CLASSIFICATION ALGORITHMS APPLIED TO A BRAIN COMPUTER INTERFACE SYSTEM BASED ON P300

A Paper
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Himanshu Gaur

In Partial Fulfillment of the Requirements
for the Degree of
MASTER OF SCIENCE

Major Department:
Computer Science

May 2017

Fargo, North Dakota

North Dakota State University
Graduate School

Title

CLASSIFICATION ALGORITHMS APPLIED TO A BRAIN COMPUTER INTERFACE SYSTEM BASED ON P300

By

Himanshu Gaur

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

MASTER OF SCIENCE

SUPERVISORY COMMITTEE:

Dr. Simone Ludwig
Chair

Dr. Jun Kong

Dr. Saeed Salem

Dr. Sanku Mallik

Approved:

May 18, 2017
Date

Dr. Kenneth Magel
Department Chair

ABSTRACT

A BCI or Brain Computer Interface is defined as a method of communication that converts neural activities generated by brain of living being (without the use of peripheral muscles and nerves) into computer commands or other device commands. BCI systems are useful for people with severe disability who have no reliable control over their muscles in order to interact with their surrounding environment. The BCI system used in this paper has used P300 evoked potential and three classifiers namely Logistic Regression (LR), Neural Network (NN), and Support Vector Machine (SVM). The system is tested with four people with severe disability and two able-bodied people. Classification accuracies obtained from LR, NN, SVM classifiers is then compared with Bayesian Linear Discriminant Analysis (BLDA) classifier and with each other. The relevant factors required for obtaining good classification accuracy in P300 evoked potential based BCI systems is also being explored and discussed.

ACKNOWLEDGEMENTS

I wish to express my deep sense of respect, and indebtedness to my adviser, Dr. Simone Ludwig for her valuable suggestions, and ceaseless encouragement during the research. I would like to thank my parents Mr. Dinesh Dutt Sharma and Mrs. Sadhana Sharma, Gopal Prasad Gaur and my entire family and my friends for their support, good wishes and encouragement. I would also like to thank all the committee members Dr. Jun Kong, Dr. Saeed Salem, Dr. Sanku Mallik for their valuable support.

TABLE OF CONTENTS

LIST OF TABLES

<u>Table</u> <u>Page</u>

LIST OF FIGURES

LIST OF APPENDIX FIGURES

CHAPTER 1. INTRODUCTION

A Brain-Computer Interface or BCI system is described as a communication system that enables people with a disability or able-bodied subjects to interact with their surroundings by utilizing only the neural activity produced by the brain without the use of peripheral muscles and nerves. BCI systems research has been investigated on a wider scale. The major reason behind the development of BCI systems is to make the communication for subjects with disabilities possible with other people, control artificial limbs, and to interact with their surrounding environment. In order to develop BCI based systems, there is a wider need of several technologies such as the processing of the patterns of neural activities generated by the brain into computer commands, algorithms for translation of signals from brain into commands that could be understood by the computer and the evaluation of BCI systems for subjects with a disability. In several papers, the research is focused on the above problems. The majority of the research is carried out for invasive and noninvasive technologies evaluation such that neural brain activity can be measured and new BCI systems can be developed. (Wolpaw et al. (2002); Lebedev and Nicolelis (2006)).

EEG or Electroencephalogram has the ability to record electrical brain activity. BCI systems for subjects with a disability have been used in this paper and are based on a noninvasive method (EEG) to measure the neural activity of the brain. Birbaumer et al. (1999) was among the first to use EEG or Electroencephalogram in the system for persons with disability. In his paper, he tested subjects with disability by using EEG to show that subjects who had Amyotrophic Lateral Sclerosis or ALS can utilize BCI systems to interact with their surrounding environment and operate via a spelling device. The research result of his work showed that subjects could acquire voluntary control of slow cortical potential. Pfurtscheller and

Neuper (2001) developed a BCI system that utilized the neural activity of brain associated to motor-imagery used as a control signal. Kubler et al. (2005) obtained excellent results for people suffering from ALS, quadriplegic persons and other forms of disability. His study depicted that people suffering from ALS have the ability to learn regulation of BCI systems utilizing motor imagery. Drawbacks of the BCI systems utilizing motor imagery was that since it used slow cortical potentials, subjects had to undergo training for a long time for several months and the subjects could only communicate slowly. Conversely to Kubler's research results, Hill et al. (2006) examined a BCI system utilizing motor imagery with a number of subjects, who were locked in, and showed that the signals obtained were not appropriate for communication. The main difference between the study of Kubler and Hill is not completely locked in subjects and a number of training sessions (Kubler) verses completely locked in subjects and one long training session (Hill). Based on the above research conducted, it has been established that subjects with disability can utilize BCI systems using motor-imagery, exceptions being locked in patients and a long training session.

P300 is defined as an Event Related Potential (ERP) and as a positive deflection in EEG. In this paper, a control signal called P300 related potential (Sutton et al., 1965) has been utilized that one can detect and it does not need lengthy training of the subjects. The P300 related potential is a positive detection in the human EEG, appearing approximately 300ms after the presentation of rare or surprising task-relevant stimuli. P300 was used by Farwell and Donchin (1988) as a control signal and a P300 speller system was utilized which enabled the subjects with disability to spell words in sequence by selecting letters from the alphabet. Here, randomly the rows and columns of the alphabet matrix were displayed in random sequence. The chosen subject was instructed to count the occurrence of the flashing symbol. This technique was adopted so that the subject can choose a symbol. Those columns of the flash, which contained the anticipated symbol, prompted EEG signals like P300. Neutral EEG signals were determined by white columns and rows. With the help of a simple algorithm, to evoke the largest P300 amplitude, the target symbol could be inferred. This algorithm searches for the row and column which evoked the largest P300 amplitude. This idea is based on the work of Farwell and Donchin.

Studies in the field of P300 based BCI systems were developed as a result of new application scenarios (Polikoff et al. (1995); Bayliss (2003)). The latest advanced algorithms for the recognition of the P300 were developed possibly from noisy data. The pioneers of this realm are Rakotomamonjy et al. (2005), Xu et al. (2004), Kaper et al. (2004), and Thulasidas et al. (2006). Now the study for the P300-based BCI systems shows good results for both the subjects with disability and able bodied subjects.

A 2D cursor regulator system having five disabled subjects and seven able-bodied subjects were tested by Piccione et al. (2006). Piccione et al. used four choice P300 paradigm as cursor control and subjects concentrating on one of four arrows flashing every 2.5s in random order. Preprocessed with independent component analysis, each of the signals was recorded from one electrooculogram electrode and four EEG electrodes. Furthermore, using a neural network the signals are categorized. The results obtained by Piccione et al. showed that the P300 was a viable control-signal for subjects with disability. However, when compared to other state-of-the-art systems the average communication speed obtained in his study was relatively low. The best examples are found by Kaper et al. (2004) and Thulasidas et al. (2006). Sellers and Donchin (2006) analyzed their system with 3 people having ALS and 3 people with able body. The system used signals from few electrodes, the small number of different stimuli and long inter stimulus intervals. The system also utilized a 4 choice paradigm. The corresponding paper that was published had 4 stimuli ('YES', 'NO', 'PASS', 'END'). These were flashed every 1.4s without following any sequential method. They were presented in different modes like either in the visual modality, in auditory modality, or in a combined auditory visual modality. Using a stepwise linear discriminant algorithm, the signals from three electrodes were categorized. Thereby, it was proved that the P300 based communication is possible in the auditory, combination of visual auditory and visual modality for ALS Patients by Sellers and Donchin.

But in the research of Piccione et al., when compared to state of the art outcomes, the classification accuracy and communication rate accomplished were low. This can be acknowledged to the small number of electrodes, the small number of different stimuli, and long ISIs (Inter-stimulus interval). Hoffmann and Vesin (2007) tested using Bayesian Linear Analysis (BLDA) and Fisher's Linear Discriminant Analysis (FLDA) for a population of 5 and 4, disabled and able-bodied subjects respectively with a paradigm of 6-choice. 6 different images were flashed in random order with a stimulus interval of 400ms. Electrode configurations consisting of four, eight, sixteen and thirty-two electrodes were verified.

For classification, Bayesian Linear Discriminant Analysis (BLDA) and Fisher's Linear Discriminant Analysis (FLDA) were tested. All the able-bodied subjects and four disabled subjects were found to have superior classification accuracy and communication rates compared to those in the research of Piccione et al. (2006), and Sellers and Donchin (2006). Discussions for disabled subjects about the factors that were relevant for good classification accuracy in BCI systems were made. To stimulate further study on data analysis methods for P300-based BCI systems and to enable other scholars to replicate results, some algorithms and datasets utilized in this research were made available for download on the EPFL BCI group website.

Hoffmann and Vesin displayed that high classification accuracies and birates can be obtained for severely disabled subjects by using BLDA and FLDA algorithms. Especially the results for BLDA were better than that of FLDA. Only a few training sessions were required to achieve good classification accuracy because P300 was used. Besides the BLDA and FLDA classifiers, various classifiers such as Neural Network, Support Vector Machine and Logistic Regression can be used.

In this MS paper, the study of the BCI system with the same control signal used by Hoffmann and Vesin is used. Through this paper, three machine algorithms are investigated; Neural Network (NN), Logistic Regression (LR) and Support Vector Machine (SVM) algorithms. In addition, a comparison between the investigated algorithms and previously used algorithm Bayesian Linear Discriminant Analysis (BLDA) is made.

CHAPTER 2. MATERIALS AND METHODS

2.1. Experimental setup

Six images displayed on the laptop screen are the following: television, telephone, lamp, door, window and radio. Users can regulate electrical appliances through BCI system where each image is selected according to the situation of the application. Images are flashed in a random sequence one after the other. Inter-stimulus interval is calculated as 400ms when each flash of an image persists for 100ms and no image is flashed during the subsequent 300ms. In the 10-20 International System, 16 electrodes are located and the EEG is recorded at a sampling rate of 2048 Hz. A bio semi-active two amplifier is used for ADC (Analog to Digital Conversion) amplification of the EEG signals. All machine learning algorithms and signal processing are implemented using MATLAB.

2.2. Subjects

Two healthy (S6, S7) and four disabled subjects (S1, S2, S3, S4) are used for testing the system. The table below illustrates data for the disabled subjects (Table 1). All the disabled subjects used in the study are wheelchair bound with different abilities for communication and limb muscle control. Disabled subjects S1 and S2 from the table perform limited action like slow and simple movements with their arms and hands. These subjects are restricted to perform or control other margins. Both subjects, S1 and S2, suffer from minor dysarthria, with the ability to speak and communicate. Performances are restricted for the only ability to make movements with his left hand. However the subject answers with eye blinks for yes or no.

Table 1: Disabled subjects and corresponding data recorded for environment control system study

	Subject 1 (S1)	Subject 2 (S2)	Subject 3 (S3)	Subject 4 (S4)
Diagnosis	Cerebral Palsy	Multiple Sclerosis	Late-Stage Amyotrophic Lateral Sclerosis	Traumatic Brain and Spinal-Cord Injury, C4 Leve1
Subject's Age	56	51	47	33
Age at Illness Onset	0 (Perinatal)	37	39	27
Sex	M	M	M	F
Speech Production (Dysarthria)	Mild	Mild	Severe	Mild
Limb Muscle Control	Weak	Weak	Very weak	Weak
Respiration Control	Normal	Normal	Weak	Normal
Voluntary Eye Movement	Normal	Mild Nystagmus	Normal	Normal

Subject 4 (S4) can communicate but can hardly move his limbs as he is suffering from mild dysarthria. Two PhD (S6, S7) male students working in the lab participated in the study aged around 30. Subject 6 (S6) and Subject (S7) did not have any recognized neurological problems.

2.3. Experimental schedule

Each and every subject finished 4 sessions of recording as scheduled. On the first day, 2 sessions were completed and another day was chosen to complete the other 2 sessions. The duration of time difference between the first and the last session was less than 2 weeks. Each session included a total of 6 runs; each run included 6 images. The below mentioned practices were used for each of the runs.

(i) Each subject was supposed to count in silence to count prearranged images flashed.

(ii) When each of the six images was flashed on the screen a warning sound was issued.

(iii) Random sequences of flashes were started and the EEG was recorded, four seconds after the warning tone. The flashes were torched in random sequences i.e.; each image was flashed once after 6 flashes, followed by flashing the image twice after he next twelve and the process progresses goes on. The blocks are chosen in random between 20 and 25 count. In one run, an average of 22.5 flashes were displayed, i.e., one run = average of 22.5 target P300 target trials and 22.5 x 4 = 90 non-target non-P300 trials.

(iv) The target image was concluded from the EEG with the aid of a simple classifier in the 2nd, 3rd, and 4th sessions. To provide response to the user, the image is flashed five times at the end of each run of the image inference. At the end of the experiment, each of the selected subjects were requested to provide their counting result. To monitor the performance of the subjects, the count results were collected in order. The calculation is approximated ('α' is used) as follows: The duration of one run is ~1 minute. The duration of between runs is ~ 30 minutes. One session trials (1 session includes setup of electrodes and short breaks) with the whole data for one subject results in 3240 trials.

2.4. Offline analysis

An offline procedure was adopted to test the machine-learning algorithms on the classification accuracy and the impact of various electrode configurations. To assess the average classification accuracy, four-fold cross-validation was used for each subject. To be precise, the data from three recording sessions and the data from the sessions that were not considered was utilized to train the classifier and validate it, respectively. Each session was run once for the validation by repeating the procedure four times.

2.4.1. Preprocessing

A number of preprocessing maneuvers were applied to the data, before the learning and validation of the classification function. The preprocessing operations followed the operations listed below in the following order:

(i) **Referencing**

For referencing, mean signal coming from 2 mastoid electrodes was utilized.

(ii) **Filtering**

To filter the data, we have used a forward-backward Butterworth band-pass filter of the 6th order. Cut-off frequencies were set to be 1.0 Hz and 12.0 Hz. Function *filtfilt* was used to filter, and the filter coefficients were computed using the function *Butterworth*.

(iii) **Down sampling**

By choosing each 64th sample from the band-pass filtered data, the EEG was down sampled to 32 Hz from 2048 Hz.

(iv) **Single trial extraction**

Single trials take place at the start of the amplification (i.e. at the stimulus onset) of an image, and ends 1000ms after the stimulus onset. For the study, 1000ms were extracted from the data (single trials of duration = 1000ms). The first 600ms of each trail were overlapping last 600ms of each preceding trial due to the ISI of 400ms.

(iv) **Windsorizing**

Any muscular activities including eye blinks, eye movements or physical subject movements can cause huge amplitude outliers in the EEG. To decrease the effect of outliers, the data from each electrode were windsorized. 10th and the 90th percentiles of the samples from each of the electrodes were calculated. Out of the computed amplitudes, some were set as the 10th percentile and others as the 90th percentile, for values which were above the 90th percentile or which were below the 10th percentile, respectively.

(v) **Scaling**

Scaling interval was chosen as [-1; 1].

(vi) **Electrode selection**

Four electrode configurations were tested with different number of electrodes.

(vii) **Feature vector construction**

Feature vectors were generated by concatenating the samples from the selected electrodes. The dimensionality of the feature vectors was Ne x Nt, where Ne is the count of the electrodes, and Nt is the count of the temporal samples present in a single trial. Ne = 4 or 8 or 16 or 32 (based on electrode configuration) and Nt =32 (based on trial time duration = 1000ms and down sampling = 32 Hz).

2.4.2. Machine learning and classification

The fourth session that was left of the training was validation. Both the training and validation data sets consisted of several sets of target and non-target trials. These average values as per Section 2.3 are as follows:

Training data sets: 405 target trials & 2025 non-target trials.

Validation data sets: 135 target trials & 675 non-target trials.

To understand the classifiers - Regression, Neural Network and Support Vector Machine were used. The corresponding acronyms used are LR, NN and SVM, respectively. Classifiers were compared with BLDA to compare the performance of these algorithms with the standard algorithm frequently used in BCI research. No user intervention of any kind was entertained to adjust hyper parameters by making all algorithms completely automatic. On a standard computer, the computation of the classifiers took not more than a minute. Followed by the training of the classifiers, the model was applied to the validation. By means of the preprocessing mechanism, the single trials, which correspond to the flashes of the first 20 blocks, were obtained for each run in the validation session. After which, the single trials were classified. This lead to an output of twenty blocks of classifiers. Each block contained six classifier outputs of which each output represents one image on the display. The classifier outputs were added over the blocks for each image. The image with the maximum classifier output was selected and was considered as the image the user was concentrating on.

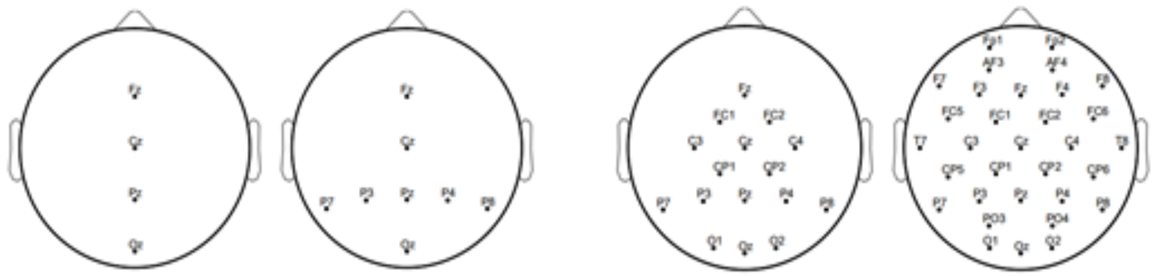

Figure 1: Electrode configurations used in the experiments.

From left to right: Configuration I (4 electrodes), configuration II (8 electrodes), configuration III (16 electrodes), and configuration IV (32 electrodes).

2.4.3. Experimental procedure

A cross validation procedure is used to evaluate the efficiency of the classification algorithm. Three session's data is used for training and the remaining session's data is used for testing. This procedure is repeated four times in such a way that every session's data is taken as test data once. The performance metrics considered are average accuracy and bitrate. They are calculated as follows.

- A subject is tested in 4 sessions, each session consist of 6 runs which is split into 20-25 blocks. Each block has 6 trials. Only the first 20 blocks in each run are considered for the performance evaluation. A single trial is of duration 400 ms, so a block is of length 2.4 s and the duration of 20 blocks is 48 s. Thus, the performance is evaluated from time 0 s to 48 s.
- For a single run, the number of correct classifications is calculated for each of the 20 blocks. This gives a 20-element array. The array elements can be either 1 or 0. A '1'

indicates correct classification in the particular block and vice versa. Repeating the process for all 6 runs and adding all the 6 arrays will give a 20-element array which has a maximum value of 6 and a minimum value of 0. If the nth element of that array is 6, that indicates the nth block is correctly classified in all 6 runs. If it is 0, then the nth block is wrongly classified in all 6 runs. Any in between value directly gives the number of times the nth block is correctly classified in 6 runs.

- The above-mentioned 20-element array is obtained from each of the 4 cross validation sessions and concatenated to form a 4x20 matrix. Then, the average across all sessions is computed by calculating the column mean for the matrix, and further dividing each element by 6. The resulting 20-element array gives the average accuracy obtained during every block of period 2.4 ms.
- The bits transferred during every block is $\log_2(6)$ if the accuracy is 1; here 6 is the number of images, and is 0 if the accuracy is 0 and it varies logarithmically for intermediate values. The bitrate is calculated by finding the ratio between the numbers of bits transferred in a particular block by the finish time of the block. So, even if two blocks transfer same number of bits, the block that is earlier in time will have more bitrate.
- The performance is plotted for block time (0 to 48 s) versus average accuracy (0 to 1) and bitrate.

CHAPTER 3. RESULTS

3.1. General observation

Classification accuracy averaged over the sessions and the corresponding bit rates versus the time required to come to a decision are depicted in Figure 2, Figure3 and Figure 4. The electrode configuration II with eight electrodes in conjunction with Logistic Regression, Neural Network, and Support Vector Machine (LR, NN and SVM) are displayed.

After 9 or more blocks of stimulus presentations were averaged for the LR classifier and it was determined that Subjects S3, S4 and S7 had achieved 100% classification accuracy whereas Subjects S1, S2, S6 did not achieve 100% accuracy. This is demonstrated in Figure 2.

After 3 or more blocks of stimulus presentations were averaged for the NN classifier, all the subjects, except S6 achieved an average classification accuracy of 100%. This is illustrated in Figure 3.

For the SVM classifier, all of the subjects, except Subject 6 achieved an average classification accuracy of 100% after 6 or more blocks of stimulus presentations were averaged. This is plotted in Figure 4.

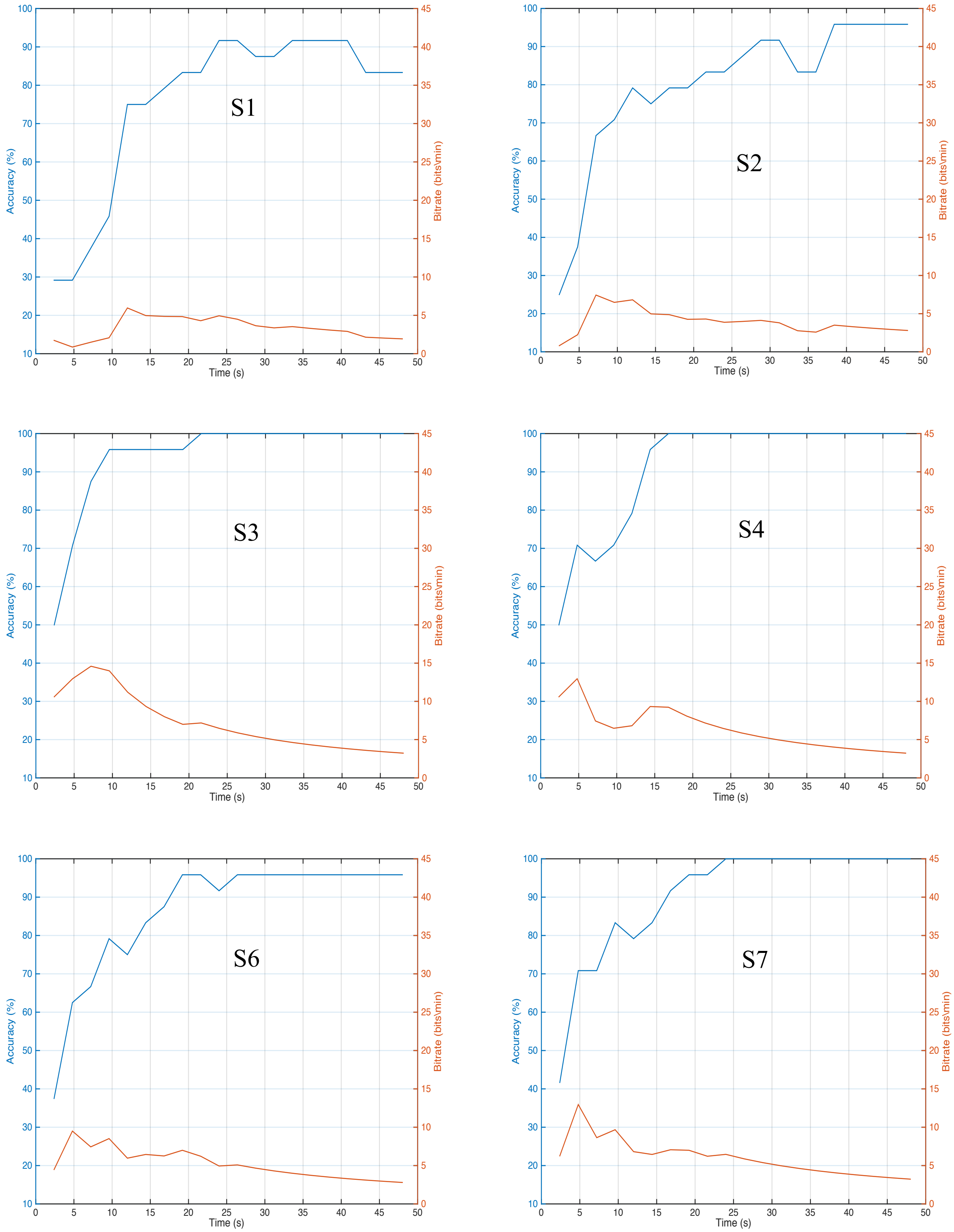

Figure 2: Logistic regression - classification accuracy & bit rate plotted vs. time

For subjects referred to as S1, S2, S3, S4, S6, and S7. The panels show the eight electrodes configuration and classification accuracy achieved with LR. Blue represents the four averaged sessions and Red represents the corresponding bitrate.

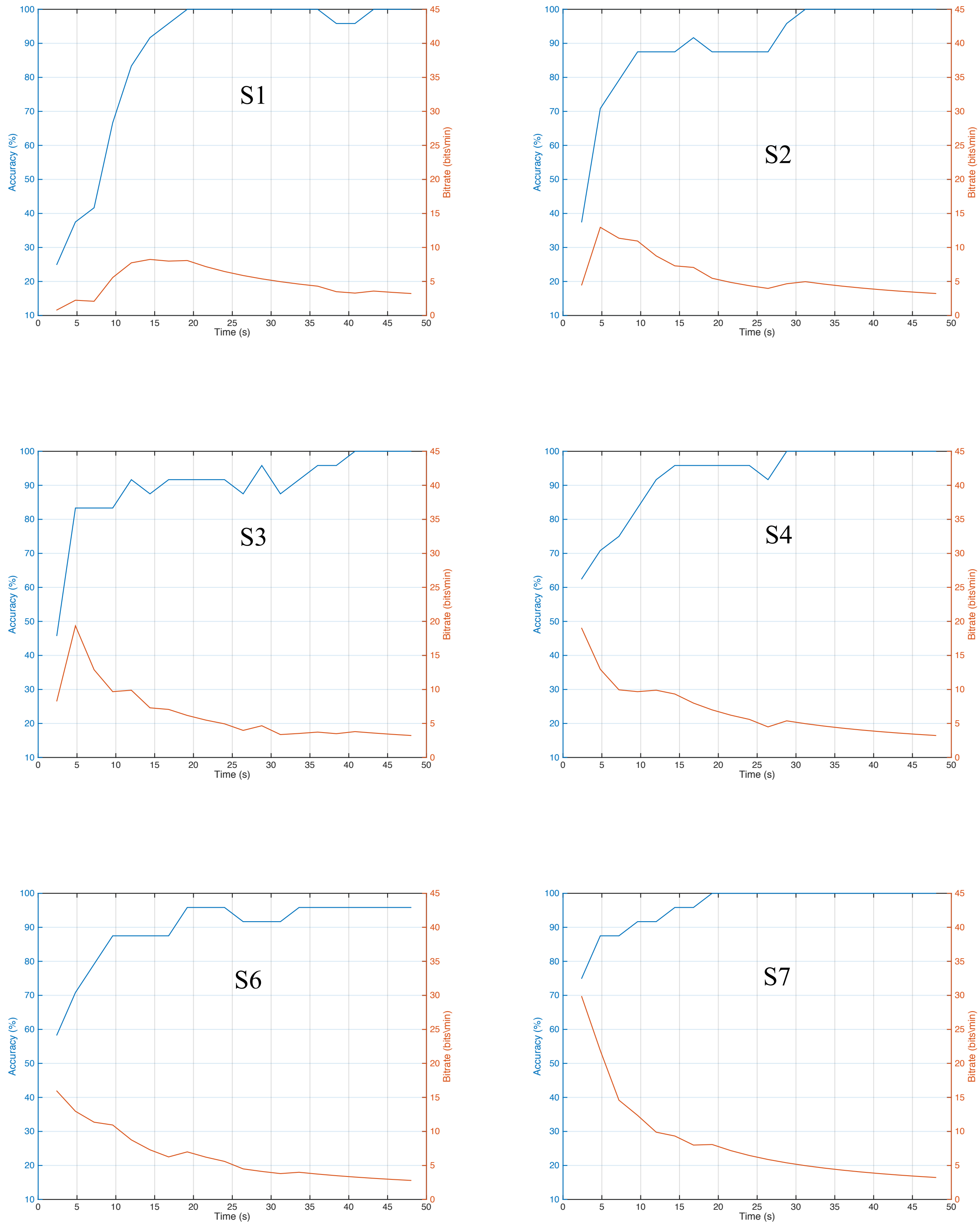

Figure 3: Neural network - classification accuracy & bit rate plotted vs. time

For subjects referred to as S1, S2, S3, S4, S6, and S7. The panels show the eight electrodes configuration and classification accuracy achieved with NN. Blue represents the four averaged sessions and Red represents the corresponding bitrate.

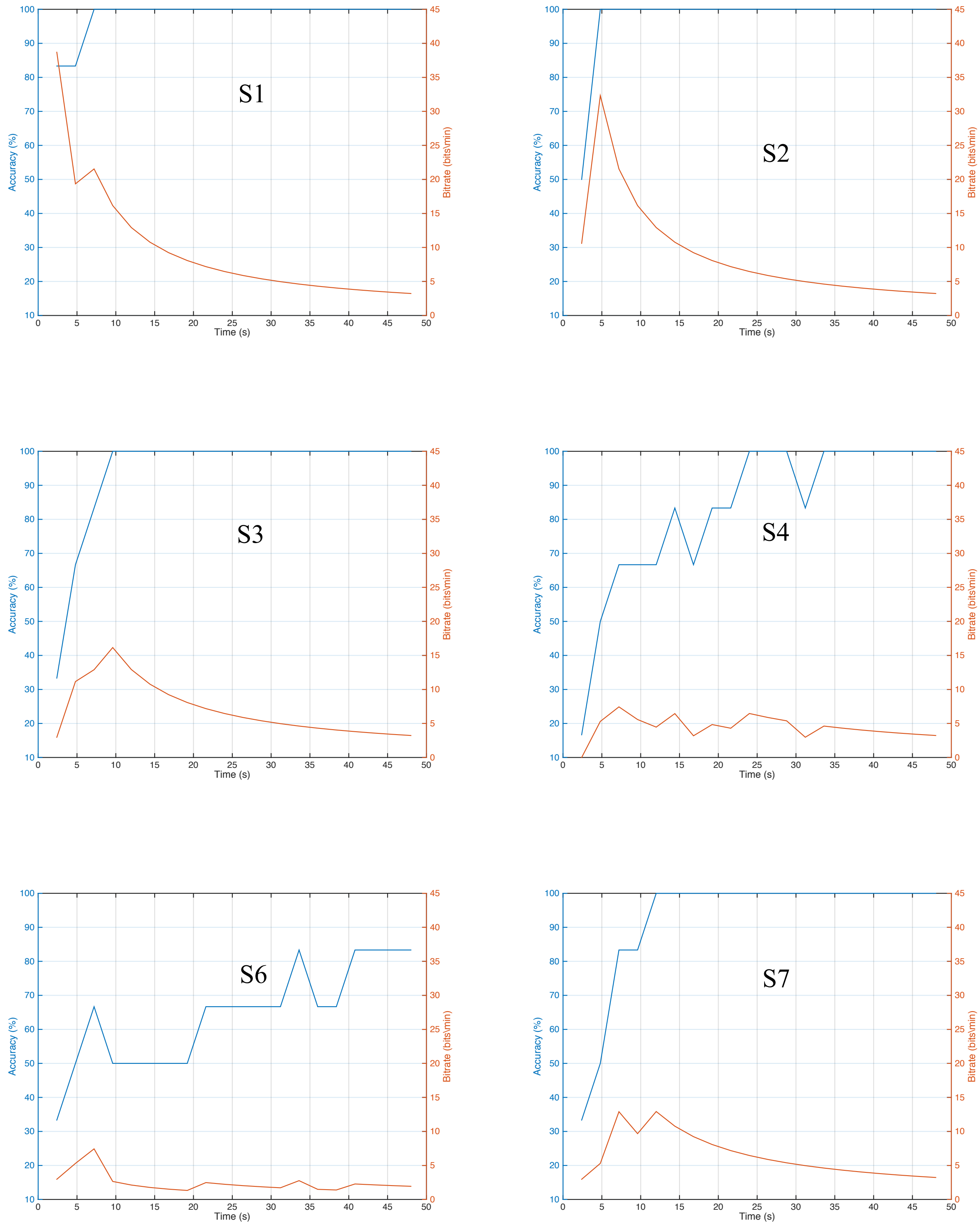

Figure 4: Support vector machine - classification accuracy & bit rate plotted vs. time

For subjects referred to as S1, S2, S3, S4, S6, S7. The panels show the eight electrodes configuration and classification accuracy achieved with SVM. Blue represents the four averaged sessions and Red represents the corresponding bitrate.

3.2. Differences between disabled and able-bodied subjects

As proved by Hoffmann and Vesin (2006), the maximum classification accuracy cannot be used as the performance measure. No dissimilarities can be found between abled and disabled subjects, if the maximum classification accuracy achieved is taken as the performance measure. Subjects are readily observed using the bitrate. It is evident from the plot that able-bodied subjects accomplished higher bitrates compared to disabled subjects.

Table 2: Bitrates according to subjects and classifier and average bitrate per minute (bpm)

Subject	Logistic Regression LR	Neural Network NN	Support Vector Machine SVM
S1	3.32	4.35	6.73
S2	3.95	3.17	5.14
S3	7.17	7.62	3.67
S4	6.39	6.85	3.83
S6	6.04	6.65	4.08
S7	9.23	9.22	7.09
Avg. (S1-S4)	5.21±1.86	5.49±2.09	4.84±1.42
Avg. (S6-S7)	6.82±2.12	7.58±1.42	4.56±2.33
Avg. (All)	5.90±1.99	6.39±2.02	4.72±1.69

The average bitrate per minute over all classifiers and the subjects are shown in Table 2. The average accuracy curves are used to compute the bitrates are shown for all classification algorithms. For both disabled subjects numbered 1 through 4, and able-bodied subjects numbered 5 through 7, the standard deviation and mean bitrate were computed.

Subject 1: The classification accuracy acquired with LR, averaged over all sessions, is calculated as 91.67%. The bitrate per minute corresponding to the calculated value is 3.32. After 11 blocks of stimulus exhibitions, the classification accuracy acquired with NN, averaged over all the sessions, was calculated as cent percent (i.e. 100%). The corresponding bitrate per minute is calculated as 4.35. For classifier SVM, after 5 blocks of stimulus presentations, unlike a NN where 6 more stimulus exhibitions were used to obtain 100% for the classification accuracy, averaged over all sessions. The corresponding bitrate per minute is calculated as 6.73. From the study it is evident that the maximum bpm (bitrate per minute) and maximum average bpm for SVM is higher than others for subject 1 (S1). Therefore it can be concluded that, the performance for SVM is the best for Subject 1 and it is determined to the worst for LR in subject 1. Subject 1 accidentally focused on the wrong stimulus during runs in sessions 1 through 4 in LR.

Subject 2: The classification accuracy with LR, averaged over all sessions, is 95.83%. The corresponding bitrate per minute is calculated as the value of 3.95. Likewise, the same 95.83% was obtained with NN, averaged over all sessions. Although the corresponding bitrate per minute is 3.17 which is less than that obtained with LR. For Subject 2, after 5 blocks of stimulus presentations the classification accuracy obtained by SVM, averaged over all sessions, is calculated as 100% and the corresponding bitrate per minute is 5.14. From the study and the computed values it can be affirmed that the performance for SVM is the best for Subject 2. The maximum value of classification accuracy is same for both LR and NN with a smaller bpm for

NN compared to that of LR. For LR, during one run in Session 4, Subject 2 concentrated on the wrong stimulus. Whereas for NN, Subject 2 concentrated on the wrong stimulus through one run in Session 2 and 3. Therefore, the performance with LR is better than the one with NN.

Subject 3: 100% accuracy is obtained over all classifiers, averaged over all sessions. However, every bpm calculated is different from each other. After 9 blocks of stimulus presentations, the classification accuracy with LR is 100%. The corresponding bpm is computed as 7.17. After 3 blocks of stimulus presentations, the classification accuracy with NN is 100% and the corresponding bpm is 7.62. As for SVM, the classification accuracy with SVM is 100% after 14 blocks of stimulus presentations. The corresponding bpm is computed as 5.14. Therefore, it can be affirmed that the performance of NN is the best in case of Subject 3.

Subject 4: The classification accuracy obtained with LR and NN, averaged over all sessions, is 100%. After 7 blocks of stimulus exhibitions, the classification accuracy with LR is obtained as 100%. The corresponding bpm is determined as 6.39. After 11 blocks of stimulus presentations with NN, the classification accuracy is 100% and the corresponding bpm is 6.85. The classification accuracy is 83.33% for SVM. The corresponding bpm is 3.83, which is slightly less than that of NN. It can therefore be confirmed that the performance of LR is the best for Subject 4.

Subject 6 and Subject 7: It was observed that performances of all classifiers were good. The classification accuracy reached 100% for all classifiers. LR has shown the best performance in this case. Figure 8 below illustrates for each of the three classifiers; LR, NN and SVM the change process of maximum average bmp according to the time. From the plot it can be determined that, the average bpm of LR and NN is greater than that of SVM. Thus, it can be concluded that the classifiers LR and NN are better compared to the classifier SVM. Of the two classifiers, LR and NN, it is determined that the maximum average bpm of NN is the highest. Consequently, it can be confirmed that the NN classifier performs best compared to the other classifiers used.

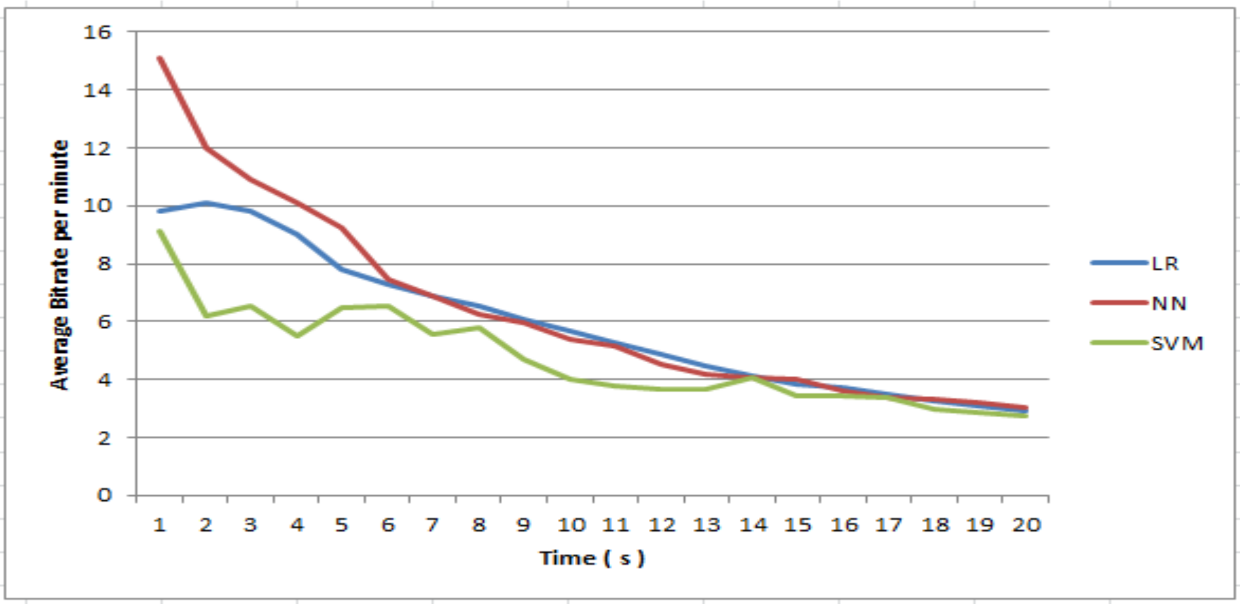

Figure 5: For 8-electrode configuration, average bitrate per minute

Acquired with LR, NN and SVM, averaged over all subjects and sessions are plotted against time.

3.3. Interpretation of results

3.3.1. Logistic regression

Out of the 6 subjects, 3 of them achieve 100% classification accuracy. The average accuracy is less during the earlier blocks and it increases to higher values as time progresses. This indicates the subjects produce an easily distinguishable P300 signal when given sufficient time. Earlier blocks give very little reaction time for the subject hence the signals are not clear enough for the classifiers to identify. The peak bitrate for subjects 1 and 2 is very less, but other subjects can achieve a peak bitrate. This indicates good response and better classification.

3.3.2. Neural networks

Compared to logistic regression, Neural Networks provide better performance. All 6 subjects achieve 100% classification accuracy, and the bitrates are fairly higher. As in the previous case the accuracy is low when a new run is started, but as time advances, the subjects tend to settle with the new target and the classifier identifies correctly.

3.3.3. Support vector machine

SVM offers better classification accuracy for disabled subjects than subjects with no known disability. The bitrates achieved by subjects 1 and 2 with SVM is higher among any classification methods. But the poor performance of SVM in subject 6 can be considered as a deterrent in selecting SVM as a classifier.

3.3.4. Comparison with BLDA

BLDA achieved 100% classification accuracy for all the subjects. Only Neural Network's achieved similar classification accuracy out of the three classifiers tried in the work. Though SVM achieved better accuracy for most of the subjects, its result for S6 is very inconsistent. There is no way to distinguish between a disabled and able bodied subject from classification accuracy while using BLDA. But the bitrates obtained by BLDA classification clearly distinguishes disability. But the 3 classifiers used in this paper are not consistent in terms of distinguishing between able bodied and disabled subjects.

CHAPTER 4. DISCUSSION

4.1. Comparison with other studies

Compared to various other P300 systems for people with disability, the classification accuracy and bitrate obtained in the current study are relatively high. In the research of Sellers and Donchin (2006) the best classification accuracy for the able-bodied subjects was on average 85% and the best classification accuracy for the ALS patients was on average 72% (values taken from Table 3 in Sellers and Donchin (2006)). In the present study the best classification accuracy for the able-bodied subjects is on average close to 100% and the best classification accuracy for disabled subjects was on average 100% (see Figures 2, 3, and 4).

In the research of Sellers and Donchin, the bitrates (bits/min) were not mentioned. The average bitrates of approx. 8 bits/min were mentioned for people with disability and people with able-body in the research of Piccione et al. (2006). The average bitrate which was found using electrode configuration (II) to be 15.9 bits/min for the people with disability and 29.3 bits/min for the able-bodied people in the current research. The classification accuracy and bitrate found in the two research works cannot be associated with that acquired in the current study due to variation in subject populations and experimental paradigms. These factors described below are some that have been recognized.

<u>Number of choices:</u>

A six-choice model had been used in the current study compared to the research of Sellers, Piccione et al. and Donchin where four-choice paradigms were used. As a consequence the target stimulus occurred with a probability of 0.25 in the study of Sellers and Donchin and Piccione et al., but in the present work it occurred with a probability of 0.16. Smaller target probabilities are equivalent to higher amplitudes in P300 systems (Duncan-Johnson and

Donchin, 1977), thus the P300 in our system might have been simpler to discover. BCI systems based on P300 were designed taking into account that disabled subjects might suffer from visual impairments. Systems such as the P300 speller in which users have to focus on a relatively small area of the display might thus not be appropriate for disabled subjects. Reducing the number of choices enlarges the area occupied by one item on the screen and thus facilitates concentration on one item. This might be particularly important for subjects who have little remaining control over their eye-movements. Such subjects might use covert changes of visual consideration (Posner and Petersen, 1990) to regulate a BCI system based on P300, which should be easier when a small number of large items is used.

Inter-stimulus interval (ISI):

Several factors have to be kept in mind when choosing an ISI for a P300-based BCI system. Regarding the classification accuracy, longer ISIs theoretically yields better results. This should be the case since longer ISIs (within some limits) cause larger P300 amplitudes. On the other hand, a consequence of lengthy ISIs is a lengthier overall duration of runs. Disabled subjects might have difficulties to stay concentrated during long runs and thus P300 amplitude and classification accuracy might actually decrease for longer ISIs.

Regarding the bitrate, the factors described above have to be considered together with the fact that for a given classification accuracy with smaller ISIs, bitrates of higher values are obtained. Moreover, it should be acknowledged that if ISI is made shorter, subjects with cognitive deficits might have problems to detect all target stimuli and classification accuracy might decrease.

Given the complex interrelationship of several factors an optimal ISI for P300-based BCIs can only be determined experimentally. Here we have shown that an ISI of 400ms yields good results. Sellers & Donchin and Piccione et al. have used an ISI of 1.4s, and 2.5s, respectively. The corresponding outcomes achieved in their conclusions signify that the ISIs to be too long. Therefore, in this paper, the classification accuracy and bitrate obtained with LR, NN and SVM are compared directly to those obtained by BLDA (Bayesian Linear Discriminant Analysis). The performance for BLDA is shown in Figure 9 and Figure 10.

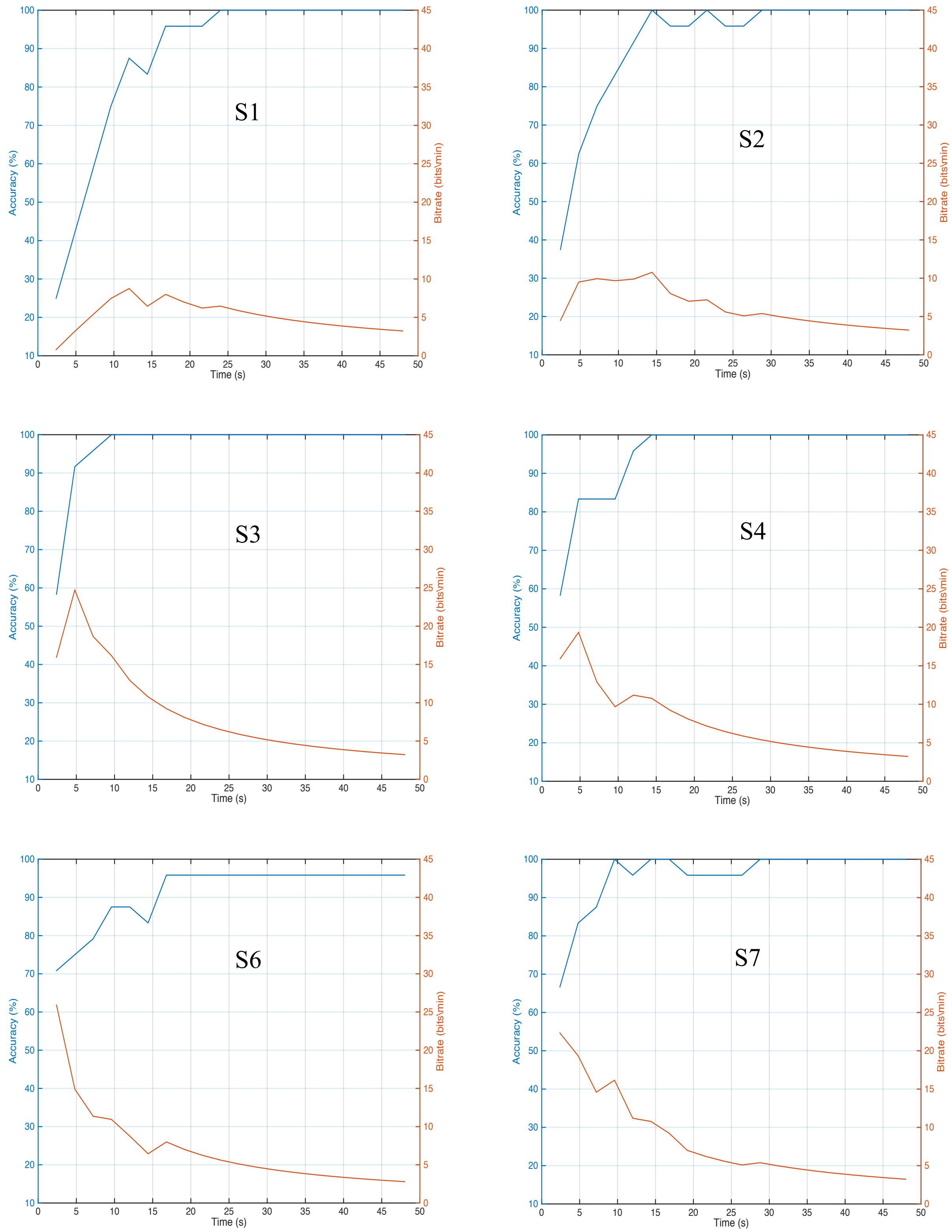

Figure 6: BLDA - classification accuracy & bit rate plotted vs. time

For subjects referred to as S1, S2, S3, S4, S6, S7. The panels show the eight electrodes configuration and classification accuracy achieved with BLDA. Blue represents the four averaged sessions and Red represents the corresponding bitrate.

For Subject 1, the classification accuracy obtained with BLDA, averaged over all sessions, is 100% after 10 blocks of stimulus presentations and the corresponding bitrate per minute is 6.46. As mentioned above, the classification accuracy obtained with LR, averaged over all sessions, is 91.67% and the corresponding bitrate per minute is 4.94; the classification accuracy obtained with NN, averaged over all sessions, is 100% after 11 blocks of stimulus presentations and the corresponding bitrate per minute is 5.87 and for SVM, the classification accuracy, averaged over all sessions, is 100% after 5 blocks of stimulus presentations, and the corresponding bitrate per minute is 12.92. As we can see from the above result, the performance of BLDA is better than the one of LR. And the classification accuracy of BLDA and NN is 100%, but the bmp of BLDA is larger than one of NN. Therefore, we can say that BLDA is better than NN. However, the bmp of BLDA is smaller than the bmp of SVM. So, the performance of BLDA is not better than the one of SVM.

For Subject 2, the classification accuracy obtained with BLDA, averaged over all sessions, is 100% after 5 blocks of stimulus presentations and the corresponding bitrate per minute is 10.77. The classification accuracy obtained with LR, averaged over all sessions, is 95.83%, and the corresponding bitrate per minute is 3.29, the classification accuracy obtained with NN, averaged over all sessions, is 95.83% and the corresponding bitrate per minute is 2.95, and the classification accuracy obtained with SVM, averaged over all sessions, is 100% after 5 blocks of stimulus presentations and the corresponding bitrate per minute is 12.92. In this case, the performance of BLDA is also better than the performance of LR and NN. But it is not better than the performance of SVM.

For Subject 3, the classification accuracy of all classifiers is 100%, but the classifier BLDA shows the best performance. The classification accuracy is 100% after 3 blocks of stimulus presentations and the corresponding bpm is 16.16, the classification accuracy with LR is 100% after 9 blocks of stimulus presentations and the corresponding bpm is 7.18, the classification accuracy with NN is 100% after 3 blocks of stimulus presentations and the corresponding bpm is 21.54, and the classification accuracy with SVM is 100% after 14 blocks of stimulus presentations and the corresponding bpm is 4.62. Among all classifiers, the bpm of BLDA is the largest. This means for the Subject 3 the performance of BLDA is the best.

For Subject 4, the classification accuracy obtained with BLDA, averaged over all sessions, is 100% after 5 blocks of stimulus presentations and the corresponding bitrate per minute is 10.7707. The classification accuracy obtained with LR and NN, averaged over all sessions, is 100%. The classification accuracy with LR is 100% after 7 blocks of stimulus presentations and the corresponding bpm is 9.23, the classification accuracy with NN is 100% after 11 blocks of stimulus presentations and the corresponding bpm is 5.87. For SVM, the classification accuracy is 83.33% and the corresponding bpm is 4.84. Therefore, for Subject 4, the performance of BLDA is also the best.

For Subject 5, the performance for all classifiers is less than 100%. The classification accuracy obtained with BLDA, averaged over all sessions, is 95.83% and the corresponding bpm is 7.99, the classification accuracy obtained with LR and NN, averaged over all sessions, is 95.83% and the corresponding bpm is 6.99, and the classification accuracy obtained with SVM, averaged over all sessions, is 83.33% and the corresponding bpm is 4.84. For BLDA, LR and

NN, Subject 5 concentrated on the wrong stimulus during one run in Session 1, but for SVM, Subject 5 concentrated on the wrong stimulus for all sessions. But the bpm of BLDA is also larger than others. Therefore, for the Subject 5, the performance of BLDA is better than the other classifiers.

For Subject 6 and 7, all performances of all classifiers are very good. For all classifiers, the classification accuracy of 100% is reached. Overall, the performance for BLDA is the best. The bpm of BLDA for Subject 6 is 16.16, and for Subject 7 is 24.72.

Table 3 shows the average bitrate per minute over all subjects and classifiers. From the table we can see the average bpm of BLDA is larger than the other classifiers. Therefore, we can conclude that BLDA is the best method compared to the other machine learning methods.

Table 3: Average bitrate per minute (bpm)

Subject	LR	NN	SVM	BLDA
S1	3.32	4.35	6.73	5.14
S2	3.95	3.17	5.14	6.22
S3	7.17	7.62	3.67	8.67
S4	6.39	6.85	3.83	7.69
S6	6.04	6.65	4.08	7.13
S7	9.23	9.22	7.09	8.24
Avg. (S1-S4)	5.21±1.86	5.49±2.09	4.84±1.42	6.93±1.56
Avg. (S6-S7)	6.82±2.12	7.58±1.42	4.56±2.33	6.93±1.56
Avg. (All)	5.90±1.99	6.39±2.02	4.72±1.69	7.62±1.68

Figure 7 shows the average bmp of each classifier according. As we can see, the average bpm of BLDA is larger than the one of the other classifiers. This means, the BLDA classifier is the better method compared to LR, NN, and SVM.

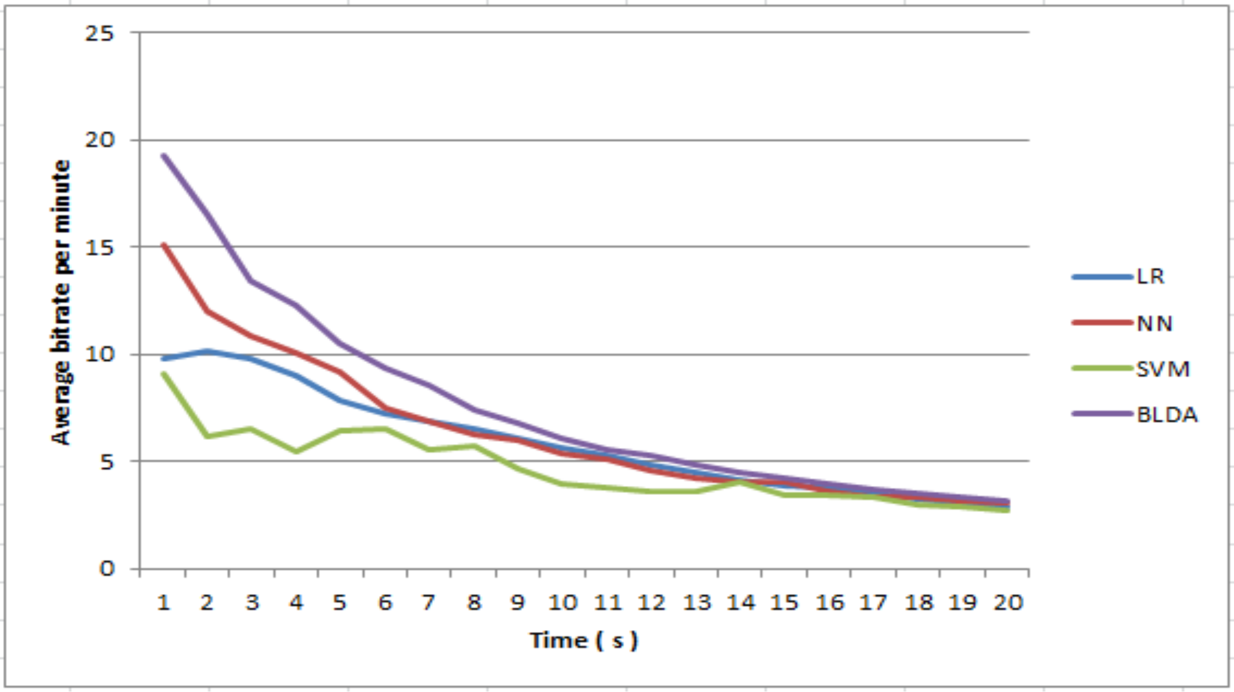

Figure 7: Average bitrate per minute vs time

4.2. Machine learning algorithm

Many of the characteristics of a BCI system depend critically on the employed machine learning algorithm. The important characteristics influenced by the machine learning algorithm are as follows: classification accuracy, communication speed, time and user intervention required to set up a classifier from the training data. FLDA is an effective and simple algorithm used in P300-based systems (Pfurtscheller and Neuper, 2001; Bostanov, 2004; Kaper, 2006). In one of Krusienski et al.'s comparisons of classification procedures (2006) for P300-based BCIs, it was determined in terms of classification accuracy and simplicity of usage, that the best method is FLDA when compared to others. However, using FLDA becomes impossible when the number of features becomes large, relative to the number of training models. This is acknowledged as the small sample size problem. The small sample size problem occurs because the between-class scatter matrix used in FLDA becomes singular when the number of features becomes larger. The answer to this setback in the current study was to use the Moore-Penrose pseudo inverse of the between-class scatter matrix. Even if the quantity of features is high it is still allowed to use FLDA in this case. However, with this approach the performance of FLDA deteriorated when the number of electrodes was increased.

In this work, LR, NN and SVM are employed. The classification accuracy of NN is higher than both LR and SVM. But the communication speed of it is very slow. The larger the dataset is, the slower the processing. When the number of features becomes large, the NN classifier is not suitable and SVM cannot reach high accuracy values. The LR classifier is the best suitable for classification accuracy and communication speed. In addition, BLDA is also considered. The common drawback of overfitting in BLDA, e.g., the small sample size, etc. are solved by using regularization. Without user intervention or time consuming cross-validation, the degree of regularization can be automatically assessed from training data through a Bayesian analysis approach.

CHAPTER 5. CONCLUSION AND FUTURE WORK

In this work an efficient P300-based BCI system for disabled subjects was presented. In the proposed algorithm, first the EEG signals were preprocessed using several stages that included referencing, filtering, downsampling, single trial extraction, windsorizing, scaling and electrode selection, to extract the feature vectors. These feature vectors were then fed to different classifiers namely Bayesian Linear Discriminant Analysis, Logistic Regression, Neural Network, and Support Vector Machine. Samples from both able bodied and disable persons were used for the experimentation. The classification accuracy averaged over sessions and the corresponding bitrates are used to determine the best classifier.

Through experimentation it is shown that high classification accuracies and bitrates can be obtained for severely disabled subjects. Due to the use of the P300, only a small amount of training data are required to achieve good classification accuracy. Bayesian Linear Discriminant Analysis has shown better performance than the other classifiers for all subjects.

Future improvements to the work presented here might consist of testing the system with completely locked-in patients and defining useful BCI applications customized to disabled users' requirements. Furthermore, it would be good to perform studies with larger numbers of subjects in order to confirm the results found in the present work.

REFERENCES

[1] A. Galushkin, “The structure of neural networks,” in *Neural Networks Theory*. Berlin: Springer, 2007, pp. 33-75.

[2] J. Bayliss, “Use of the evoked P3 component for control in a virtual apartment,” *IEEE Trans. Neural Syst. Rehab. Eng*. vol. 11, ed. 2, 2003, pp. 113-116.

[3] N. Birbaumer et al., “A spelling device for the paralyzed,” *Nature* vol. 398, Mar 1999, pp. 297-298.

[4] V. Bostanov, “Feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram.” *IEEE Trans. Biomed. Eng.* vol. 51, ed. 6, May 2004, pp. 1057-1061.

[5] M. Carrillo-de-la Pena and F. Cadaveira, “The effect of motivational instructions on P300 amplitude,” *Neurophysiologie Clinique/Clinical Neuro-physiology* vol. 30, ed. 4, Aug 2000, pp. 232-239.

[6] T. Centeno and N. Lawrence, “Optimising kernel parameters and regularisation coefficients for non-linear discriminant analysis,” *Journal of Machine Learning Research* vol. 7, Feb 2006, pp. 455-491.

[7] C. Wei, “Bayesian approach to Support vector machine,” Ph.D. thesis, Dept. of Mechanical Eng., National Univ. of Singapore, 2006.

[8] C. Duncan-Johnson and E. Donchin, “On quantifying surprise: The variation of event related potentials with subjective probability,” *Psychophysiology* vol. 14, ed. 5, Sep1977, pp. 456-467.

[9] L. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials." *Electroencephalogr. Clin. Neurophysiol.* vol. 70, ed. 6, Dec 1988, pp. 510-523.

[10] A. Galushkin, "Synthesis of Multilayer Neural Networks with Flexible Structure" in *Neural Networks Theory*. Berlin: Springer, 2007, pp. 273-293.

[11] A. Galushkin, "Continual Neural Networks" in *Neural Networks Theory*. Berlin: Springer, 2007, pp. 67-72.

[12] A. Galushkin, "Methods of Problem Solving in the Neural Network Logical Basis" in *Neural Networks Theory*. Berlin: Springer, 2007, pp. 341-368.

[13] A. Galushkin, "Neural Network Adjustment Algorithms" in *Neural Networks Theory*. Berlin: Springer, 2007, pp. 163-187.

[14] N. Hill, T. Lal et al., "Classifying EEG and ECoG signals without subject training for fast BCI implementation: Comparison of nonparalyzed and completely paralyzed subjects," *IEEE Trans. Neural Syst. Rehab. Eng*. vol. 14, ed. 2, Jun 2006, pp. 183-186.

[15] U. Hoffmann, G. Garcia et al., "A boosting approach to P300 detection with application to brain-computer interfaces" in *Proceedings of the 2nd International IEEE EMBS Neural Engineering Conference, Arlington, VA, USA, March 16-19, 2005,* Arlington: IEEE, 2005, pp. 97-100.

[16] U. Hoffmann, G. Garcia et al., 2004. "Application of the evidence framework to brain-computer interfaces" in *Proceedings of the 29th Annual International Conference of the IEEE, Lyon, France, August 22-26, 2007*, Lyon: IEEE, 2007. pp. 46-49

[17] U. Hoffmann et al., “Spatial filters for the classification of event-related potentials” in *Proceedings of the 14th European Symposium on Artificial Neural Networks (ESANN), Bruges, Belgium, April 26-28, 2006,* Bruges: ESANN, 2006, pp. 47-52

[18] U. Hoffmann et al., “An efficient P300-based brain-computer interface for disabled subjects,” *Journal of Neurosci. Methods*, vol. 167, ed. 1, Jan 2008, pp. 115-125.

[19] M. Kaper, “P300-based brain-computer interfacing,” Ph.D. thesis, Faculty of Technology, Bielefeld Univ., Germany, 2006.

[20] M. Kaper et al., “Support vector machines for the P300 speller paradigm,” *IEEE Trans. Biomed. Eng*, vol. 51, ed. 6, Jun 2004, pp. 1073-1076.

[21] D. Krusienski et al., “A comparison of classification techniques for the P300 speller” *J. Neur. Eng.* vol. 3, ed. 4, Dec 2006, pp. 299-305.

[22] A. KÄubler et al., “Patients with ALS can use sensorimotor rhythms to operate a brain-computer interface,” *Neurology* vol. 64, May 2005, pp. 1775-1777.

[23] M. Lebedev and M. Nicolelis, “Brain-machine interfaces: past, present and future,” *Trends in Neurosciences,* vol. 29, ed. 9, Sep 2006, pp. 536-546.

[24] D. MacKay, “Bayesian interpolation,” *Neural Comput.* vol. 4, ed.3, Jan 1992, pp. 415-447.

[25] G. Pfurtschelle and C. Neuper, “Motor imagery and direct brain-computer communication,” *Proceedings of the IEEE* vol. 89, ed. 7, Jul 2001, pp. 1123-1134.

[26] F. Piccione et al.,“P300-based brain-computer interface: Reliability and performance in healthy and paralyzed participants,” *Clin. Neurophysiol,* vol. 117, ed. 3, Mar 2006, pp. 531-537.

[27] J. Polikoff et al., "Toward a P300-based Computer Interface" in *Proceedings of the RESNA '95 Annual Conference, June 9-14, 1995, Vancouver, BC, Canada*, Arlington: RESNAPRESS, 1995, pp. 178-180.

[28] M. Posner and S. Petersen, "The attention system of the human brain," *Annual Review of Neuroscience*, vol. 13, ed. 1, Mar 1990, pp. 25-42.

[29] A. Rakotomamonjy et al., "Ensemble of SVMs for improving brain-computer interface P300 speller performances" in *Proceedings of 15th International Conference on Neural Networks, ICANN 2005, Warsaw, Poland, September 11-15, 2005,* Eds. Berlin: Springer, 2005. vol. 3696.

[30] E. Sellers and E. Donchin, "A P300-based brain-computer interface: Initial tests by ALS patients," *Clin. Neurophysiol,* vol. 117, ed. 3, Mar 2006, pp. 538-548.

[31] H. Serby et al., "An improved P300-based brain-computer interface," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 13, ed. 1, Mar 2005, pp. 89-98.

[32] S. Marinai and H. Fujisawa, *Machine Learning in Document Analysis and Recognition*. Firenze, Italy: Springer, 2008, pp. 72-138.

[33] S. Sutton et al., "Evoked-potential correlates of stimulus uncertainty," *Science,* vol. 150, ed. 700, Nov 1965, pp. 1187-1188.

[34] M. Thulasidas et al., "Robust classification of EEG signals for brain-computer interface," *IEEE Trans. Neural Syst. Rehab. Eng.* vol. 14, ed. 1, Mar 2006, pp. 24-29.

[35] Q. Tian et al., "Comparison of statistical pattern-recognition algorithms for hybrid processing. II. Eigenvector-based algorithm," *J. of the Optical Society of America A,* vol. 5, Oct 1988, pp. 1670-1682.

APPENDIX A. SUPPORT MACHINE LEARNING (SVM)

As computationally dominant gears for supervised learning, support vector machines (svms) are extensively employed in classification and regression issues. Assume that a data set $D = \{(x_i, y_i) \mid i = 1,..,n\}$ is known for training, where the target value is the input vector $x_i \in R^d$. A linear machine is built by decreasing a normalized functional when SVMs maps these input vectors into a high dimensional Reproducing Kernel Hilbert Space (RKSH). $f(x) =< w \cdot \phi(x') > +b$ is the form taken by the linear machine . Here, $\phi(\cdot)$ is the mapping function, b is known as the bias, and the dot product $< \phi(x) \cdot \phi(x') >$0i is also the reproducing kernel $K(x, x')$ in the RKHS. The regularized functional is commonly defined as;

$$R(w,b) = C \cdot \sum_{i=1}^{n} l(y_i, f(x_i)) + \frac{1}{2} \| w \|^2$$

Where the regularization parameter is C, which is greater than zero, the norm of w in the RKSH is the stabilizer and $\sum_{i=1}^{n} l(y_i, f(x_i))$is the empirical loss term.

The standardized functional can be minimized in standard SVMs, by resolving a convex quadratic programming optimization problem that assurances an exclusive global minimum result. In SVMs for binary classification (SVC) in which the target values $y_i \in \{-1, +1\}$, the hard margin loss function is defined as:

$$l_h(y_x, f_x) = \begin{cases} 0 & if \quad y_x \cdot f_x \geq +1; \\ +\infty & otherwise \end{cases}$$

For noise-free data sets, the best and appropriate function is the hard margin loss function. Whereas for other general cases, a soft margin loss function is commonly employed in classical SVC. This is equated as:

$$l_\rho(y_x, f_x) = \begin{cases} 0 & if \quad y_x \cdot f_x \geq +1; \\ \frac{1}{\rho}(1 - y_x \cdot f_x)^\rho & otherwise \end{cases}$$

Here, ρ is a positive integer.

The minimization of the regularized functional with the soft margin as loss function leads to a convex programming problem for any positive integer ρ; for L1 (ρ = 1) or L2 (ρ = 2) soft margin, it is also a convex quadratic programming problem. The L1 and L2 soft margin loss functions are simplified and handled as the le soft margin loss function, which is defined as:

$$l_e(y_x, f_x) = \begin{cases} 0 & if \quad y_x \cdot f_x \geq +1; \\ \frac{(1 - y_x \cdot f_x)^2}{4\varepsilon} & if \quad +1 \geq y_x \cdot f_x \geq +1 - 2\varepsilon; \\ (1 - y_x \cdot f_x) - \varepsilon & otherwise \end{cases}$$

where the parameter $\varepsilon > 0$. By introducing slack variables $\xi_i \geq 1 - y_i(< w \cdot \phi(x_i) > +b) \forall i$, the minimization hitch in SVC with the Le soft margin loss function can be modified as the following equivalent optimization problem. This is referred to as the primal problem:

$$\min_{w,b,\xi} R(w,b,\xi) = C \cdot \sum_{i=1}^{n} \psi_e(\xi_i) + \frac{1}{2} \| w \|^2$$

Subject to

$$\begin{cases} y_i(< w \cdot \phi(x_i) > +b) \geq 1 - \xi_i \\ \xi_i \geq 0, \ \forall i \end{cases}$$

Where

$$\psi_e(\xi) = \begin{cases} \frac{\xi^2}{4\varepsilon} & if \quad \xi \in [0, 2\varepsilon] \\ \xi - \varepsilon & if \quad \xi \in (2\varepsilon, +\infty) \end{cases}$$

Standard Lagrangian techniques are used to derive the dual problem. For the inequalities in the primal problem let us make an assumption that $\alpha_i \geq 0$. Also, let $\gamma_i \geq 0$ be the corresponding Lagrange multipliers. Now for the primal problem, the Lagrangian would be equated as:

$$L(w,b\xi) = C \cdot \sum_{i=1}^{n} \psi_e(\xi_i) + \frac{1}{2} \| w \|^2 - \sum_{i=1}^{n} \alpha_i (y_i \cdot (< w \cdot \phi(x_i) > +b) - 1 + \xi_i) - \sum_{i=1}^{n} \gamma_i \cdot \xi_i$$

The KKT conditions for the primal problem require:

$$w = \sum_{i=1}^{n} y_i \cdot \alpha_i \cdot \phi(x_i)$$

$$\sum_{i=1}^{n} y_i \cdot \alpha_i = 0$$

$$C \frac{\partial \psi_e(\xi_i)}{\partial \xi_i} = \alpha_i + \gamma_i \quad \forall i$$

On Lagrange multipliers, an equality constraint can be evidently equated as follows:

$$C \cdot \frac{\xi_i}{2\varepsilon} = \alpha_i + \gamma_i \ \ if \ \ 0 \leq \xi_i \leq 2\varepsilon \ \ and \ \ C = \alpha_i + \gamma_i \ \ if \ \ \xi_i > 2\varepsilon \ \ \forall i$$

If we collect all items involving ξ_i and let $T_i = C\psi_e(\xi_i) - (\alpha_i + \gamma_i)\xi_i$ .Fom the above formulas, we have:

$$T_i = \begin{cases} -\frac{\varepsilon}{C}(\alpha_i + \gamma_i)^2 & if \ \ \xi \in [0, 2\varepsilon] \\ -C\varepsilon & if \ \ \xi \in [2\varepsilon, +\infty] \end{cases}$$

Thus, the ξ_i can be eliminated if we set $T_i = -\frac{\varepsilon}{C}(\alpha_i + \gamma_i)^2$ and introduce the additional limitations. In expressions of the positive dual variables α_i and γ_i, the dual problem can be quantified as a maximization problem. This is represented as:

$$\max_{\alpha,\gamma} R(a,\gamma) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i y_i \alpha_j y_j < \phi(x_i)\cdot\phi(x_j) > +\sum_{i=1}^{n}\alpha_i - \frac{\varepsilon}{C}\sum_{i=1}^{n}(\alpha_i + \gamma_i)^2$$

Subject to:

$$\alpha_i \geq 0, \gamma_i \geq 0, 0 \leq \alpha_i + \gamma_i \leq C, \forall i \; and \; \sum_{i=1}^{n}\alpha_i y_i = 0$$

It is noted that $R(\alpha,\gamma) \leq R(\alpha,0)$ for any α and γ. Hence the maximization of the above formula over (α, γ) can be found as maximizing R (α,0) over $0 \leq \alpha_i + \gamma_i \leq C, \; and \; \sum_{i=1}^{n}\alpha_i y_i = 0$.

Therefore the dual problem can be finally simplified as:

$$\min_{\alpha} R(a) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i y_i \alpha_j y_j K(x_i, x_j) - \sum_{i=1}^{n}\alpha_i + \frac{\varepsilon}{C}\sum_{i=1}^{n}\alpha_i^{\,2}$$

subject to $0 \leq \alpha_i + \gamma_i \leq C, \forall i \; and \; \sum_{i=1}^{n}\alpha_i y_i = 0$. The classifier can be obtained from the solution of the above problem as $f(x) = \sum_{i=1}^{n}\alpha_i y_i K(x_i, x) + b$. In the solution, ‘b’ is achieved as a byproduct.

APPENDIX B. LOGISTIC REGRESSION (LR)

Logistic regression model can be express as the following:

$$\log it\ p(\Pi_1 \mid \mathbf{x}) = \log_e\left(\frac{p(\Pi_1 \mid \mathbf{x})}{1 - p(\Pi_1 \mid \mathbf{x})}\right) = \beta_0 + \beta^T \mathbf{x}$$

Similarly, we now write p ($\Pi_1 \mid \mathbf{x}$) as $p_1(\mathbf{x}, \beta_0, \beta)$ and p($\Pi_2 \mid \mathbf{x}$) as $p_2(\mathbf{x}, \beta_0, \beta)$. Then we can estimate β_0, β as the following method.

Given n training data, (x_i, y_i), the conditional likelihood for (β_0, β) can be written as:

$$L(\beta_0, \beta) = \prod_{i=1}^{n} (p_1(x_i, \beta_0, \beta))^{y_i} (1 - p_1(x_i, \beta_0, \beta))^{1-y_i}$$

hence, the conditional log-likelihood is:

$$l(\beta_0, \beta) = \sum_{i=1}^{n} \{y_i \log_e p_1(x_i, \beta_0, \beta) + (1 - y_i) \log_e (1 - p_1(x_i, \beta_0, \beta))\}$$

The maximum likelihood estimates (β_0, β) are obtained by maximizing $l(\beta_0, \beta)$ with respect to β_0 AND β .

There various algorithms for calculating the maximum likelihood estimates such as EM-method and weighted least-squares procedure are widely used. In this paper, its algorithms are not given.

After the maximum likelihood estimates ($\tilde{\beta}_0, \tilde{\beta}$) of (β_0, β) is obtained, the classification rule:

IF $\tilde{L}(x) > 0,\ assign\ \ x\ to\ \Pi_1$

Otherwise, $assign\ \ x\ to\ \Pi_2$

APPENDIX C. NEURAL NETWORK (NN)

The neural network mode is the following:

$assign\ \ x\,to\,\Pi_2\,assign\ \ x\,to\,\Pi_2$

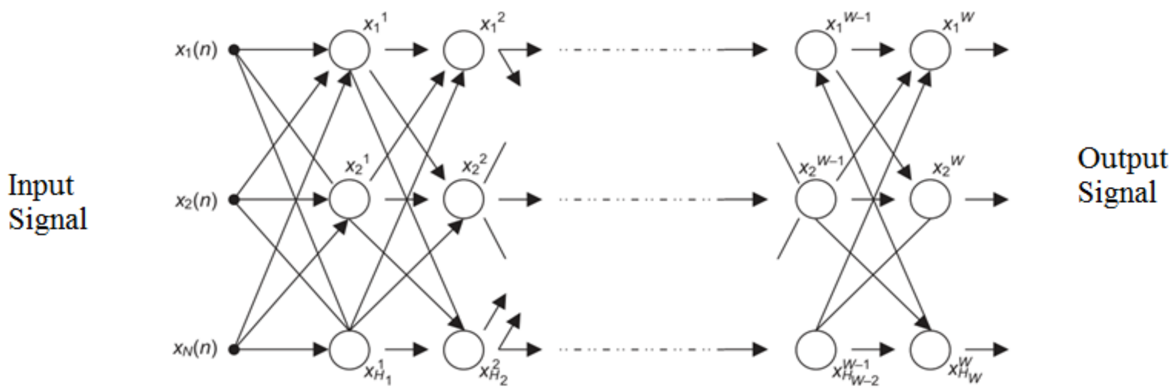

Figure C1: Graph of a multilayer neural network

As we can see in the Figure C1, the structure of the neural network consists of input, multilayers and output. In each layer, the output is calculated by the following formula:

$$\begin{cases} u_j^{H_{k-1}} = \sum_{i=1}^{H_{k-1}} w_{ji}^k z_i^{H_{k-1}} + h_j^{H_k} \\ z_j^{H_k} = f_j^{H_k}(u_j^{H_{k-1}}) \end{cases} \quad k = \overline{1,\ W},\ \ j = \overline{1,\ H_k}$$

where $z_j^{H_k}$ is the output of the j-th cell of the k-th layer, $h_j^{H_k}$ is the critical value of the j-th cell of the k-th layer, w_{ji}^k is the weight coefficient and $f_j^{H_k}$ is the output function of the j-th neural cell at the k-th layer.

Given the training data $(x(m), y(m))\ (x, y \in R^p, m = \overline{1, n})$, NN improve the weight coefficients, as the square sum of all errors of output is smaller. The formula of as the square sum of all errors of output is the following:

$$E(m) = \frac{1}{2}\sum_{j=1}^{p} | y_j(m) - z_j(m) |^2$$

The algorithm for training the given data is the following:

Step1: Set the initial weight (we can select any values for initial weights).

Step2: For k-th training data input the input signal and find all output signals.

Step3: By using the following formula, improve the weight coefficients.

First improve the weight coefficients of the last layer (W-th layer).

$$\begin{cases} e_j(m) = y_j(m) - z_j(m) \\ \delta_j^W(m) = (f_j^W(u_j^W(m)))' e_j(m) \\ \Delta w_{ji}^W(m) = \eta \delta_j^W(m) z_j^W(m) \\ \Delta h_j^W(m) = \eta \delta_j^W(m) \\ w_{ji}^W(m+1) = w_{ji}^W(m) + \Delta w_{ji}^W(m) \\ h_j^W(m+1) = h_j^W(m) + \Delta h_j^W(m) \end{cases}$$

Improve the weight coefficients of every layer in reverse order.

$$\begin{cases} \delta_j^W(m) = (f_j^W(u_j^W(m)))' \sum_{i=1}^{H_{L+1}} \delta_i^{L+1}(m) w_{ij}^{L+1}(m) \\ \Delta w_{ji}^L(m) = \eta \delta_j^L(m) z_j^L(m) \\ \Delta h_j^L(m) = \eta \delta_j^L(m) \\ w_{ji}^L(m+1) = w_{ji}^L(m) + \Delta w_{ji}^L(m) \\ h_j^L(m+1) = h_j^L(m) + \Delta h_j^L(m) \end{cases}$$

With improved weight coefficients, calculate the sum of the output error of the all-training data.

$$E = \sum_{m=1}^{n} E(m) = \frac{1}{2} \sum_{m=1}^{n} \sum_{j=1}^{p} | y_j(m) - z_j(m) |^2$$

Step 5: Until $E \le \varepsilon$, repeat Step 2- Step 4.