

Classification of EEG-P300 Signals Extracted from Brain Activities in BCI Systems using ν -SVM and BLDA Algorithms

Ali MOMENNEZHAD, Mousa SHAMSI, Hossein EBRAHIMNEZHAD*,
Hamidreza SABERKARI

Faculty of Electrical Engineering, Sahand University of Technology, Tabriz, Iran
Emails: a_mommenzhad@sut.ac.ir, shamsi@sut.ac.ir, ebrahimnezhad@sut.ac.ir*,
h_saberkari@sut.ac.ir

* Author to whom correspondence should be addressed; Tel.:+984113459343; Fax:+984113444322

Received: 18 April 2014/Accepted: 2 June 2014 / Published online: 25 June 2014

Abstract

In this paper, a linear predictive coding (LPC) model is used to improve classification accuracy, convergent speed to maximum accuracy, and maximum bitrates in brain computer interface (BCI) system based on extracting EEG-P300 signals. First, EEG signal is filtered in order to eliminate high frequency noise. Then, the parameters of filtered EEG signal are extracted using LPC model. Finally, the samples are reconstructed by LPC coefficients and two classifiers, a) Bayesian Linear discriminant analysis (BLDA), and b) the ν -support vector machine (ν -SVM) are applied in order to classify. The proposed algorithm performance is compared with fisher linear discriminant analysis (FLDA). Results show that the efficiency of our algorithm in improving classification accuracy and convergent speed to maximum accuracy are much better. As example at the proposed algorithms, respectively BLDA with LPC model and ν -SVM with LPC model with 8 electrode configuration for subject S1 the total classification accuracy is improved as 9.4% and 1.7%. And also, subject 7 at BLDA and ν -SVM with LPC model algorithms (LPC+BLDA and LPC+ ν -SVM) after block 11th converged to maximum accuracy but Fisher Linear Discriminant Analysis (FLDA) algorithm did not converge to maximum accuracy (with the same configuration). So, it can be used as a promising tool in designing BCI systems.

Keywords: EEG signal classification; BCI system; Visual Evoked Potential; P300; BLDA; ν -SVM.

Introduction

Brain Computer Interference (BCI) provides a way for brain to communicate with outside world. In a BCI system, human brain activities are transformed into computers usable commands. The goal of a BCI system is improving and expanding the systems for establishing communication with outside world for disabled people and controlling different organs [1]. BCI performance doesn't depend on healthiness of brain muscular output channels [2]. User intention is transferred by brain signals, independent of peripheral nerves and muscles. These signals are considered in a BCI system.

Nowadays, there are a lot of challenges in designing a BCI system in spite of developing technologies. This leads to be need a BCI system which can be used in out of laboratorial environment by disabled people [3]. Today, there are different techniques for recording brain signals, which Electroencephalogram (EEG) signal is especially important due to noninvasive property and easy implementation. This signal reflects electrical activities of large group of the

nervous signal in brain. These electrical activities are recorded in skull via many electrodes in special arrangements. Basic structure of a BCI system is shown in Figure 1 which includes five stages as follows [4]:

- System Input includes raw EEG information which is received from electrodes connected to brain,
- Preprocessing stage consists of filtering the input EEG signal in order to noise reduction and increasing the signal to noise ratio,
- Translation process includes two parts; extraction and classification the features. Feature extraction includes extraction of valuable signals from input and classifying them into useable outputs for the next stage,
- Feature classification includes identifying feature patterns for simplifying the user's commands clustering, and
- The classifier output is used for controlling the device. Device control process convert the classifier output into an action of device.

Many types of mental activities may be used in BCI system designing. These methods are totally classified into two main groups based on type of their production [5]: a) using of stimulus input such as visual evoked potential (VEP), b) use of cortical potentials not requiring external stimulation. In this paper, our goal is to design P300-based BCI system. P-300 wave is an event-related potential (ERP) [6]. One of the features of this signal is that personal training is not needed to record this signal and obtained by recording brain signals. This wave is corresponding to a positive reflection in voltage which appears 300ms after stimulation in EEG signal [1]. In other words this reflection appears approximately 300 milliseconds after stimulation of brain via a stimulator like a light bulb [7]. Its main application is for the disabled peoples who suffering from severe muscular disturbance [8] so that making possible for them to communicate with the outside world and control their different organs and regarding as a suitable rehabilitation tool for them [9]. The P-300 was firstly applied in words spelling systems which helped disabled people to spell words by it. It was done by selecting some terms on computer screen which contained alphabets or signs [10].

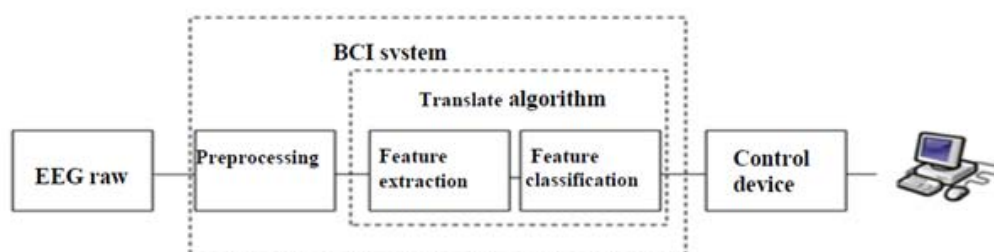


Figure 1. BCI system structure [1]

Many researches have been done on EEG signals classification. It's provided an algorithm for classifying EEG signals using wavelet transform and classification of SVM in [11]. Classification of these signals have been done in [12], using signal spectral analysis and classifiers of RNN recursive networks. AR model and nervous system classifiers are used for this purpose in [13]. LDA and FLDA methods in [14] and wavelet transform and nervous system in [15] are used for EEG signals classification.

Signal level obtained from the P-300 is much smaller than noise level. Therefore, it is necessary to use an optimal method in order to extracting and classifying components of P-300 from EEG signal. In this paper, first of all, an optimal model has been proposed for extracting main features of EEG signal and removing noise which is called Linear Predictive Model (LPC). Linear Predictive Model is one of the powerful tools in analysis of speech signals which is used for estimating main parameters of these signals [16]. As we mentioned speech signals are very similar to EEG signals in terms of their nature i.e. noise pollution and non-stationary features. Two classifiers of Bayesian

Linear Discriminator Analysis (BLDA) and ν -Support Vector Machine (SVM) with RBF kernel are used for better classifying of signals extracted from LPC model too.

The paper is organized as follows: The proposed algorithm is introduced in section 2 and its different stages are mentioned in detail. Criteria of evaluating function of the proposed algorithm including maximum bitrates, accuracy of classification and convergent speed to maximum accuracy are mentioned in Section 3. Results of implementation are presented in Section 4 and finally we will have conclusion in Section 5.

The Proposed Algorithm

The proposed algorithm for designing a BCI system is shown in Figure 2. The main steps that would be discussed in the following are as follows:

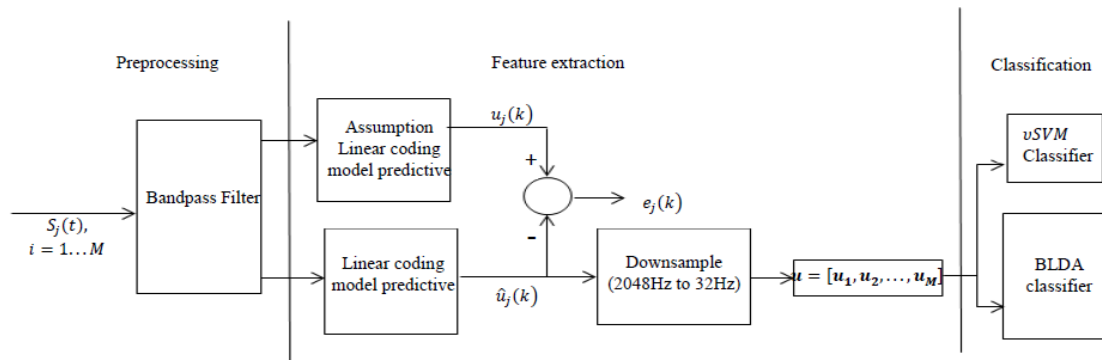


Figure 2. Block diagram of proposed algorithm

EEG Signal Preprocessing

Before extracting features of signal, a 6th order Butterworthband-pass filter with cut off frequencies of 1 Hz and 12Hz is used for removing signal noise [17].

Extracting Main Features by Applying Linear Predictive Model

The goal of extracting features is to find out brain signals related to the mental activities. Extracted signals should be noiseless as much as possible and have no other redundancy patterns. Because it leads to reduce the classification accuracy and also will be difficult for analysis the EEG signals. For this purpose, the linear predictive coding (LPC) model is applied on filtered input signals. LPC is one of the strongest techniques for non-stationary signals analyzing. The primary idea of this model is to estimate the signal based on a linear combination of the previous samples. The prediction coefficients are computed by minimizing the summation of the samples of main signals and the estimated samples errors.

Assume that $s_j(t)$, $j = 1 \dots M$ ($M =$ number of electrodes) is the filtered input signal at time t and $\hat{u}_j(k)$, $j = 1 \dots M$ is the estimated signal by applying LPC model in which k is the number of signal samples. $\hat{u}_j(k)$ is obtained by linear combination of p previous calculated samples [18]:

$$\hat{u}_j(k) = \sum_{i=1}^p a_j(i) u_j(k-i) \quad (1)$$

in which $\{a_j(i)\}$ is called linear estimated coefficient (in this paper we consider p as 6). Prediction error $e(k)$ between the observed signal $u_j(k)$ and the estimated signal $\hat{u}_j(k)$ is defined as follows [18]:

$$e(k) = u_j(k) - \hat{u}_j(k) = u_j(k) - \sum_{i=1}^p a_j(i) u_j(k-i) \tag{2}$$

Estimated coefficients $\{a_j(i)\}$ can be optimized by minimizing the summation of squared errors as Eq. (3):

$$E_j = \frac{1}{N} \sum_{k=0}^{N-1} e_j^2(k) = \frac{1}{N} \sum_{k=0}^{N-1} \left(u_j(k) - \sum_{i=1}^p a_j(i) u_j(k-i) \right)^2 \tag{3}$$

In order to solve Eq. (3), we can do as follow by differentiating from E to a_k and setting it equal to zero:

$$\frac{\partial E_j}{\partial a_j(i)} = 0, i = 1, \dots, p. \tag{4}$$

By solving Eq. (4) a p set of linear equations is obtained as follows:

$$\sum_{i=1}^p a_j(i) r_j(m-i) = r_j(m), m = 1, \dots, p \tag{5}$$

where $r_j(m)$ is the autocorrelation of $u_j(k)$:

$$r_j(m) = \sum_{k=1}^N u_j(k) u_j(k+m) \tag{6}$$

Matrix form of Eq. (5) is as follows:

$$\mathbf{R}\mathbf{a} = \mathbf{r} \tag{7}$$

where \mathbf{R} is the autocorrelation matrix with $p \times p$ dimension, \mathbf{r} is the autocorrelation vector with $p \times 1$ dimension and \mathbf{a} is a vector with $p \times 1$ dimension which contains the prediction coefficients. So, each parameters \mathbf{R} , \mathbf{r} and \mathbf{a} are defined as follows:

$$\mathbf{R} = \begin{bmatrix} r(0) & r(1) & r(2) & \dots & r(P-1) \\ r(1) & r(0) & r(1) & \dots & r(P-2) \\ r(2) & r(1) & r(0) & \dots & r(P-3) \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ r(P-1) & r(P-2) & r(P-3) & \dots & r(0) \end{bmatrix}; \mathbf{a} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \cdot \\ \cdot \\ \cdot \\ \alpha_p \end{bmatrix}; \mathbf{r} = \begin{bmatrix} r(1) \\ r(2) \\ r(3) \\ \cdot \\ \cdot \\ \cdot \\ r(P) \end{bmatrix} \tag{8}$$

The LPC model used in this paper is shown in Figure 3. The frequency rate of signal would change from 2048 Hz to 32 Hz after applying the LPC model. Figures 4 (a) and (b) show the EEG signal for six pictures before and after applying the LPC model, respectively.

Proposed Classification Algorithms

In this paper, Bayesian Linear Discriminant Analysis (BLDA) classifier and ν -support vector machine (ν -SVM) is used for classifying the main extracted features from LPC model that we attempt to express them in this part.

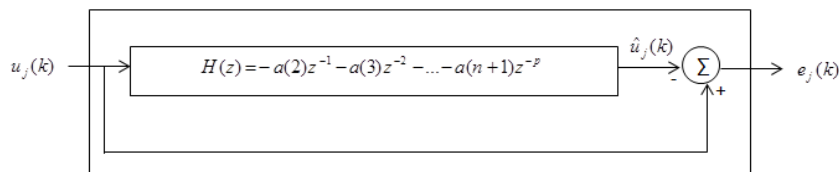


Figure 3. Block diagram of proposed algorithm

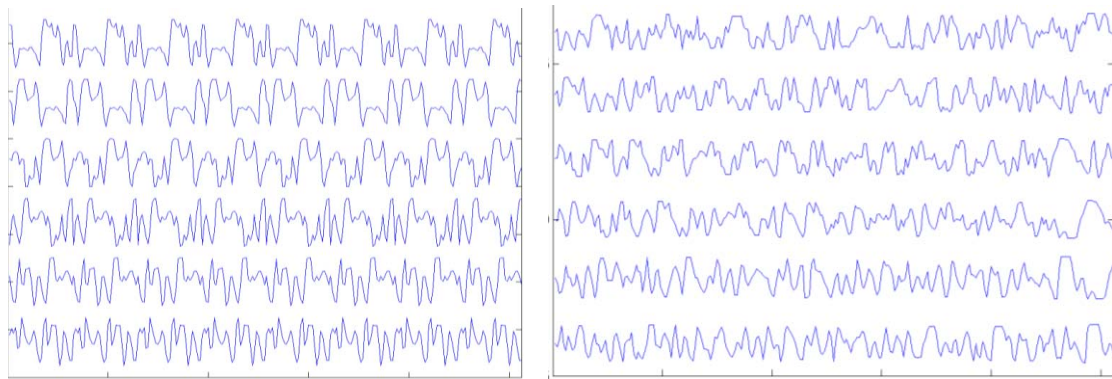


Figure 4. The signal models extracted from 6 pictures in one block before and after applying the LPC are shown in (a) and (b), respectively

Bayesian Linear Discriminant Analysis (BLDA). The BLDA algorithm is an adjustable algorithm, using to prevent over-fitting in high dimension data. By this algorithm, the adjustment degree could be estimated automatically and rapidly through training data and without using validation. This classifier is used for classifying the noisy data and the features which could not be classified correctly [14]. Main idea of this classifier is performing regression in Bayesian framework [3]. For this reason, targets and feature vector have linear relation. This relation is as follows:

$$t = w^T x + n \tag{9}$$

in which t is the target vector, x is the feature vector, w is the weight vector ($w \in R^D$) and n is the white noise. So, the Likelihood function is mentioned as follows for weights of w in regression:

$$p(D | \beta, w) = \left(\frac{\beta}{2\pi}\right)^{\frac{N}{2}} \exp\left(-\frac{\beta}{2} \|X^T w - t\|^2\right) \tag{10}$$

$X (X \in R^{D \times N})$ is a row-matrix containing feature vectors, D indicates two parameters of $\{X, t\}$, β is the reverse variance of the noise and N indicates the number of samples in training set. It is necessary to note that N_1 is the sample of the first class (C_1) and N_2 is the sample of the second class (C_2) by considering two classes of C_1 and C_2 in BLDA classifier ($N = N_1 + N_2$). The Label of C_1 and C_2 classes, changes as $\frac{N}{N_1}$ and $-\frac{N}{N_2}$, respectively [19].

In order to describe a Bayesian set, the prior distribution for weight vectors should be determined. This distribution obtains primary information about weight vectors and is defined as follows:

$$p(w | \alpha) = \left(\frac{\alpha_i}{2\pi}\right)^{\frac{D}{2}} \exp\left(-\frac{1}{2} w^T I'(\alpha) w\right) \tag{11}$$

in which α_i indicates the reverse of primary distribution variance for W_i weight vectors, I'_α is a square diagonal matrix with $D+1$ dimensions where D is the number of features. By using the prior distribution and Likelihood function and also applying Bayesian law, posterior distribution could be obtained as follows:

$$p(w | \beta, \alpha, D) = \frac{p(D | \beta, w) p(w | \alpha)}{\int p(D | \beta, w) p(w | \alpha) dw} \tag{12}$$

Since the prior distribution and Likelihood function are Gaussian, the posterior function would be Gaussian too, so that the average and covariance of this distribution calculate as follows:

$$m = \beta(\beta XX^T + I'(\alpha))^{-1} X t \tag{13}$$

$$C = (\beta XX^T + I'(\alpha))^{-1} \tag{14}$$

Therefore, the Posterior distribution can be used to distribute probability computation of regression targets, \hat{t} , for the new feature vector \hat{x} . The predictive distribution can be obtained by integrating respect to w :

$$p(\hat{t}|\beta, \alpha, \hat{x}, D) = \int p(\hat{t}|\beta, \hat{x}, w) p(w|\beta, \alpha, D) dw \tag{15}$$

The predictive distribution is again Gaussian and can be characterized by its mean μ and its variance σ^2 as follows:

$$\mu = w^T \hat{x} \tag{16}$$

$$\sigma^2 = \frac{1}{\beta} + \hat{x}^T C \hat{x} \tag{17}$$

Represented distributions in equations (16, 17) are Gaussian distributions with the average and variance of μ and σ^2 , respectively, that can be used for determining the different amounts of \hat{t} for new input vector \hat{x} . It should be noted that regression targets in *BLDA* algorithm are regulated for the first class samples in $\frac{N}{N_1}$ and for second class samples in $\frac{N}{N_2}$, in which N is the total training samples, N_1 is the number of first class samples, and N_2 is the number of second class samples.

To compute β and α , we need to write their Likelihood function. So we have:

$$p(D|\beta, w) = \int p(D|\beta, w) p(w|\alpha) dw \tag{18}$$

The quantity of $p(D|\beta, w)$ is the marginal likelihood and computes the probability of β and α . The integral in equation (18) can be solved by considering that everything is Gaussian. After solving the integral and maximizing that by partial derivation with respect to α and β and equate to zero, at last we have:

$$\alpha = \frac{D}{\sum_{i=1}^D c_{ii} + m_i^2} \tag{19}$$

$$\beta = \frac{N}{tr(XX^T C) + \|X^T m - t\|^2} \tag{20}$$

***v* Support Vector Machine (SVM).** *v* Support Vector Machine (*v*-SVM) is a common method for classification, prediction and regression. Its main idea is to use a divider super plate to maximize the distance between two classes in order to make the desired classifier (Figure 5). In a binary SVM, training data includes n ordered pairs of $(x_1, y_1), \dots, (x_n, y_n)$ such that [20]:

$$y_i \in \{-1, 1\} \quad i = 1, \dots, n \tag{21}$$

As a result, SVM standard formula is as follows [20]:

$$\min_{\omega, b, \zeta} \frac{1}{2} \omega^T \omega + C \sum_{i=1}^n \zeta_i \tag{22}$$

And also:

$$y_i (\omega^T f(x_i) + b) \geq 1 - \zeta_i, \quad \zeta_i \geq 0, \quad i = 1, \dots, n \tag{23}$$

$\omega \in R^m$ is a vector of training samples weights. B is a numerical constant; C is a fixed parameter with real value and finally ζ is a slack variable. If we have $\phi(x_i) = x_i$, then the relation (23) indicates a linear super plate with the maximum distance. Relation (23) is a nonlinear SVM if x_i could be mapped to a space with different dimensions of x_i space by ϕ , the common method uses the relation (24), [20]:

$$\min_a \frac{1}{2} a^T Q a - e^T a \tag{24}$$

where:

$$y^T a = 0, \quad 0 \leq a_i \leq C, \quad i = 1, \dots, n \tag{25}$$

e is a vector of ones, C is the upper bound, α_i is a Lagrange multiplying variable which its efficacy depends on C . Q is also a self-defined and positive matrix so that $Q_{ij} K(x_i, x_j) \equiv y_i y_j K(x_i, x_j)$ is a core function. It could be proven that if we got α as an optimal factor for relation (24), then $\omega = \sum_{i=1}^n \alpha_i y_i \phi(x_i)$ would be optimal too. The training vector of x_i is mapped into a space with different dimensions by function ϕ , so the decision function would be as follows [20]:

$$\text{sgn}(\omega^T f(x)+b) = \text{sgn}(\sum_{i=1}^n y_i \alpha_i K(x_i, x) + b) \quad (26)$$

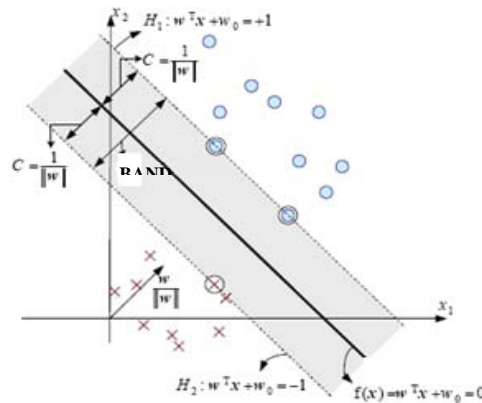


Figure 5. Hyperplane for two classes' problem [20]

The main problem of Support Vector Machine is the constant and uncontrollable nature of parameter C in relation (22). In order to solve this problem, ν Support Vector Machine (ν -SVM) has been used in this paper that we used LIBSVM for this. This algorithm was introduced by Scholkopf in 2000 [21]. In this algorithm, a pair of super plate with $\omega^T x + \omega_0 = \pm \rho$, $\rho \geq 0$ and a new parameter called $\nu \in (0, 1]$ has been used. By applying this algorithm, relation (22) would be corrected as follows [22]:

$$\min_{\omega, b, \zeta} \frac{1}{2} \omega^T \omega - \nu \rho + \frac{1}{l} \sum_{i=1}^l \zeta_i \quad (27)$$

And we have:

$$y_i (\omega^T f(x_i) + b) \geq \rho - \zeta_i, \zeta_i \geq 0, i = 1, \dots, n \quad (28)$$

In [23], it has been proved that ν is an upper bound on a part of training errors and a lower bound on a part of support vectors. To understand the role of ρ , we consider $\zeta_i = 0$. Then the constraint is $\frac{2\rho}{\|w\|}$. So, the margin can be controlled by ρ . To have a modified cost function we use

its Lagrange coefficients. If we suppose λ_i as Lagrange coefficients, at last we have[20]:

$$f(x) = \text{sgn}(\sum_{i=1}^N \lambda_i y_i k(x_i, x) + b) \quad (29)$$

$k(x_i, x)$ is depends on the type of kernel. In this paper, RBF function has been used($K(x_i, x_j) = \exp(-\frac{\|x_i - x_j\|^2}{\sigma^2})$). Also the value of σ^2 in 8-electrode configuration is equal to 250 and in 16-electrode configuration is 500. And we considered value of parameter ν as equal to 0.24 in this paper.

Criteria for Evaluating Function of the Proposed Algorithm

Parameters of evaluating the proposed algorithm performance in this paper include classification total accuracy, maximum bit rate and convergent speed to maximum accuracy. The convergent speed to maximum accuracy is the most important criteria because criterion of maximum bit rate generally depends on the initial accuracy of blocks while there is no guaranty for increasing and keeping the next blocks accuracy. Bitrate is defined as the number of the sent bits to BCI system in a defined time unit by each under experiment person and it is obtained from the following relation [23]:

$$b(N, p, t) = \left(\log_2(N) + p \log_2(p) + (1-p) \log_2 \left(\frac{1-p}{N-1} \right) \right) \frac{60}{t} \quad (30)$$

N is the number of different commands which are sent by user to the system. P is probability of correct command recognizing by the system and t is the required time to send a single command. Based on relation (30), as the classification accuracy increased, the sent bit rate would increase too, while this will be more effective on the primary blocks of the proposed algorithm.

Present your results in logical sequence in the text, tables, and illustrations, giving the main or most important findings first. Do not repeat in the text all the data in the tables or illustrations; emphasize or summarize only important observations.

Results and Discussion

The database using in this paper includes recorded EEG signals relating to 5 healthy people and 4 disabled is introduced in [19]. Subjects S1 and S2 were able to move their hands slowly and it was possible to communicate with them verbally. These people were suffering from speech disorders. Subject S3 was able to move his left hand but it was not possible to communicate with him verbally and was only able to communicate by telling yes or no. Subject S4 had low ability to control his hand movement while he has had ability of verbal communication. Subject S5 has no ability to control movement of his hand and it was too difficult to communicate with. Subject S6 to S9 had no problem in their physical condition. Everyone is tested in four stages, two of them are performed at one day and the next two stages at another day with less than two weeks' time interval. In this test, 6 images would be randomly shown to every person with time interval of 400 *ms* and they would be requested to count the number of times an image was seen. In the second stage of test and after observing 6 images by subjects, order of appearance of images would change. All 6 images in this test make a block and total number of blocks in all 6 tests is between 20 and 25. EEG signal is recorded by the electrodes connected to these people while they are seeing the images. Four configurations of electrodes have been used in this test, which include 4-electrode, 8-electrode, 16-electrode and 32-electrode configurations [8].

In this paper, obtained results validation is based on *k-fold* method in which k refers to the number of repetitions and it is considered to be 4. Hence, 4-fold validation is done so that the samples are classified into four parts at first and 75% of all data are regarded as training samples and the remaining is regarded as test samples in each time of algorithm applying. This trend would be done for four times by using training data and different tests and the conclusion would be taken by averaging the 4 test repetitions.

Classification accuracy and convergent speed to maximum accuracy, as the two main parameters have been studied. Classification accuracy is shown in 16 and 18-electrodes configurations in terms of time for all tested subjects in the proposed algorithms and FLDA algorithm in Figures 6 and 7. As observed, the proposed LPC-BLDA algorithm is much better than FLDA algorithm in terms of convergent speed to maximum accuracy, so that the algorithm reaches maximum accuracy more rapidly than FLDA in 8-electrode configuration in all subjects except S2. Also in 16-electrode configuration; LPC-BLDA algorithm has reached maximum accuracy in fewer blocks for all subjects except S7. In the LPC- v SVM proposed algorithm, convergent speed to maximum accuracy was much better than FLDA algorithm.

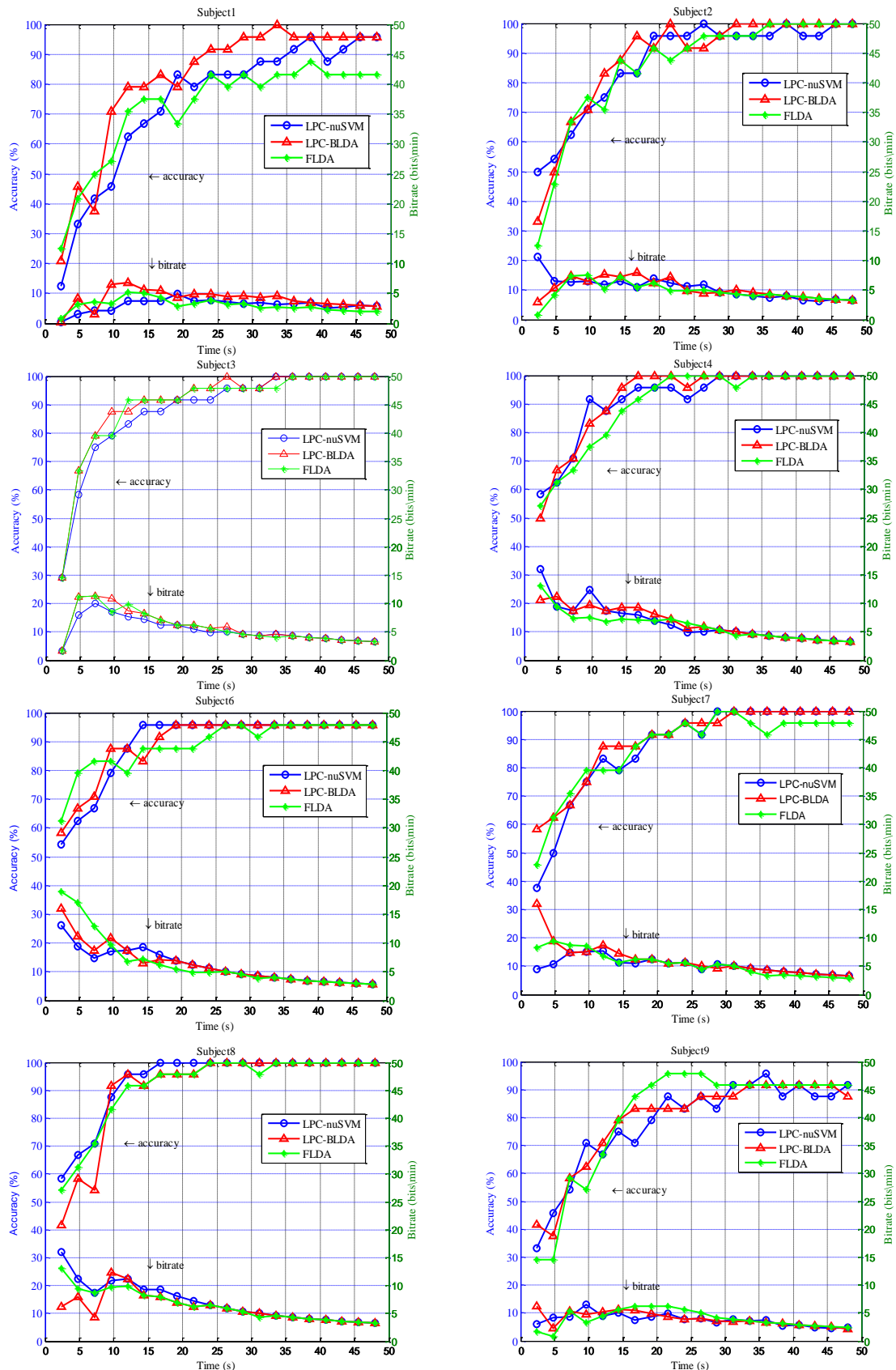


Figure 6. Comparison of classification accuracy and transmitting bitrates in proposed algorithms and FLDA by selecting 8-electrode configuration

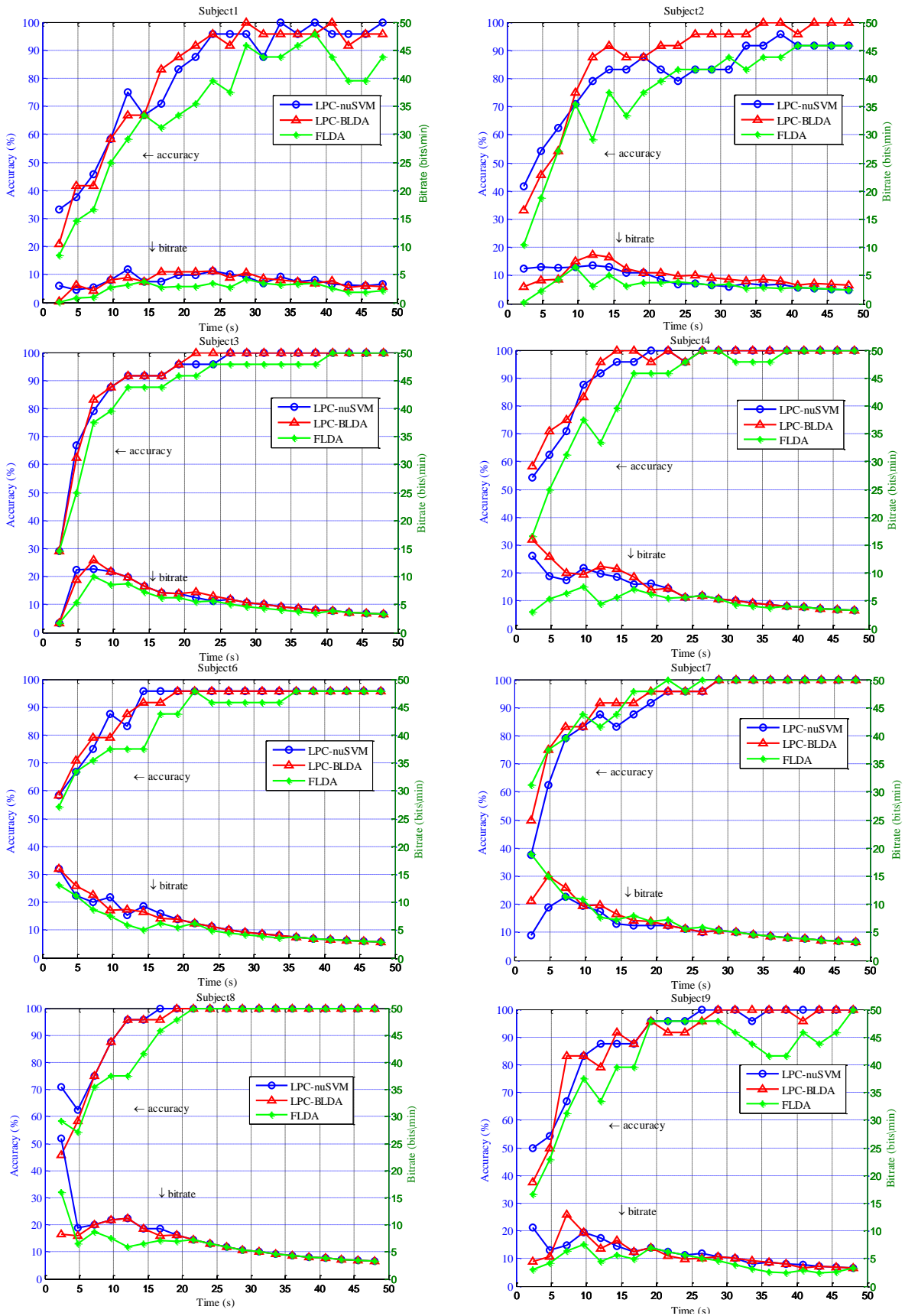


Figure 7. Comparison of classification accuracy and transmitting bitrates in proposed algorithms and FLDA by selecting 16-electrode configuration

For all subjects who reached maximum accuracy, the 8-electrode configuration algorithm has done better. Convergent speed to maximum accuracy was better in 16-electrode configuration for all who have reached maximum accuracy, except S7.

Classification accuracy of the proposed algorithms and FLDA algorithm in 8 and 16-electrode configurations is illustrated in Tables 1 and 2. According to the Table 1, the LPC-BLDA algorithm was better than FLDA algorithm in 8-electrode configuration for all disabled persons. For example, Classification accuracy improvement rate for subjects S1 and S2 in LPC-BLDA algorithm compared to the FLDA algorithm is 13% and 3%, respectively. For classification accuracy improvement, similar results obtained in LPC-BLDA algorithm by selecting 16-electrode configuration for the handicaps which are shown in Table 2.

Table 1. Comparison of quantitative results of the classification accuracy in proposed algorithm for 8-electrode configuration

Subject	FLDA	LPC + v-SVM	LPC+BLDA
S1	72.3	74.0	81.7
S2	85.4	87.1	87.9
S3	89.8	88.1	90.4
S4	90.4	91.9	92.5
S6	89.2	89.4	89.6
S7	87.1	87.3	89.8
S8	91.9	93.7	91.0
S9	80.4	78.1	79.2
Average (S1-S4)	84.5±7.3	85.3±6.7	88.1±4.0
Average (S6-S9)	87.1±4.2	87.1±5.7	87.4±4.8
Average (all)	85.8±6.1	86.2±6.3	87.8±4.4

Table 2. Comparison of quantitative results of the classification accuracy in proposed algorithm for 16-electrode configuration

Subject	FLDA	LPC + v-SVM	LPC+BLDA
S1	69.8	80.8	80.6
S2	75.0	81.0	86.2
S3	87.5	91.2	91.7
S4	86.2	92.7	93.7
S6	86.0	90.4	90.2
S7	93.1	90.0	92.5
S8	90.2	94.4	92.7
S9	81.9	90.0	89.2
Average (S1-S4)	79.6±7.5	86.4±5.5	88.0±5.1
Average (S6-S9)	87.8±4.2	91.2±1.8	91.1±1.5
Average (all)	83.7±7.3	88.8±4.8	89.6±4.1

The LPC-BLDA algorithm is better than the FLDA algorithm for healthy subjects by 8-electrode configuration for all, except S8 and S9, and in 16-electrode configuration except S7, too. By selecting 8-electrode configuration for all handicaps, the obtained classification accuracy in the LPC-vSVM algorithm was better than to the FLDA one, except S3. Also classification accuracy of LPC-vSVM algorithm compared to FLDA one improves for all handicaps in 16-electrode configuration. The improvement rate for subjects S1 and S2 is 15.7% and 8% in this configuration, respectively. The LPC-vSVM algorithm performance improves for healthy people in 8-electrode configuration except S9 and in 16-electrode configuration except S7.

Conclusions

A BCI system based on visual evoked potential of EEG-P300 was evaluated in this paper in terms of classification accuracy, convergent speed to maximum accuracy and bitrates. As mentioned before, EEG signals have a lot of noise. So, it's necessary to utilize an efficient technique to increase the amount of signal to noise ratio. Using a LPC model can greatly leads to reduce of noise in BCI system. In this paper, this model was used to extract the main features and eliminate artifacts of EEG signal. Also, two proposed classifiers, BLDA and ν -SVM, were applied in order to classify. Implementation results showed that the proposed algorithm was totally better than the FLDA. For example, in LPC- ν SVM and LPC-BLDA, the average classification accuracy for all subjects is improved by the amount of 0.8% and 2%, respectively relative to FLDA method. The same result can be seen for configuration of 16 electrodes. The average classification accuracy for all subjects in our two proposed algorithms, LPC- ν SVM and LPC-BLDA, is improved by the amount of 5.1% and 5.9%, respectively. This implies the superiority of our two proposed algorithms.

References

1. Khemri NA. P300 Wave Detection Using a Commercial non-Invasive EEG Sensor: Reliability and Performance in Control Applications, Oklahoma State University 2012.
2. McFarland DJ, Wolpaw JR. Brain-Computer Interfaces for Communication and Control. *Communications of the ACM* 2011;54(5):60-66.
3. Hoffmann U. Bayesian Machine Learning Applied in a Brain-Computer Interface for Disabled Users [PhD thesis]. École polytechnique fédérale de lausanne; 2007.
4. Wessel M. Pioneering Research into Brain Computer Interfaces [PhD thesis]. Delft University of Technology; 2006. Available from: URL: http://www.kbs.twi.tudelft.nl/docs/MSc/2006/Wessel_Mark/thesis.pdf, 2006.
5. Cabrera AR. Feature Extraction and Classification for Brain-Computer Interfaces [PhD thesis]. Aalborg University, Denmark; 2009. Available from: URL: http://vbn.aau.dk/files/18941236/phdthesis_alvarocabrera.pdf, 2009.
6. Wolpaw JR., Birbaumer . Brain-Computer Interfaces for Communication and Control. *Clinical Neurophysiology* 2002;113(6):767-791.
7. Bakhshi A, Ahmadi Fard A. A Comparison Among Classification Accuracy of Neural Network, FLDA and BLDA in P300-based BCI System. *Int J of Computer Applications* 2012;46(19):11-15.
8. Zhang H, Guan C. Asynchronous P300-based Brain-Computer Interfaces: A Computational Approach with Statistical Models. *IEEE Trans on Biomed Eng* 2008;55(6):1754-1763.
9. Birbaumer N. Breaking the Silence: Brain-Computer Interfaces (BCI) for Communication and Motor Control. *Psychophysiology* 2006;43(6):517-532.
10. Donnerer M, Steed A. Using a P300 Brain-Computer Interface in an Immersive Virtual Environment. *Teleoperators and Virtual Environments* 2010;19(1):12-24.
11. Panda R., Khobragade P. Classification of EEG Signal using Wavelet Transform and Support Vector Machine for Epileptic Seizure Diction. *IEEE International Conference on Systems in Medicine and Biology*, 2010.
12. Naderi MA, Mahdavi-Nasab H. Analysis and Classification of EEG Signals using Spectral Analysis and Recurrent Neural Networks. *17th Iranian Conference of Biomedical Engineering (ICBME)*, 2010.
13. Anderson CW, Sijercic Z. Classification of EEG Signals from Four Subjects During Five Mental Tasks. *Solving Engineering Problems with Neural Networks: EANN Symp Proc.* 2005.
14. Hoffmann U, Vesin JM. An Efficient P300-based Brain-Computer Interface for Disabled Subjects. *Journal of Neuroscience Methods* 2008;167(1):115-125.
15. Subasi A, Erçelebi E. Classification of EEG Signals using Neural Network and Logistic Regression. *Computer Methods and Programs in Biomedicine* 2005;78(2):87-99.

16. Gray RM. Linear Predictive Coding and the Internet Protocol, Now Publishers Inc; 2010.
17. Kugelstadt T. Active Filter Design Techniques 2001;Chapter 16, p. 271-281.
18. Rabiner LR,Schafer RW. Digital Processing of Speech Signals, Pearson Education; 1979.
19. Hoffmann U, Yazdani A. Bayesian Feature Selection Applied in a P300 Brain-Computer Interface2008.
20. Theodoridis S,Koutroumbas K. Pattern Recognition and Neural Networks. Machine Learning and Its Applications, Springer 2001; pp. 169-195.
21. Sellers EW,Krusienski DJ. A P300 Event-Related Potential Brain-Computer Interface (BCI): the Effects of Matrix Size and Inter Stimulus Interval on Performance. Biolog Psychology 2006;73(3):242-252.
22. Chang C, Lin CJ. LIBSVM: a Library for Support Vector Machines. ACM Trans on Intell Sys and Tech (TIST) 2011;2(3):27-35.
23. Settles M. An Introduction to Particle Swarm Optimization. Department of Computer Science, University of Idaho 2005.