 April 2012; accepted 11 September 2012

Infrequent or particularly significant auditory, visual, or somatosensory stimuli, when interspersed with frequent or routine stimuli, typically evoke a positive peak at about 300 ms in the EEG over parietal cortex. Donchin and his colleagues have used this ‘P300’ or ‘oddball’ response in a BCI [1]. Inbar used the P300 component for communication between the brain and computer. The BCI developed in his work was based on the BCI described by Farwell and Donchin in 1988, which allows a subject to communicate one of 36 symbols presented on a 6 × 6 matrix [2].

A P300-based BCI has an apparent advantage in that it requires no initial user training: the P300 is a typical or naive response to a desired choice.

The P300 speller paradigm is a kind of BCI which uses the P300 potential to spell the intended character of the user. The BCI Competition is a competition that is held every two years since 2000 and the P300 speller system is one of its parts. The P300 speller paradigm used to produce the dataset IIb of the BCI Competition 2003 and the dataset II of the BCI Competition 2005 is basically the same as that of Farwell and Donchin’s [3].

Since that time, many research groups and investigators in this line of research have worked on the P300 Speller system. For example Sellers’ group have carried out several studies and examined the effect of some cases such as the P300-Speller matrix size, expanding the classical P300 feature space, the performance of different linear and nonlinear classifiers on the BCI system accuracy [4, 5]. In their latest research, this group has tried to make changes in the stimulus paradigm to reduce the probability of error in the P300 detection [6].

Seyyedsalehi in 2008 [7] used a feature set as inputs into committee machines (CM) based on the LDA, MLP and SVM. This algorithm achieved an accuracy of 94% in the P300 detection.

Rakotomamonjy and Guigue proposed a method that detects the P300 through an ensemble of classifiers approach. Each classifier is composed of a linear support vector machine trained on a small part of the available data for which a channel selection procedure has been performed. Performance of their algorithm has been evaluated on the dataset II of the BCI Competition III and has yielded the best performance (96.5% accuracy) in the competition [8].

In 2009 Salvaris [9] tried to introduce a novel classification method included discrete-wavelet transform (DWT) preprocessing and an ensemble of Fisher’s Linear Discriminants for classification. The performance of the proposed method was slightly worse than the state of the art method for the BCI competition III data sets. But the proposed method was far less computationally expensive than the current state of the art method. The best accuracy that he could achieve was 95% on the dataset II of the BCI Competition III.

In this paper a pattern recognition system depicted in Fig.1 is used for detection of the P300 component. In this study, emphasis is on the feature extraction block. Thus, taking into account priori Neurophysiologic knowledge and different available processing methods, we decide to extract features through some new and suitable methods that have been used far less in this area. (Like intelligent signal segmentation and common spatial pattern (CSP)). In addition, morphological features – being common features in this field – have been used.

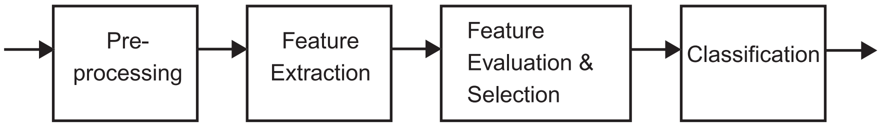

Fig. 1. Block diagram of the processing system

The outline of the rest of the paper is as follows. In Section II, the data set and the methods used in each block of Fig.1 are explained. The definition of the used features, the method of feature selection and the classification algorithm will be described in this section. Section III presents the performance and discussion about the results. Finally, Section IV concludes the paper.

2. Materials and Methods

In this section the proposed P300 detection system and its different parts (according to Fig. 1) are introduced.

2.1. Data Acquisition

The speller system studied in this article is based on the P300 speller paradigm that has been provided by Wadsworth center for the BCI Competition 2005 [3].

Farwell and Donchin developed a protocol whereby a subject is presented in a 6 × 6 character matrix as illustrated in Fig. 2.

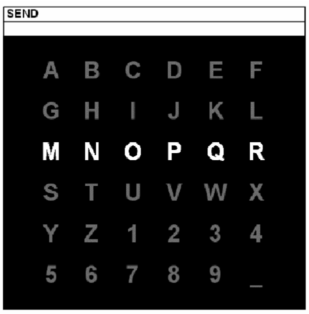

Fig. 2. User display for Farwell and Donchin Paradigm [3]

For the spelling of a single character, each of the 12 rows and columns of the matrix is then intensified according to a random sequence. All rows and columns of this matrix were successively and randomly intensified at a rate of 5.7 Hz. That is, each row and column in the matrix was randomly intensified for 100ms and after intensification of a row/column, the matrix was blank for 75ms. The subject is asked to focus its attention on the character he wants to spell and then it is expected that a P300 evoked potential appears in the EEG in response to the intensification of a row or column containing the desired character. In order to make the spelling procedure more reliable, this sequence of intensifications is repeated 15 times for each character to spell.

In the BCI Competition 2005, data has been recorded from two subjects and 5 different spelling sessions. Each session is composed of runs, and for each run, a subject is asked to spell a word. For a given acquisition session, all EEG signals of a 64-channel scalp have been continuously collected. The locations of the channels are defined based on the 10–20 standard. Before digitization at the sample rate of 240 Hz, the signals have been band pass filtered from 0.1–60 Hz.

For the competition data set the recorded data includes 85 characters. So there are 85*10 = 850 trials without the P300 and 85*2 = 170 trials with the P300 and totally we have 1020 data for our study.

2.2. Preprocessing

In this step, in order to eliminate high frequency and low frequency noise, the signal is passed through a high pass elliptic filter with 3 dB cut off frequency of 0.5 Hz and a low pass elliptic filter with 3 dB cut off frequency of 35 Hz. Then all the filtered data are normalized in the interval of [–1, 1] and finally for each channel the continuous signal is divided into epochs. Each epoch starts at the time of stimulation (the flashing of a row or column) and lasts for 1000 ms after it.

Another important issue is choosing appropriate channels. Because of the large number of channels used for data recording and to avoid complicated calculations, this selection is necessary. Therefore in this paper at all methods, except for the CSP method, we use seven more suitable channels including Po7, Po8, Fz, C3, Cz, C4, Pz. These channels were introduced as appropriate channels by the second place winner of the BCI Competition.

Finally, the data were averaged over 15 repeated intensifications for each character to spell. This increases the performance of our system significantly.

2.3. Feature Extraction

After the preprocessing, suitable features are extracted from the raw signal. The goal of feature extraction is to remove noise and other unnecessary information from the input signals, while at the same time retaining the information that is important in discriminating of different classes of the signals. A priori neurophysiologic knowledge can aid to decide which brain signal features are to be expected to hold the most discriminative information for the chosen paradigm.

In this study, four different groups of features are proposed and extracted that will be explained in this part.

2.3.1. Intelligent Segmentation

In recent years, there has been an explosion of interest in mining time series databases. As with most computer science problems, data representation is the key to the efficient and effective solutions. Several high level representations of time series have been proposed, including Fourier Transforms, Wavelets, Symbolic Mappings and Piecewise Linear Representation (PLR) [10].

In all cases, describing the data in a compact mathematical way is useful towards at least three goals: feature extraction (shape descriptors), data compression, and noise filtering [11].

The feature vector for the P300 detection, in some studies is constructed by concatenating the EEG time series data acquired from each electrode. Constructing the feature vectors with all sampled data points might result in a very high-dimensional vector, which might make suffer from the curse of dimensionality.

To avoid these problems, many investigators have applied downsampling to reduce the dimension of the feature vectors. Typically, two approaches have been primarily used, a decimator approach [8], and a downsampling approach with uniform interval segments [12].

But our signal segmentation refers to the approximation of a time series T, of length n, with K straight lines. Because K is typically much smaller than n, this representation makes the storage, transmission and computation of the data more efficient [10].This process can be used as a stage of feature extraction in the pattern recognition systems.

In the current study, we propose a new intelligent segmentation method to improve accuracy of the P300-based BCI classification. With this approach, we perform segmentation at non-uniform intervals as illustrated in Fig. 3 and take the average in each segment. The idea behind this strategy is based on our assumption that effective down sampling for the P300-based BCI should provide clear distinctions between the two classes (target and non target), and at the same time the degree of separation between the two classes should become larger.

To assign the samples into the classes, we need a measure of similarity or dissimilarity. In this research, we employ Fisher Discriminant Criterion (FDC) for this purpose. In the FDC, we provide a mapping vector that maximizes the ratio of the squared average difference between classes to the sum of between class variances. We first define the average vectors, between-class covariance and within-class covariance.

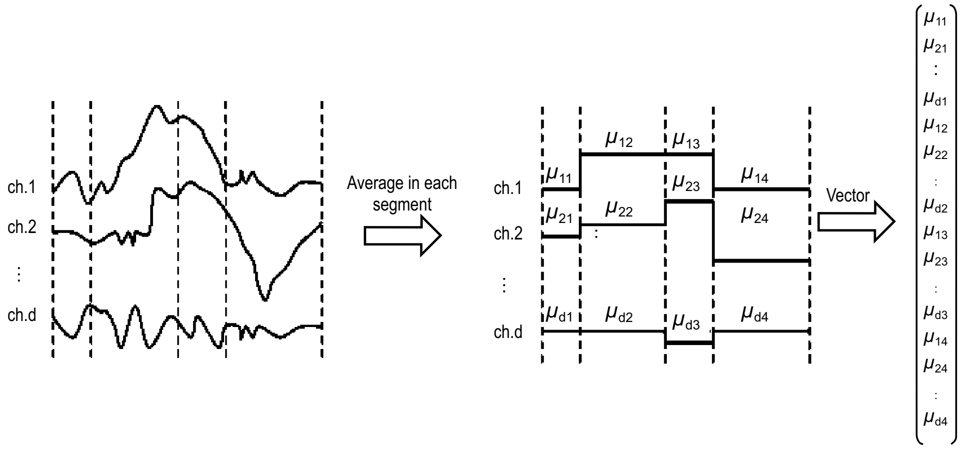

Fig. 3. Intelligent Segmentation and feature vector generation [13]

$$m_0 = \frac{1}{N_0} \sum_{i:y_i=1} x_i, \qquad m_1 = \frac{1}{N_1} \sum_{i:y_i=1} x_i, \tag{1}$$

$$W = \sum_{l=0}^{1} \sum_{i:y_i=l} (x_i - m_l)(x_i - m_l)^T, \tag{2}$$

$$B = (m_0 - m_1)(m_0 - m_1)^T,$$

where N_0 and N_1 represent the number of training data and $y_i \in \{0,1\}$ is a label for each class.

Then, the Fisher Discriminant Analysis (FDA) mapping vector is given by a vector that maximizes the variable denoted by "**a**" in the following expression:

$$\mathrm{J}(\mathbf{a}) = \frac{a^T Ba}{a^T Wa}. \tag{3}$$

The $J(\mathbf{a})$ in this expression is an FDC. Based on derivatives of equation (3) and after performing a series of mathematical operations, we obtain Fisher Discriminant Criterion that we will use in this research:

$$J = (m_0 - m_1)^T W^{-1}(m_0 - m_1). \tag{4}$$

Now we formulate the problem of obtaining the "optimal segment." Let $E = \{(S_i, y_i) / i = 1, \ldots, N\}$ be training samples, where $y_i \in \{0,1\}$ is a label and $\mathbf{S}_i$ represents the EEG data. Thus, if a single trial EEG consists of n data points, $\mathbf{S}_i$ is defined as $S_i = \{s_i(1), s_i(2), \ldots, s_i(n)\}$, where $s_i(t) \in R^d$, and d is the number of channels. Assume that the number of segments is denoted by k, the breakpoints between segments can be defined as $T = (\tau_0, \tau_1, \ldots, \tau_k)$ where, $\tau_j \in N, \tau_j < \tau_{j+1}, \tau_0 = 0$ and $\tau_k = n$.

We define a *Make Vector* ($\mathbf{S}_i$; T) function that down samples an $\mathbf{S}_i$ and converts it into a vector:

$$X_i(T) = Make\ Vector\ (S_i, T) = \begin{pmatrix} \frac{1}{\tau_1 - \tau_0} \sum_{t=\tau_0+1}^{\tau_1} s_i(t) \\ \frac{1}{\tau_2 - \tau_1} \sum_{t=\tau_1+1}^{\tau_2} s_i(t) \\ \vdots \\ \frac{1}{\tau_k - \tau_{k-1}} \sum_{t=\tau_{k-1}+1}^{\tau_k} s_i(t) \end{pmatrix}^t . \tag{5}$$

X_i obtained by this function is a $d \times k$ dimensional vector. Consider the training data set $\tilde{E}(T) = \{(X_i\,(T), y_i) / i = 1, \ldots, N\}$, which is obtained by the transformation defined by equation (5).

Since $\boldsymbol{m}_0$, $\boldsymbol{m}_1$, and $\boldsymbol{W}$ which we defined previously are generated by $\tilde{E}(T)$, they are all functions of T. The degree of separation between two classes based on the feature vectors generated by T is rephrased by equation (4) as follows:

$$J(T) = (m_0(T) - m_1(T))^t\, W(T)^{-1}(m_0(T) - m_1(T)). \tag{6}$$

Finally, the problem of finding "segments" that maximize the degree of separation between classes is reduced to finding $\widehat{T}$, which is given by the following formula [13]:

$$\widehat{T} = arg\ max_T J(T) = arg\ max_{\tau_0, \tau_1, \ldots, \tau_k} J(T). \tag{7}$$

Subject to $\tau_j \in N$, $\tau_j < \tau_{j+1}$, $\tau_0 = 0$, $\tau_k = n$.

To solve equation (7), we seek the optimal $\widehat{T}$ by modifying Local Iterative Replacement (LIR) into our time series data segmentation. LIR is a simple greedy procedure where the new place for a breakpoint is selected optimally between the neighboring two break points [14].

In LIR, we first provide the initial value for T, and then seek the optimal solution. LIR consists of repeated computation of the boundary selection between segments, and the search of a new boundary location before or after the selected boundary, until we obtain a solution that satisfies equation (7), which is assured to be a local optimum [13].

The goal is maximizing the FDC (J), so this algorithm will be stopped when the amount of FDC cannot be increased by any admissible move of a breakpoint.

After the convergence of the algorithm, the feature vector is obtained. In this vector the number of features for each channel is equal to the number of segments (k) and the average of each segment is taken as a feature (according to Fig. 3). For each character, the feature vector is obtained by concatenating the features from seven chosen channels. In other words, the length of this vector is $7k$.

Choosing the value of k is important because a small k may produce an inaccurate approximation of the signal and important parts of the data may be lost. On the other hand, a large k cannot reduce the volume of data and cannot avoid the curse of dimensionality. Therefore, we examined some values of k and finally chose $k = 10$ as the optimum value providing the best accuracy.

2.3.2. CSP

The CSP is a technique of spatial filtering that finds the directions of optimal discrimination between two classes through variance. It maximizes the variance of one condition and at the same time minimizes the variance of the other condition [15].

This method has been originally used for detection of the abnormal EEG signals [16], and was successfully applied to classification of the movement related EEG signals [17].

Within the P300 oddball principle context, we consider two spatio-temporal matrices **Xt** and **Xnt** with dimension N × L, where N is the number of channels and L is the number of samples of the time series epoch of each channel. The matrix **Xt** represents the P300 potential evoked by the target event and **Xnt** represents the ongoing EEG for non target events. The CSP method is based on the principal component decomposition of the averaged covariance matrix, $\overline{R}$ [18]. $\overline{R}$ is obtained by taking the sum of the target and non target covariances $(\overline{R}_t)$ and $(\overline{R}_{nt})$. The averaged covariance matrix $\overline{R}$ is factorized through the application of the PCA as follows:

$$\overline{R} = \overline{R}_t + \overline{R}_{nt} = \mathrm{A}\lambda\mathrm{A}', \tag{8}$$

where A is the orthogonal matrix of eigenvectors of $\overline{R}$ and λ is the diagonal matrix of eigenvalues of $\overline{R}$. A whitening transformation matrix **W,** transforms the covariance matrix $\overline{R}$ to I (identity matrix)

$$\mathrm{W} = \lambda^{\frac{1}{2}}\mathrm{A}', \tag{9}$$

$$S = W\overline{R}W' = I. \tag{10}$$

Applying the whitening transform to each individual class, we obtain

$$\begin{aligned} S_t &= W\overline{R}_t W'. \\ S_{nt} &= W\overline{R}_{nt} W'. \end{aligned} \tag{11}$$

From the above two equations it is straightforward that

$$S_t + S_{nt} = I. \tag{12}$$

Performing a PCA factorization on (11) we have

$$S_t = A_t \lambda_t A'_t \qquad S_{nt} = A_{nt} \lambda_{nt} A'_{nt}. \tag{13}$$

From (12) and (13) then

$$A_t = A_{nt}. \tag{14}$$

$$\lambda_t = I - \lambda_{nt}. \tag{15}$$

This means that both class patterns share the same eigenvectors and the respective eigenvalues are reversely ordered. The eigenvector with the largest eigenvalue for one class has the smallest eigenvalue for the other class and vice versa. The first and last eigenvectors are optimal eigenvectors to discriminate the two classes. Defining A_t and A_{nt} as the first and last eigenvectors with dimension $N \times 1$ the following spatial filters are designed:

$$H_{t\,=} A'_t W.$$

$$H_{nt\,=} A'_{nt} W. \tag{16}$$

The spatially filtered data is given by:

$$\mathrm{Y} = \mathrm{HX}, \tag{17}$$

where H is the matrix or vector with the selected filters.

In our study, the third set of features is obtained through the Common Spatial Pattern (CSP) technique. In this method, instead of using many channels in the original space we can obtain better separation between two classes by using fewer alternative channels in a new space. Therefore, in this method we use all 64 channels and after anti-aliasing filtering, we down sample the signal from 240 samples per second to 24 samples per second. Then, according to the pervious section, the eigenvalues and vectors are calculated (equation (14) and (15)) and the $N \times 4$ dimensional matrix H is obtained (By using the first two and the last two eigenvectors).

For feature extraction by the CSP method, according to previous research [15, 18], two sets of features are more common and we use these two sets. The first set is filtered projection features, signals that are filtered by the spatial filter, H, and are converted to $4 \times L$ dimensional signals (All parameters like H, N and L are as defined previously).

The second set, which is usually used for classification in motor imagery, is the ratio between the variance of one filtered projection and the sum of the variances of all filtered projections.

For building this set, according to equation (18) for each signal, four features are obtained.

$$f_k = \log\left[\frac{\operatorname{var}(H_k S)}{\sum_{i=1}^{4} \operatorname{var}(H_i S)}\right], \tag{18}$$

where S is an $L \times N$ dimensional signal and H_k is the $\mathbf{k^{th}}$ row of H in equation (17).

Thus the feature vector of this part is obtained by concatenation of these two feature sets.

2.3.3. Combined Features (CSP+ Segmentation)

This kind of feature is a combined feature group. In this group we first reduce the number of channels by going to a CSP's space and then decrease the volume of the data in this new space using intelligent segmentation and choosing better samples for classification. We also compare the result of this combined feature vector with the results of each of the groups CSP and intelligent segmentation alone.

2.3.4. Morphological Features

In this paper, morphological features are used in two parts, once for the raw signals and another time after the intelligent segmentation.

At first, 11 morphological features are chosen. Some of these features were previously used by Kalatzis et al. in discriminating of depressed patients from healthy controls using the P600 component of the ERP signal [19]. These features are defined and calculated as follows:

1) Latency $(LAT, t_{s_{\max}})$ – the ERP's latency time, i.e. the time where the maximum signal value appears:

$$t_{s_{\max}} = \{t \,/\, s(t) = s_{\max}\},$$

where $s(t)$ is the ERP single trial during 0–1000ms after stimulus and s_{max} is the maximum signal value in this time interval.

2) Amplitude $(AMP, s_{\max})$ – the maximum signal value:

$$s_{\max} = \max\{s(t)\}$$

3) Positive area (PAR, A_P) – the sum of the positive signal values:

$$A_P = \sum_{t=0ms}^{1000ms} 0.5\,(s(t) + |s(t)|).$$

4) Negative area (*ANR*, A_n) – the sum of the negative signal values:

$$A_n = \sum_{t=0ms}^{1000ms} 0.5(s(t) - |s(t)|).$$

5) Peak-to-peak (PP, pp):

$$pp = s_{\max} - s_{\min},$$

where $s_{\max}$ and $s_{\min}$ are the maximum and the minimum signal values, respectively:

$$s_{\max} = \max\{s(t)\} \qquad s_{\min} = \min\{s(t)\},$$

6) Peak-to-peak time window (*TPP*, t_p):

$$t_{pp} = t_{s_{\max}} - t_{s_{\min}}.$$

7) Peak-to-peak slope $(PPS, \dot{s}_{pp})$:

$$\dot{s}_{pp} = \frac{pp}{t_{pp}}.$$

8) Peak of N100 (P_{N100}) – the minimum signal value in [50, 180] time interval.

$$P_{N100} = \min\{s(t), 50 \le t \le 180\}.$$

9) Latency of N100 (t_{N100}) – the time where P_{N100} appears.

$$t_{N100} = \{t | s(t) = N100\}$$

10) *P3N4* – difference between the maximum signal value in [185, 500] time interval and the minimum signal value in [320, 500] time interval (corresponding to the P300 amplitude and the N400 amplitude respectively).

11) *N1P3* – difference between the maximum signal value in [185, 500] time interval and the minimum signal value in [50, 170] time interval (corresponding to the P300 amplitude and the N100 amplitude respectively).

Secondly, after the intelligent segmentation, we introduce and calculate four features for each segmented signal including:

1) *max* – Maximum signal value
2) *min* – Minimum signal value
3) *mean* – Average over signal values
4) *range* – Peak to peak

2.4. Feature Evaluation

Until now, four different feature groups are prepared. For evaluating these features we use two methods. First a statistical analysis is applied for evaluating the fitness of each feature in discriminating between target and non target signals. For this goal, all 315 extracted features of all feature extraction methods (70 intelligent segmentation+100 CSP+ 40 combined+ 105 morphological) are fed into the SPSS program. To determine the most relevant features differing between target and non target signals, the features are subjected to an independent sample t-test, with the target and non target as the grouping variable [20].

Then the ability of P300 detection for each feature group is looked into analyzed by an LDA classifier. Through these two methods, features can also be evaluated individually and as a group.

2.5. Feature Selection

In any classification task, there is a possibility that some of the extracted features might be redundant. These features can increase the cost and running time of the system, and decrease its generalization performance. In this way, the selection of the best discriminative features plays an important role when constructing the classifiers.

In this study, we employ a method using an SWLDA by the SPSS software to identify the best subset of features for classification. The SWLDA is a technique for selecting suitable features. In this method a combination of forward and backward stepwise regression is implemented. Starting with no initial model terms, the most statistically significant feature is added to the model. After each new entry is added to the model, a backward stepwise regression is performed to remove the least significant features. This process is repeated until the model includes a predetermined number of terms, or until no additional terms satisfy the entry/removal criteria. In this research we used the second stopping criteria for the SWLDA.

The SWLDA algorithm can be considered efficient because the terminating heuristic is implemented in such a way that suitable features are selected in a non-exhaustive manner. In a sense, the SWLDA has the advantage of having automatic feature extraction [12].

2.6. Classification

After feature extraction and selection, we need a classifier to distinguish the target and non-target signals from each other. In this study, features are subjected to the LDA. The aim of the LDA (also known as Fisher's LDA) is to use hyperplanes to separate the data representing the different classes. For a two-class problem, the class of a feature vector is determined by the side of the hyperplane that the vector is. This technique has very low computational requirements, which makes it suitable for many pattern recognition problems. Moreover this classifier is simple to use and generally provides good results.

Consequently, the LDA has been used with success in a great number of the ERP and EEG processing researches such as the motor imagery based Brain–Computer Interface (BCI), the P300 speller, the asynchronous BCI and the P300 detection [21, 22].

In this study, the performance of the LDA algorithm is estimated using a leave-one-out (LOO) method.

3. Results and Discussion

In this section, the numerical results are shown and explained.

3.1. Feature Evaluation

3.1.1. t-test Results

To determine the significance level of the difference of each feature's average in the two classes (target/non target), the t-test was applied to the data. According to the results of the test, 140 of 315 defined features for the subject 'A' and 143 of 315 defined features for the subject 'B' provide significant difference between the two classes (that is these features had p-value < 0.05). The results for the 20 most significant features for the subject 'A' and 'B' are presented in Table 1 and 2 respectively.

Table 1. Results of comparing between target and non-target group using statistical *t*-test (for Subject A)

Feature No. in Group	Feature Group	p_value	t_value
15	Segmentation	1.10E-38	–13.571
14	Segmentation	8.53E-30	–11.708
2	Morphological	1.62E-27	–11.195
45	Segmentation	3.17E-27	–11.128
37	Morphological	5.09E-22	–9.873
25	Segmentation	4.43E-21	–9.633
14	CSP+Seg.	5.3E-21	9.612
33	CSP	1.01E-20	9.541
35	Segmentation	3.36E-20	–9.404
36	CSP	2.49E-19	9.173
15	CSP+Seg.	6.59E-19	9.059
25	CSP+Seg.	1.28E-18	–8.980
35	CSP	4.80E-18	8.822
44	Segmentation	3.25E-17	–8.589
34	CSP	1.50E-16	8.398
60	CSP	2.09E-16	8.357
65	Morphological	1.40E-15	–8.114
5	Morphological	6.41E-15	–7.915
31	CSP	8.92E-13	7.239
5	Segmentation	1.33E-12	–7.181

These tables display the *t*-values and significance levels (*p*-values) measured for the difference between the targets' and the non targets' corresponding features. In these tables the first column shows the feature number in its group. For example in the first row of the table 1, the number "15" indicates that between 70 intelligent segmentation 15th feature is selected.

Table 2. Results of comparing between target and non-target group using statistical *t*-test (for Subject B)

Feature No. in Group	Feature Group	p_value	t_value
21	Morphological	3.33E-22	9.920
56	Morphological	1.38E-20	–9.905
14	Segmentation	6.41E-20	–9.330
13	Segmentation	1.52E-16	–8.397
91	Morphological	1.93E-15	8.073
44	Segmentation	2.15E-14	–7.754
37	Morphological	6.94E-14	–7.595
105	Morphological	7.59E-13	–7.262
38	Morphological	7.127E-12	–6.937
4	Segmentation	1.63E-11	–6.813
2	Morphological	1.87E-11	–6.792
23	Segmentation	7.20E-11	–6.587
24	Segmentation	1.04E-10	–6.528
3	Morphological	7.12E-10	–6.223
3	Segmentation	7.173E-10	–6.222
69	Segmentation	9.90E-10	–6.169
33	Segmentation	1.28E-09	–6.126
14	Morphological	2.59E-09	–6.009
34	Segmentation	3.41E-09	–5.963
36	Morphological	3.92E-09	–5.939

For a clear discussion and better comparison, in Fig. 4 the percent of each feature group among the features with *p*-value < 0.05 is shown for each of subjects, A and B.

According to Fig. 4, for both subjects, the most of appropriate features (regarding the t-test evaluation) belong to the morphological features. The Intelligent segmentation and the CSP features come next with little difference. And finally the combined features have the lowest share of the suitable features.

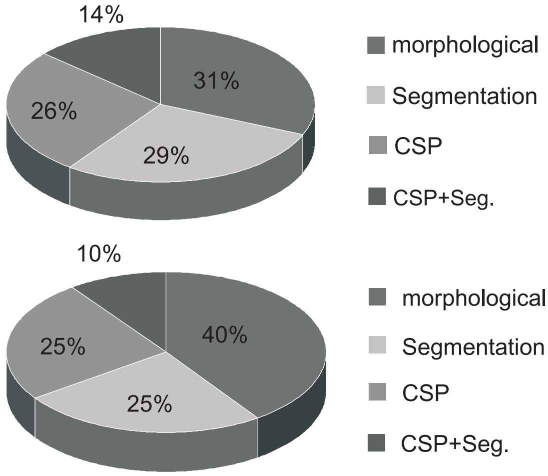

Fig. 4. The percent of each feature group among features with p-value < 0.05 for each subject; Up – subject A, Down – subject B

3.1.2. Comparison with a Common Classifier Results

Afterwards, each of the four feature groups was evaluated by a common classifier. We used an LDA as the common classifier. In Table 3 the results of this comparison between the feature groups for each subject are illustrated. According to this table, among these four feature vectors, the intelligent segmentation was seen to be most efficient in classification of these signals for both subjects.

Table 3. The results of P300 classification for each feature group by a common classifier, LDA

Accuracy / Feature Group	Subject A	Subject B
Intelligent Segmentation	94%	96.1%
Morphological	89.2%	87.3%
Combined (CSP+Seg.)	86.1%	84.3%
CSP	85%	79.9%

3.2. Feature Selection Results

After applying of the SWLDA, the best feature set was determined for each subject. This feature set included, for the subject ‘A’: 4 features from the morphological group, 7 features from the CSP group, 12 features from the combined group and 21 features from the intelligent segmentation group (a total of 44 features) and for subject ‘B’: 10 features from the morphological group, 12 features from the CSP group, 5 features from the combined group and 20 features from the intelligent segmentation group (47 features overall).

As explained before, the SWLDA method was used for selection of the best subset of features. Figure 5 shows the percent of each feature group among the features selected by this method for each subject.

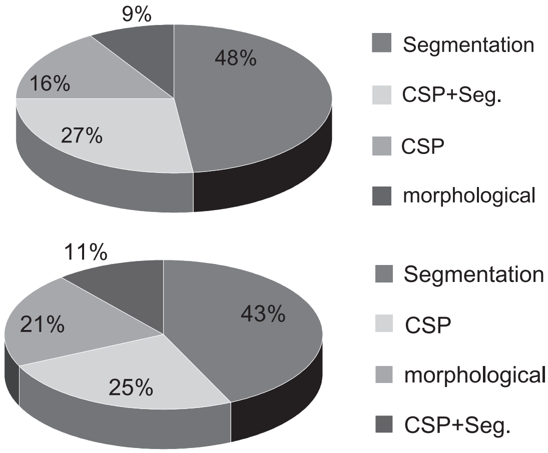

Fig. 5. The percent of each feature group among features selected by the SWLDA method for each subject; Up – subject A, Down – subject B

Figure 5 shows that in both subjects nearly half of all the selected features belong to the intelligent segmentation set and this result emphasizes the ability of this feature extraction method for our purpose (P300 detection).

3.3. Classification Results

The selected features (44 features for subject A and 47 features for subject B) were then used in the design of our classification system and in the analysis of the data as explained before. Table 4 displays the final result of our proposed system for the P300 detection. These results are obtained by using the selected features as an input for the LDA classifier.

Table 4. Results of classification for each subject

Accuracy / Subject	Target Accuracy	Non-target Accuracy	Total Accuracy
Subject A	96.50%	97.10%	97%
Subject B	95.30%	98.40%	97.80%
Mean	95.90%	97.75%	97.40%

4. Conclusion

The main purpose of this study was to evaluate the performance of a new classifying method in a P300-based BCI. For this purpose, we used a pattern recognition system including four main blocks: Preprocessing, feature extraction, feature evaluation and classification. Our emphasis was on the feature extraction block, so after preprocessing, a feature set consisting of four groups of features (including intelligent segmentation, CSP, combined and morphological features) were defined and extracted from the data. Then the features were compared with regard to the target and non target discrimination, by statistical analysis and the performance of a common classifier. After the feature evaluation, the optimal subset of the feature set was selected using a stepwise linear discriminant method. Then the selected features were used for the classification of data using the linear discriminant analysis.

By comparing our performance with the top of previous investigations, we can say that our proposed method by providing 97.4% mean accuracy for the P300 detection, outperforms other researches done in this line of research until now. For example the state of the art in the BCI competition 2005 by using ensemble of the SVMs yielded the 96.5% accuracy. Or Salvaris with wavelet and ensemble of the FLDs could obtain 95% accuracy on this dataset.

As demonstrated before the use of pattern recognition systems for the classification of the P300 has been very common. But not a very large number of these studies have used the features of this diversity that we used. In this paper we tried to extract some kind of features that have not been so useful in the P300 detection so far. So by a suitable selection of these features and a very simple classifier, we could achieve a high performance. The algorithm described here is efficient but we think there is still room for improvements. For instance, using the developed methods for the feature selection or the channel selection can lead to better results.

References

1. Farwell L.A., Donchin E.: Talking of top of your head: Toward a mental prosthesis utilizing event--related brain potentials. J. Electroensephalogr. Clin. Neurophy 1988, 70, 510–523.
2. Serby H., Yom-Tov E., Inbar G.F.: An improved P300-based brain-computer interface. IEEE Trans. on Neural Systems and Rehabilitation Eng. 2005, 13, 1, 89–98.
3. BCI Competition 2005. ida.first.fraunhofer.de/projects/bci/competition_ii
4. Sellers E.W., Krusienski D.J., McFarland D.J., Vaughan T.M., Wolpaw J.R.: A P300 event-related potential brain–computer interface (BCI): The effects of matrix size and inter stimulus interval on performance. J. Biological Psychology 2006, 73, 242–252.
5. Nijboer F., Sellers E.W., Mellinger J., Jordan M.A., Matuz T., Furdea A., Halder S., Mochty U., Krusienski D.J., Vaughan T.M., Wolpaw J.R., Birbaumer N., Kübler A.: A P300-based brain–computer interface for people with amyotrophic lateral sclerosis. J. Clin. Neurophysiol. 2008, 119, 1909–1916.

6. Townsend G., LaPallo B.K., Boulary C.B., Krusienski D.J., Frye G.E., Hauser C.K., Schwartz N.E., Vaughan T.M., Wolpaw J.R., Sellers E.W.: A novel P300-based brain–computer interface stimulus presentation paradigm: Moving beyond rows and columns. J. Clin. Neurophysiol. 2010, doi:10.1016/j.clinph.2010.
7. Seyyedsalehi Z., Nasrabadi A.M., Abootalebi V.: Committee Machines and Quadratic B-spline Wavelet for the P300 Speller Paradigm. IEEE/ACS Conf. on Comp. Sys. And Appl., 2008, 866–869.
8. Rakotomamonjy A., Guigue V.: BCI Competition III: Dataset II- Ensemble of SVMs for BCI P300 Speller. IEEE Trans. Biomed. Eng. 2008, March, 55, 3, 1147–1154.
9. Salvaris M., Sepulveda F.: Wavelets and ensemble of FLDs for P300 classification. 4th Int. IEEE/EMBS Conf. on Neural Eng., 2009. NER '09, April 29 2009-May 2 2009, 339–342.
10. Keogj E., Chu S., Hart D. and Pazzani M.: An Online Algorithm for Segmenting Time Seris. First IEEE Int. Conf. Data Mining 2001,289–296.
11. Pavlidis T., Horowitz S.L.: Segmentation of Plane Curves. IEEE Trans. Comp. 1974, C-23(8), 860–870.
12. Krusienski D.J., Sellers E.W., McFarland D.J., Vaughan T.M., Wolpaw J.R.: Toward enhanced P300 speller performance. J. Neuroscience Methods 2008, 167, 15–21.
13. Sakamoto Y., Aono M.: Supervised Adaptive Downsampling for P300-based Brain Computer Interface. 31st Int. Conf. on IEEE EMBS**,** September 2–6, 2009, 567–570.
14. Himberg J., Kalle K., Mannila H., Tikanmaki J.: Time series segmentation for context recognition in mobile devices. IEEE Int. Conf. Data Mining, December 2001, 203–210.
15. Combaz1 A., Manyakov N.V., Chumerin1 N., Suykens J.A.K., Van Hulle M.M.: Feature Extraction and Classification of EEG Signals for Rapid P300 Mind spelling. IEEE Int. Conf. Machine Learning Appl. 2009, 386–391.
16. Koles Z. J.: The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG. Electroenc. Clin. Neurophys. 1991, 79, 440–447.
17. Müller-Gerking J., Pfurtscheller G., Flyvbjerg H.: Designing optimal spatial filters for single-trial EEG classification in a movement task. Electroenc. Clin. Neurophys. 1999, 110, 5, 787–798.
18. Pires G., Nunes U., Castelo-Branco M.: P300 spatial filtering and coherence-based channel selection," 4th IEEE EMBS Conf. on Neural Eng., NER '09, April 29 2009–May 2 2009, 311–314.
19. Kalatzis I., Piliouras N., Ventouras E., Papageorgiou C.C., Rabavilas A.D., Cavouras D.: Design and implementation of an SVM-based computer classification system for discriminating depressive patients from healthy controls using the P600 component of ERP signals. J. Computer Methods and Programs in Biomedicine 2004, 75, 11–22.
20. Abootalebi V., Moradi M.H., Khalilzadeh M.A.: A new approach for EEG feature extraction in P300--based lie detection. J. Computer Methods and Programs in Biomedicine 2009, 94, 48–57.
21. Pfurtscheller G.: EEG event-related desynchronization (ERD) and event-related synchronization (ERS), in: E. Niedermeyer, F. Lopes da Silva (Eds.), Electroencephalography: Basic Principles, Clinical Applications and Related Fields, 4th edn, Williams and Wilkins, Baltimore, 2000, MD 958–967.
22. Bostanov V.: BCI competition 2003-data sets Ib and IIb: feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram. IEEE Trans. Biomed. Eng. June 2004, 51, 6, 1057–1061.