MACHINE LEARNING METHODOLOGIES IN BRAIN-COMPUTER INTERFACE SYSTEMS

A. E. Selim[1,2], M. Abdel Wahed[2], Y. M. Kadah[2,3]
[1] CTDC (Cairo Technology Development Center), IBM Egypt
[2]Systems and Biomedical Engineering Department, Cairo University, Egypt
[3]Systems and Biomedical Engineering Department, Nile University, Egypt
E-mail: abeers@eg.ibm.com

Abstract-**Brain-Computer Interfaces (BCI) is a one kind of communication system that enables control of devices or communication with others only through brain signal activities without using motor activities. The main application for BCI is to provide an alternative channel for helping disabled persons, hereafter mentioned as subjects, to communicate with the external world. This paper tries to demonstrate the performance of different machine learning algorithms based on classification accuracy. Performance has been evaluated on dataset II from BCI Competition III for the year 2004 for two subjects 'A' & 'B' and dataset IIb from BCI Competition II for the year 2003 for one subject 'C'. As a primary stage, a preprocessing was applied on the samples in order to extract the most significant features before introducing them to machine learning algorithms. The algorithms applied are Bayesian Linear Discriminant Analysis (BLDA), linear Support Vector Machine (SVM), Fisher Linear Discriminant Analysis (FLDA), Generalized Anderson's Task linear classifier (GAT), Linear Discriminant Analysis (LDA). BLDA and SVM yielded the highest accuracy for all 3 subjects. BLDA algorithm achieved classification accuracy 98%, 98% and 100%, SVM algorithm achieved 98%, 96% and 100% for subjects 'A', 'B' and 'C' respectively.**
***Keywords* - BCI, SVM, BLDA, Linear Classifiers**

I. INTRODUCTION

Brain-Computer Interface (BCI) operation is based on two adaptive controllers, the subject's brain, which produces the activity that encodes the subject's thoughts, intent or reflects the brain function, and the system, which decodes or translates this activity into control signals or device commands that control devices or computer applications.

Some people who suffer neurological diseases can be highly paralyzed and incapable of any motor functions but still have some cognitive abilities. Their only way to communicate with their environment is by using their brain activities. BCI research aims at developing systems that help disabled subjects. Moreover, as the interest in developing a new method of man-to-machine communication BCI research has grown steadily over the past few decades.

Many factors determine the performance of a BCI system. These factors include the brain signals measured, the signal processing methods that extract signal features, the algorithms that translate these features into device commands, the output devices that execute these commands, the feedback provided to the user, and the characteristics of the user.

BCI systems offer different paradigms to help disabled subjects to manipulate their brain activities and consequently different brain activity patterns can be obtained. Associated to BCI paradigms, there is the problem of classifying these patterns in order to be employed to translate the subject's intent into a control signal that controls devices or computer applications.

BCI systems can control a variety of devices such as wheelchair and an artificial limb or computer applications such as a specialized graphical user interface, simple word processing software or a computer application that is used as an environment control system.

Nearly all BCI systems contain as a core part a machine learning algorithm, which learns from training data and yields a function that can be used to discriminate different brain activity patterns. It adapts the BCI system to the brain of a particular subject. This decreases the learning load imposed on the subject. For simplicity and practical reasons, machine learning algorithms are usually divided into two modules: feature extraction and classification.

The feature extraction module serves to transform raw brain signals into a representation that makes classification easy. In other words, the goal of feature extraction is to remove noise and other unnecessary information from the input signals, while at the same time retaining information that is important to discriminate different classes of signals. Feature vectors are extracted from the brain signals by signal processing methods. Neurophysiologic a priori knowledge can aid to decide which brain signal feature is to be expected to hold the most discriminative information for the chosen paradigm. These features are translated into a control signal by machine learning algorithms.

BCI tools and techniques such as signal acquisition, signal processing, feature extraction, machine learning algorithms and classification techniques shares in the development and improvement of BCI technology.

Since few years now, several BCI competitions have been organized in order to promote the development of BCI and to evaluate the current state of the art of BCI system techniques and tools. Well versed laboratories in EEG-based BCI research provided data sets in a documented format. These data sets classified into labeled training sets and unlabeled test sets. The goal in the competition was to maximize the performance measure for the test labels. These competitions allow the community to benchmark several classification techniques. In this paper the data sets acquired using BCI2000's [1] P300 Speller Paradigm provided by BCI competitions II (2003) [2] & III (2004) organizers was used. Data sets has been recorded from two subjects 'A' & 'B' in competition III and from one subject 'C' in competition II. These dataset represents a complete record of P300 evoked related potentials (ERPs) [3].

P300 Speller is a BCI paradigm that helps disabled subjects to spell words by means of their brain signal activities. This paradigm based on the so-called oddball paradigm which states that rare expected stimuli produce a positive deflection (ERPs) in the electroencephalogram (EEG) signals after about 300 ms. The change occurs in the EEG signals called P300 component which is present in nearly every human.

Farwell and Donchin (1988) [4] were the first to employ the P300 as a control signal in BCI systems. Then much of the researches in the area of P300 based BCI systems have concentrated on developing new algorithms for the detection of the P300 from possibly noisy data [5-9].

In this paradigm the subject is presented with a 6 by 6 matrix of characters as shown in Fig. 1. The subject's task is to spell the word displayed on top of the matrix one character at time. For the spelling of a single character, each of the 12 rows and columns (6 rows and 6 columns) of this matrix were successively and randomly intensified. Row/column intensifications were block randomized in blocks of 12. In order to make the spelling procedure more reliable, the sets of 12 intensifications were repeated 15 times for each character sequence (i.e., any specific row/column was intensified 15 times and thus there were 180 total intensifications for each character sequences). The subject focuses on one out of 36 different characters of the matrix. Two out of 12 intensifications of rows or columns contained the desired character (i.e., one particular row and one particular column). Each row or column has a code from 1-12 as shown in Fig.2. P300 ERP appears in the EEG as a response to the intensification of a row or column containing the desired character. The EEG signals have been acquired using 64-channels. A more detailed description of the dataset can be found in the BCI competition online web site [10]. This P300 based paradigm can be considered as a "Virtual Keyboard on computer screen" BCI system.

The problem addressed in this paper is to predict if the post-intensification segments (i.e., the 64-channel signals collected after the intensification of a row or column, named a post-intensification segments) contains P300 ERP or not. This first part is a binary classification problem that applied 15 times corresponding to the number of sequences in each character spelling. The second part of the problem deals with a 36-class classification problem to recognize a symbol from a 6 by 6 matrix. Determining the target row and column can be used for the prediction of the character that the subject was focusing on. The goal is to correctly predict the desired character using the fewest sequences as possible.

Classification is a challenging problem due to the low signal-to-noise ratio of EEG signals, the variance of EEG signals for a given subject, and the variance between different subjects. The preprocessing and enhancement techniques used to enhance the EEG signals and remove the noise and extract the relevant features. Next section discusses the BCI tools and techniques tried in this work.

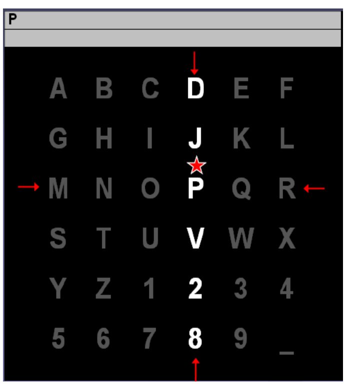

Fig. 1. P300 Speller matrix with a highlighted column.

Fig. 2. Encoding of the matrix rows & column.

II. Methodology

A. Data preprocessing and feature extraction

Since the significant part of the EEG signal is the part that occurs after intensification, and because some “high-level” features that can be fed to a classifier are needed to be built, the EEG signals have been preprocessed. The preprocessing operations were applied to the raw data provided by the competition organizers in the order stated below.

A.1. Single trials extraction

As mentioned before for a subject to spell a character there are a block of 12 post-intensification segments that are repeated 15 times therefore there are total of (12x15) 180 post-intensification segments for each character over 64-channels. The data provided by the competition organizers are not separated into post-intensification segments as discussed. They are provided as one signal. So the first step is to separate the provided signals in both data sets training and test into post-intensification segments. This is achieved by extracting the data samples between 0-650 ms posterior to the beginning of each intensification. According to the prior knowledge that P300 ERP appears about 300 ms after the stimulus this window is considered to be large enough to capture the required time features for an efficient classification.

A.2. Filtering

Filtering is a crucial step in noise reduction since certain types of artifact occur at known frequencies. The collected signals are bandpass filtered from 0.1-60 Hz and digitized at 240 Hz. After extracting data extra filtering was applied to the post-intensification segments using bandpass filter with cut-off frequencies 2-8, 0.1-10, 0.1-20, 0.1-30 & 0.1-40 Hz. These frequency ranges were chosen as the cognitive activity very rarely occurs outside of the range 3-40Hz [11].

A.3. Decimation

The filtered signals have been decimated according to the high cut-off frequency. At this point, an extracted filtered post-intensification segments from a single channel is composed of 6, 7, 14, 20, 27 measurements for signals filtered at high cut-off frequencies 8, 10, 20, 30, 40 Hz respectively.

A.4. Feature vector Construction

After this preprocessing stage, post-intensification segments have been transformed into a vector from the concatenation of the measurements of all 64 channels. For each character there are 180 post-intensification segments. For subjects 'A' & 'B' the training set of each subject consists of 85 characters therefore there are total 15300 (85x180) and the test set consists of 100 characters therefore there are 18000 (100x180). For subject 'C' the training set consists of 42 characters therefore there are total 7560 (42x180) post-

978-1-4244-2695-9/08/$25.00 ©2008 IEEE

intensification segments for training phase and the test set consist of 31 characters. Also from the provided data a vector of dimension 180 for each character that carries the code of the row or column corresponding to each post-intensification segment for both test and training data sets for each character was obtained. Also a vector of labels of the same dimension that carries either 1 or -1 for target and non-target post-intensification segments respectively for training set only was obtained. For test set the goal is to predict the label of each post-stimulus segment and consequently predict the correct character.

A.5. Normalization:

Prior to training, all feature vectors from a given training set obtained from the previous step have been normalized to zero mean and unit variance. Test set has also been transformed according to the resulting normalization parameter obtained from normalization of the training set.

B. Machine learning and classification:

Machine learning methods main role is to discriminate EEG patterns representing different types of brain activity. The performance of a machine learning system depends on both the features and the classification algorithm employed. Hence, in this work special emphasis is given to algorithms that learn from a set of training data how to discriminate EEG segments containing a P300 ERP from other EEG segments.

Classification was guided by two general approaches. First approach, follows the concept of "simple methods first" by employing only linear classifiers. As In BCI studies linear classification methods were never found to perform worse than non-linear classifiers [12]. Second approach, depends on extending the functionality of machine learning algorithms by regularization and also by combining multiple classifiers. Bayesian Linear Discriminant Analysis (BLDA) which is an extension of Fisher Linear Discriminant Analysis (FLDA) [13] and combination of multiple linear Support Vector Machines (SVM) classifiers [14] which also contains regularization parameter selection.

B.1. Linear Classifiers

There are pairs (x, y) of inputs $x \in X$ and desired outputs $y \in Y$. The problem which a learning algorithm has to solve is to choose, based on the training examples, a function f: $X \longrightarrow Y$ such that new examples, not contained in the training set, are correctly mapped to the corresponding output. For practical reasons the functions F is usually indexed by a set of parameters θ, i.e. $y= f(x; \theta)$. Hence, the task of choosing a function is equivalent to choosing parameters θ.

In the binary case $Y = \{1, -1\}$, the linear classifier is represented by a single discriminant function given by the parameter vector ω and bias b. This is the discriminant function

$$f(x) = (\omega \cdot x) + b. \qquad (1)$$

The input vector x is assigned to class $y \in \{1, -1\}$ as follows;

$$q(x) = \begin{cases} 1 \text{ if } f(x) = (\omega . x) + b >= 0 \\ -1 \text{ if } f(x) = (\omega . x) + b < 0. \end{cases} \qquad (2)$$

Different linear classifiers algorithms determine the parameters of the linear classifier vector ω and bias ***b*****.** Then these parameters obtained from the training or learning phase are used in the testing phase to predict each test example belongs to which class. The implemented linear classifiers algorithms can be distinguished according to its performance.

In this work the implemented algorithms are Fisher Linear Discriminant Analysis (FLDA), Linear Discriminant Analysis (LDA) and Generalized Anderson's Task (GAT). In this paper Generalized Anderson's Task linear classifier is introduced as one of the linear classifiers that can be used in BCI systems.

B.2. Linear Support Vector Machines (SVM)

SVM has been used in BCI researches since it is a powerful approach for pattern recognition especially for high-dimensional problems [14]. The signals provided are high-dimensional signals with low signal-to-noise ratio. There is also another problem which is signal responses varies due to spelling-unrelated EEG signal components or other brain activities within a single subject. There are two approaches to overcome these problems [15].

The first approach to cope with this problem is through multiple classifiers combination approach where each single classifier has its own training data set. The training signals in several partitions were clustered so that each partition has "similar" noisy components. For subjects 'A' and 'B' time chronology of spelled characters has been lost because organizers decided to scramble them. So the training data were clustered into different partitions each partition consists of the post-intensification segments of 5 consecutive characters of the training set. For subject 'C' the training data has the same sequence as the sequence of their acquisition they were not scrambled. So each word characters are used as a training partition the training data consists of 11 words and consequently 11 training partitions were obtained. A multiple classifier system for each subject was designed. Each single classifier of the system is a linear Support Vector Machine trained on one of the training partitions. Also each single SVM training involves the choice of the classical SVM hyperparameter (regularization parameter) C through cross-validation. The validation set was subset of the remaining training partitions. In fact the cross-validation showed that for values of C which were close to zero classification accuracy was optimal while lower classification accuracy was obtained for larger values of C. So the following values have been tried C = [0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1] and value that maximizes the score in each classifier was selected [14], [15]. The outputs of these classifiers are fused together in order to produce a single predicted character. Two procedures for output fusion were tried. Each classifier assigns a real-valued score to a post-intensification segment vector of the test set associated to a given row or column. It is considered that the most probable row and column is the one that maximizes the score. Averaging of the scores obtained from all classifiers was performed. In the first procedure the so-called "SVM method I" signals from each row or column were averaged over sequences. The idea of averaging SVM outputs over sequences has already been applied by Kaper et al. [6]. In the second procedure the so-called "SVM method II" averaging was done over sequences and also over the classification score. Averaging of SVM outputs over sequences and classification scores has been applied by Rakotomamonjy et al. [14].

This latter procedure leads to a more robust classification scheme since a classifier that assigns a bad score to a test data can be corrected by other classifiers. Moreover, classifier outputs averaging approach helps to reduce subject variance.

The second approach is to perform signal averaging which is a classical method for enhancing signal-to-noise ratio. So after preprocessing the post-intensification segments that have the same code (i.e. which belongs to the same row or column) were averaged which means that all 15 post-intensification segments for each row or column yields one post-intensification segment. For each character instead of having 180 post-intensification segments there are 12 segments. Then the training signals are now 1020 (12x85) these feature vectors were fed to linear SVM classifier. Concerning the hyperparameter C selection, in the first approach it was observed that almost the values of hyperparameter C selected by cross-validation were C = [0.1, 0.5, 0.05] so these values were tried here in this approach the so-called "Linear SVM".

B.3. Bayesian Linear Discriminant Analysis

Bayesian Linear Discriminant Analysis (BLDA) is an extension of Fisher Linear Discriminant Analysis (FLDA) [13], which is a simple but efficient method for machine learning. In particular, a framework from Bayesian machine learning, the so-called evidence framework. Bayesian version of FLDA outperforms plain FLDA in terms of classification accuracy. In this algorithm regression targets were computed from class labels of the training data sets in order to apply FLDA via regression. Then an iterative estimation of parameters procedure was applied. For the test phase to predict each test example associated to a certain row or column belongs to which class the mean of predictive distribution was computed for each test example which was considered as the score for the associated row or column. This algorithm has been applied by Hoffmann et al. [8].

C. Character prediction

All the previous machine learning algorithms yield scores for each row or column. Those columns and rows with high score are considered as the candidates [5]. The target character was determined by searching the column and row with the highest score for each of the 15 sequences.

III. RESULTS

This section presents the results achieved in this work. The classification performance yielded on the test data of the competition II and III is demonstrated. Test sets have been processed similarly to the training set and then are fed to the machine learning algorithms. For the competition, performances have been evaluated based on the percentage of correctly predicted characters in the test sets during the 5^{th} and the 15^{th} sequences. For subjects 'A' & 'B' (Competition III data sets) the 1^{st} ranked algorithm was an ensemble of SVM classifiers which moreover includes channel selection procedure. It achieved 73.5% and 96.5% for 5^{th} and 15^{th} sequences respectively. Based on the competition evaluation criteria, Table I depicts the performance results achieved in this work on the test sets with respect to the 5^{th} and 15^{th} sequences. The SVM method II achieved 75% and 97% as an average for both subjects 'A' & 'B' for 5^{th} and 15^{th} sequences respectively. The SVM method II algorithm is similar to the 1^{st} ranked one but it doesn't include channel selection procedure i.e. using all 64-channels. BLDA outperforms SVM method II as it achieved 75% and 98%. For subject 'C' results achieved in this paper are compared to those obtained by BCI competition II (2003) winner [8]. Table II includes the results obtained by all algorithms.

TABLE I
PERFORMANCES OF THE ALGORITHMS USED IN THIS WORK EVALUATED ON BCI COMPETITION DATA SET II RECORDED ON SUBJECTS 'A' AND 'B'

Algorithms	5^{th} Sequence			15^{th} Sequence		
	A	B	avg	A	B	avg
BLDA	69%	81%	75%	98%	98%	98%
SVM method II	70%	80%	75%	98%	96%	97%
SVM method I	55%	78%	66.5%	95%	96%	95.5%
GAT	49%	64%	55%	89%	91%	90%
Linear SVM	53%	66%	59.5%	89%	87%	88%
FLDA	49%	67%	58%	84%	90%	87%
LDA	29%	69%	49%	75%	91%	83%

TABLE II
NUMBER OF MISSPELLINGS OF CHARACTERS IN THE TEST WORDS WITH RESPECT TO THE NUMBER OF SEQUENCES AND THE ALGORITHM EVALUATED ON BCI COMPETITION II DATA SET IIb RECORDED ON SUBJECT 'C

Algorithms	Number of sequences										
	1	2	3	4	5	6	7	8	9	10	15
BLDA	5	2	0	0	0	0	0	0	0	0	0
SVM method II	7	4	4	0	0	0	0	0	0	0	0
SVM method I	11	5	4	1	0	0	0	0	0	0	0
LDA	17	15	9	4	3	3	0	0	0	0	0
FLDA	22	10	7	4	4	4	2	0	0	0	0
GAT	20	16	11	6	7	4	4	2	1	0	0
Linear SVM	17	16	11	10	6	8	6	2	3	3	1

The winner reached to zero misspelling of characters in the test words starting from the 5^{th} sequence. In this work zero misspelling of characters was achieved from the 3^{rd} sequence, the 4^{th} sequence and the 5^{th} sequence with BLDA, SVM method II and SVM method I algorithms respectively.

IV. DISCUSSION

Concerning feature extraction module, several preprocessing operations were applied to the data including filtering. Data filtered within different frequency ranges were tested with all previously mentioned classification algorithms. From the results obtained it was observed that almost all classification algorithms yield the best results at certain frequency range which differ from subject to another.

978-1-4244-2695-9/08/$25.00 ©2008 IEEE

It is found that when data filtered within 0.1-30, 0.1-10 and 0.1-20 Hz for subjects 'A', 'B' and 'C' respectively they yield the best results with almost all classification algorithms. For most of the previous trials done in BCI the same filtering frequency range was applied to all subjects. The results achieved in this work outperforms the results of the competition winners although the same machine learning algorithms have been used due to the selection of the appropriate frequency ranges at which data to be filtered for each subject separately.

Concerning the classification module, almost all linear classifiers yield results on the same range approximately 80-90%. Generalized Anderson's Task (GAT) linear classifier is introduced in this paper to BCI classification algorithms and it outperforms all other linear classifiers evaluated here with most of the subjects.

For SVM three procedures were tried to address the variance between subjects. The first is combination of multiple linear SVM classifiers done by single averaging over sequences only which is SVM method I. The second is combination of multiple linear SVM classifiers done by double averaging over sequences and classifiers scores which is SVM method II. The third post-intensification segments of the same row or column were averaged, then they were fed to linear SVM classifier which is linear SVM. SVM method I & II which represent the combination of linear SVM classifiers approach outperform the linear SVM which represent the classical approach of signal averaging. While for SVM method I & II as expected double averaging approach (SVM method II) outperforms the single averaging approach (SVM method I).

For the regularization parameter (hyperparameter) C it was observed that the values 0.1, 0.5 & 0.05 were usually selected during model selection. It was also observed that for linear SVM approach the results obtained from the data filtered using the appropriate frequency range were not affected by changing the value of hyperparameter C.

Finally, Bayesian Linear Discriminant Analysis outperforms all machine learning algorithms evaluated in this work. BLDA uses regularization to prevent over-fitting to high-dimensional and possibly noisy data sets. Through Bayesian analysis the degree of regularization can be estimated quickly, robustly and automatically from training data without the need for time consuming cross-validation and model selection procedures.

V. Conclusion

This paper presented methodologies for classifying event related potential which is P300 for a P300-based BCI system for disabled subjects. Different ways were demonstrated to address the variance in EEG signals of one subject and variance between different subjects. And also to over come the low signal-to-noise ratio of EEG signals problem.

Machine learning algorithms are one of the main themes of this work. For feature extraction phase it was found that selecting the appropriate frequency rage for each subject can help in addressing this problem and yields better performance. For classification phase the most important result reached is that the extension of functionality of linear classifier algorithms outperforms linear classifiers. For that reason BLDA which is an extension to FLDA and combination of linear SVM classifiers which is an extension to linear SVM both achieve the best results. Future improvement to the work could be by trying to achieve higher recognition rate with fewer sequences.

References

[1] G. Schalk, D. McFarland, T. Hinterberger, N. Birbaumer, and J. Wolpaw, "BCI2000: a general-purpose brain-computer interface (BCI) system," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1034–1043, June 2004.

[2] B. Blankertz, K.-R. Mueller, G. Curio, T. Vaughan, G. Schalk, J. Wolpaw, A. Schloegl, C. Neuper, G. Pfurtscheller, T. Hinterberger, M. Schroeder, and N. Birbaumer, "The BCI competition 2003: Progress and perspectives in detection and discrimination of EEG single trials," *IEEE Transactions Biomedical Engineering*, vol. 51, no. 6, pp. 1044–1051, June 2004.

[3] B. Blankertz, K.-R. Müeller, D. Krusienski, G. Schalk, J. Wolpaw, A. Schlögl, G. Pfurtscheller, J. del R. Millan, M. Schröder, and N. Birbaumer, "The BCI competition iii: Validating alternative approaches to actual BCI problems," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 14, no. 2, pp. 153–159, June 2006.

[4] L. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalography and Clinical Neurophysiology*, vol. 70, no. 6, pp. 510–523, 1988.

[5] N. Xu, X. Gao, B. Hong, X. Miao, S. Gao, and F. Yang," BCI Competition 2003—Data Set IIb: Enhancing P300 Wave Detection Using ICA-Based Subspace Projections for BCI Applications," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1067 – 1072, June 2004.

[6] M. Kaper, P. Meinicke, U. Grossekathoefer, T. Lingner, and H. Ritter, "BCI competition 2003 - dataset IIb: Support vector machines for the P300 speller paradigm," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1073– 1076, June 2004.

[7] A. Rakotomamonjy, V. Guigue, G. Mallet, and V. Al-varado, "Ensemble of SVMs for improving brain-computer interface P300 speller performances.," International Conference on Artificial Neural Networks (ICANN)., Springer Berlin / Heidelberg, vol. 3696 pp. 45-50, November 2005.

[8] U. Hohffmann, G. Garcia, J.-M. Vesin, and T. Ebrahimi, "Application of the evidence framework to brain-computer interfaces.," *IEEE Engineering in Medicine and Biology Conference*., vol. 1, pp. 446 – 449, September 2004.

[9] M. Thulasidas, C. Guan, and J. Wu, "Robust classification of EEG signal for brain-computer interface," *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 1 pp. 24–29, March 2006.

[10] B. Blankertz, BCI Competition III Webpage. [Online]. Available: http://ida.first.fraunhofer.de/projects/bci/competition_iii

[11] B. Allison,"P3 or not P3: Toward a Better P300 BCI", ph.D. dissertation, University of California, San Diego 2003.

[12] F. Vojtěch, H. Václáv "Statistical Pattern Recognition Toolbox for Matlab User's guide," June 24, 2004.

[13] U. Hoffmann, J-M Vesin, T. Ebrahimi, and K. Diserens, " An efficient P300-based brain-computer interface for disabled subjects ", *Journal of Neuroscience Methods,* vol. 167, no. 1, pp. 115-125, 2008.

[14] A. Rakotomamonjy, and V. Guigue, " BCI Competition III: Dataset II- Ensemble of SVMs for BCI P300 Speller," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 3, pp. 1147 – 1154, March 2008.

[15]A. Rakotomamonjy, "Variable Selection Using SVM-based Criteria", *Journal of Machine Learning Research*, vol. 3, pp. 1357–1370, 2003.

978-1-4244-2695-9/08/$25.00 ©2008 IEEE