An implementation of SSVEP-BCI system based on a cluster measure for feature selection

Luisa Suarez, Eleri Cardozo
and Romis Attux
FEEC, UNICAMP. Campinas, Brazil
Email: lsuarez@dca.fee.unicamp.br,
eleri@dca.fee.unicamp.br, attux@dca.fee.unicamp.br

Diogo Soriano
Center of Engineering, Modeling and Applied Social Sciences
ABC Federal University, Sao Paulo, Brazil
Email: diogo.soriano@gmail.com

***Abstract*—The main objective of a brain-computer interface (BCI) is to create alternative communication channels between the brain and a machine using information from cerebral responses. Among the possible paradigms to design a BCI system, this work focuses on Steady-State Visually Evoked Potentials (SSVEP). SSVEP are brain responses synchronized with fast repetitive external visual stimuli. The SSVEP-BCI system is able to meet many of the requirements of a strict BCI, but still needs to reduce the influence of noise on the Electroencephalogram (EEG) signal in order to improve its performance. In this paper, a novel SSVEP-BCI system is presented and analyzed in detail. The system is based on three pillars: spectrum estimation, systematic feature selection - for which different heuristics were proposed here -, and linear classification.**

Keywords*—BCI - Brain-Computer Interface, EEG - Electroencephalogram, SSVEP - Steady-State Visual Evoked Potential, Power Spectral Density, Feature selection, Davies Bouldin index, Linear Discriminant Analysis

I. INTRODUCTION

Brain-computer interface (BCI) systems play a very important role in the context of modern assistive technology [1]. The reason is that this kind of system allows the control of devices by commands generated according to the recognition of certain patterns of brain activity [2], as illustrated in Figure 1.

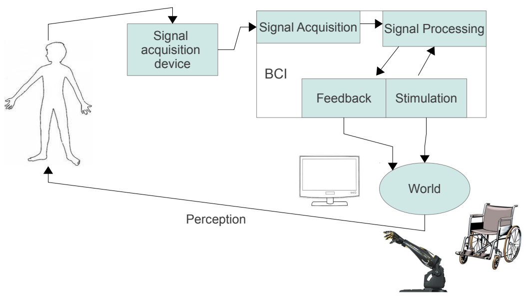


Fig. 1. Overview of a BCI system.

Distinct paradigms can be followed in the design of a BCI system such as motor imagery, P300, and steady state visually evoked potentials (SSVEP) [3]. Among these options, the use of SSVEP is interesting from the standpoint of operational straightforwardness and robustness if the target users have no limitation regarding eye movement.

In particular, the SSVEP-BCI system consists of different light sources with distinct flashing frequencies, with each frequency encoding a control command. The SSVEP response is classically observed in the occipital region of the brain as an oscillating wave induced by repetitive visual stimulation with frequencies higher than 6 Hz. In general, stimuli at a given frequency lead to an electrical oscillation in the visual cortex at the same frequency with its harmonics and sub-harmonics [4]. This identification can be performed by a frequency analysis. After this stage, a feature selection followed by a classification structure is used in order to better recognize the brain activity, and, afterwards, the generation of commands for the external devices takes place [5].

The result of the spectral analysis is a number of features (coefficients in the frequency domain), which are passed forward to the classification stage. Although it may seem prima facie that the use of as many features as possible is a sound strategy, this is not the case: there can be useless and even noxious features from a classification perspective. This explains the relevance of adopting systematic feature selection methods.

In the present work, a complete SSVEP-BCI system is implemented, with a special focus in a technique devoted for selecting features before classifying it. To accomplish this task, a cluster measure defined by the Davies Bouldin (DB) index [6] was used in order to choose the attributes that better separate the classes (maximizing the distance of the classes centroids), and, at the same time, minimize the dispersion of the classes. The results obtained illustrate that such feature selection technique helps to effectively improve the attained classification performance in a five command SSVEP-BCI system. In this case, different heuristics concerning the attributes ranked by DB measure are presented and the convergence towards a minimal classification error is evaluated as the number of attributes increases. In particular, it is shown that despite of the variability of the features selected by the strategies adopted here, the information obtained from the analysis is sufficient for reaching classification results with 90% of accuracy for each class, providing a clear picture of how the classification error decreases when information of the attributes (selected according different heuristics) were added. In this work, all the processing framework was carried in a offline scenario, being the real time application a natural future perspective.

II. Modeling

The SSVEP response was modeled as an oscillation $Y_i(t)$ defined as the voltage between the i-th electrode and a reference electrode at time t. $Y_i(t)$ can be described as a sinusoidal function of stimulus frequency f_i and phase $\theta_{i,k}$, and their k harmonics with coefficients $a_{i,k}$, as shown in the first term of equation 1. The second term corresponds to a set of disturbing signals, for example, brain processes or other external disturbances. These disturbing signals $(z_j(t))$ are added to each electrode signal with weighting $(b_{i,j})$ factors. Finally, there is an interfering component given by the signal $e_i(t)$. N_h is the number of harmonics considered.

$$y_i(t)=\sum_{k=1}^{N_h} a_{i,k}\sin(2\pi k f_i+\theta_{i,k})+\sum_j b_{i,j} z_j(t)+e_i(t) \quad (1)$$

III. Methods

A. Preprocessing

The preprocessing stage consisted in the use of analog notch filters with cutoff frequency at 60 Hz to remove noise from the power grid and a band-pass filter from 0.1 Hz to 50 Hz. Afterwards, the signals from each electrode were normalized by the respectives peaks in order to mitigate the scale discrepancies. The mean of each electrode was also subtracted and the normalization to the unitary variance were also employed as described in [7].

B. Feature extraction

The spectral content of a SSVEP response was determined by estimating the power spectral density (PSD). This estimation step can be carried out according to a variety of methods, like the classic periodogram, the Bartlett method and the Welch method [8]. The last of these methods was chosen due to its desirable properties insofar as estimator variance is concerned.

Briefly, the PSD coefficients were evaluated here considering 6s of time series sampled at 128 Hz, which implies in segments of 768 points. These segments were divided in 8 overlapping sub-blocks with 50% of intersection. The square of the absolute value of the discrete Fourier transform (DFT) was calculated and the average of all sub-blocks was taken. The DFT calculation employed 128 points, providing a spectral resolution of 1 Hz.

C. Feature selection

After extracting the features by means of spectral analysis, the Davies Bouldin index was used to rank the attributes capabilities in discriminating the classes. This measure combines in a single expression two main relevant aspects of data clustering: the minimization of the intraclass distance and the maximization of the distance between the classes w_i with $i = 1, 2, ..., m$, which can be described by:

$$DB_m = \frac{1}{m}\sum_{i=1}^{m} \max_{j=1,\dots,m \ j\neq i} \frac{s_i + s_j}{d_{ij}} \quad (2)$$

in which s_i is a measure of dispersion around the mean of class i and d_{ij} is the distance between the centroids of classes i and j. Therefore, small values of DB are indicative of the presence of compact and well-separated groups, being, the inverse of this index (DBinv) used here to rank the features. Thus, the feature selection heuristics will seek by combination of attributes with high DBinv measures. A comprehensive and detailed introduction of DB index can be found in reference [7].

For the heuristics introduced here, the DBinv index was firstly used in order to seek channels (electrodes) with improved response at stimulation frequencies. Later it was also used in the search of the best features scattered in all channels and for all values of the frequency spectrum of interest (8-30 Hz). It is important to note that analyzing each attribute independently is far from ideal, but is simpler in comparison with other more elaborate techniques, avoiding, for example, exhaustive searches or even multi-objective optimization problems.

Three strategies have been proposed in order to define the attributes that would compose the attribute vector: 1) the selected attributes were restricted to the points of the spectrum at frequencies of visual stimulation taking the coefficients without using any ranking for the attributes or electrodes. In this case, if we take the four coefficients of each of the 16 electrodes it would perform an amount of 64 descriptors for the feature vector; 2) the classifier is fed with vectors of four descriptors of the electrodes ranked by DBinv index. The electrodes are ranked according to the sum of its respectives PSD coefficients. For an amount of 16 electrodes, a total of 64 descriptors is also attained; 3) a rank list based on DBinv index of all possible attributes considering PSD coefficients in the range 8 and 30 Hz for 16 electrodes was created and the attributes were progressively added to the feature vector until reach 64 descriptors.

D. Classification

Finally, the classification stage was based on linear discriminant analysis (LDA) [9] since it is easy to implement, fast to train and widespread in the BCI literature [10] [3]. In this approach, a linear combination $\mathbf{w}$ of the features $\mathbf{x}$ that better separates the classes is found, providing a decision surface in the form: $\mathbf{w}^T\mathbf{x} + c = 0$, for a threshold c. Considering two normal multivariate distributed classes with means $\boldsymbol{\mu_1}$ and $\boldsymbol{\mu_2}$ and covariance matrices C_1 and C_2, respectively, the LDA approach consists in finding the weights $\mathbf{w}$ that maximize the ratio concerning the variance between the classes and the variance within the classes:

$$S = \frac{\sigma^2_{between}}{\sigma^2_{within}} = \frac{(\mathbf{w}^T(\boldsymbol{\mu_1} - \boldsymbol{\mu_2}))^2}{\mathbf{w}^T(C_1 + C_2)\mathbf{w}} \quad (3)$$

It is possible to show that this criteria is satisfied for $\mathbf{w} \propto (C_1 + C_2)^{-1}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2})$ and the threshold c is given by $\frac{1}{2}\mathbf{w}^T(\boldsymbol{\mu_1} + \boldsymbol{\mu_2})$. Once the training stage has been completed, trials with attribute vectors $\mathbf{x}$ were classified according to their position in the attribute space relative to the achieved decision hyperplane. The multi-class case was treated here analyzing all pair of classes.

A cross-validation process to evaluate the performance of classification was performed using the technique of k-fold with k = 200 different partitions, separating 250 trials for training and 100 trials for testing the classifier for each subject.

IV. Materials

Three healthy young subjects (lab colleagues, one woman) with no previous history of neurological diseases participated in this study. All subjects signed an informed consent previous to data acquisition. The reference to each subject is done by naming the first letter of their names, with P, J and L.

In order to acquire, digitize and amplify the EEG signals, the g.tec's g.USBamp equipment was used. The EEG was recorded using a cap of dry electrodes. The electrode system is known as g.SAHARA, and consists of 16 8-pin electrodes made of a special alloy of gold. These electrodes were placed at 16 sites defined by the international 10-20 system, some of these positions were chosen due to their good SSVEP responses [11]. They are: Fz, Cz, Pz, Oz, PO3, PO4, O1, O2, P3, P4, Iz, POZ, PO7, PO8, O9, O10. Two reference electrodes were placed on each mastoid. For visual stimulation, a set of four light-emitting diodes (white LEDs) was used. These LEDs were driven at frequencies 13, 18, 21 and 25 Hz. Studies have shown that stronger SSVEP responses are evoked in this range [8]. The subjects were placed on a comfortable chair at a distance of 0.5 m from the LEDs and the computer screen. An ambient light level was maintained during the execution of the experiments.

The software used for signal acquisition was BCI2000. This software allowed the creation of experiments with particular settings for captured data. The captured data is stored in files in the BCI2000 [1] own format and were subsequently imported into MATLAB were the signal processing was conducted.

The data acquisition sessions were performed at two different dates. In each acquisition date, four different sessions were held with intervals of five minutes. The first three sessions comprised four runs separated by a one-minute pause; the last session had only three runs. Each run is composed of 25 trials associated to 5 different tasks (4 visual flicking stimuli and a rest state).

During a typical experiment run, the subject is asked to perform a task to focus on a visual stimulus or to stay relaxed according to images displayed on a computer screen. During each run, the subject was instructed to remain as motionless as possible, thus minimizing the generation of artifacts. Each image was displayed continuously for 6s; afterwards, the screen was blank for 1.5s before the next image appeared. In resting condition, the subject could rest his/her eyes, and, in the intervals, he/she had time to move around a little.

V. Results

The analysis of the signal of 16 channel EEG was performed to find spectral patterns related to each subject. Figures 2 to 4 show the estimated PSD for L, P and J subjects for some of the channels considered. A comparison between the frequency spectrum in resting state vs. the response when the subject was stimulated is observed in these figures. The difference between the spectrums of each subject is remarkable. In addition to that, it can also be realized that the response state is usually close to the rest state for almost all scenarios. This probably follows from the fact of the rest state required attention only to the center of the computer screen, a situation that can be affect by the interferences of all visual stimulus in the peripheral vision. However, note that certain frequencies in specific channels could be used to distinguish the response and rest state, as, for instance, 25 Hz in electrode 7 for subject J. Despite of that, a better pattern characterization is only performed using the information of several PSD coefficients and from different electrodes, which justifies the use of the feature selection stage as performed here.

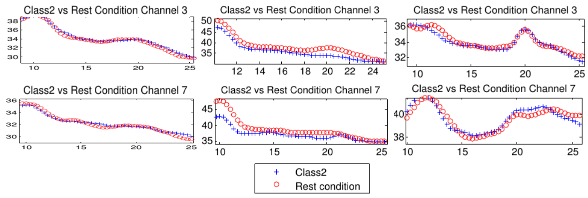


Fig. 2. Power spectral density coefficients for subject J (column 1), L (column 2) and P (column 3) to channels Pz and O1. X Axis (Power, V^2Hz^{-1}). Y axis (Frequency, Hz). Class2 (21 Hz).

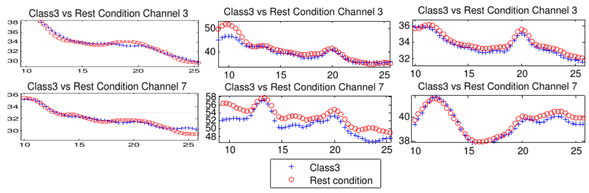


Fig. 3. Power spectral density coefficients for subject J (column 1), L (column 2) and P (column 3) to channels Pz and O1. X Axis (Power, V^2Hz^{-1}). Y axis (Frequency, Hz). Class3 (25 Hz).

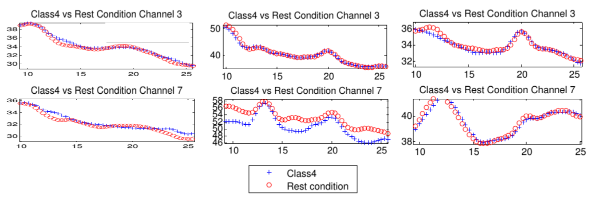
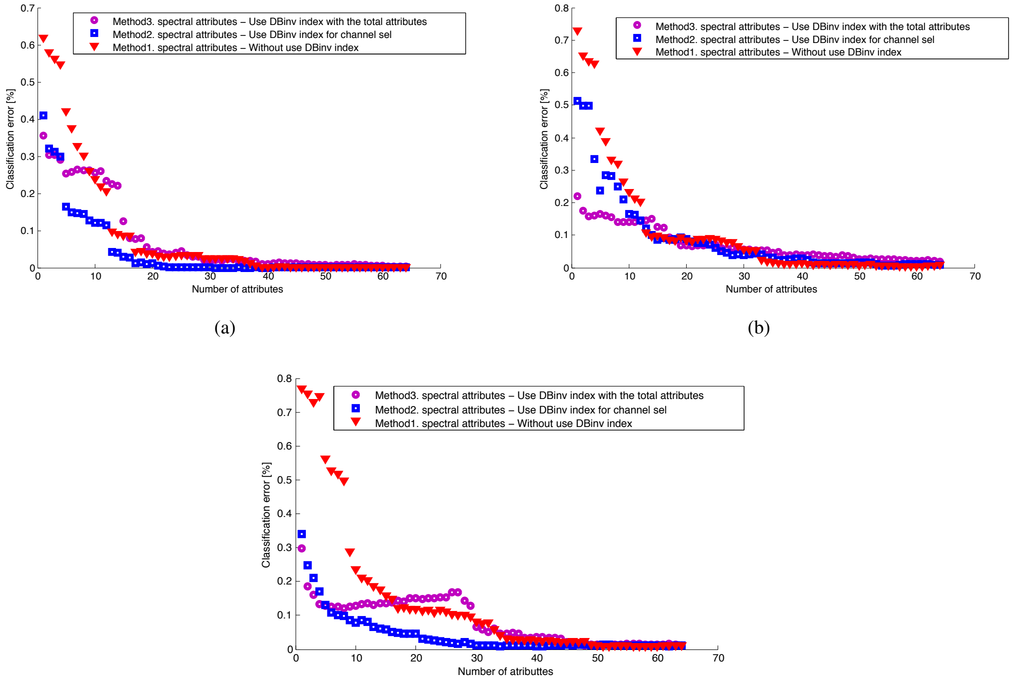


Fig. 4. Power spectral density coefficients for subject J (column 1), L (column 2) and P (column 3) to channels Pz and O1. X Axis (Power, V^2Hz^{-1}). Y axis (Frequency, Hz). Class4 (13 Hz).

In order to illustrate the different performances attained for the feature selection strategies, Figure 5 shows the evolution of the mean classification error of each strategy for each subject under progressive increase of the number of attributes. In this case, it can be observed that all the subjects reached an overall successful classification rate of 90% using between 10 and 20 attributes. Each subject had a different strategy as the most successful, and, in general, the use of the DBinv index for feature selection resulted in misclassification results that fell faster in comparison with strategies that did not use any technique for ranking the attributes, i.e. just took the coefficients in the visual stimulation frequencies (see the blue and purple dots vs. red dots on the graph).

Table I shows the results of comparing the number of attributes necessary for each one of the techniques used for

[1] BCI2000 is a general purpose system for research with brain-computer interface (BCI). Can be used for data acquisition, stimulus presentation, and brain monitoring applications. http://www.bci2000.org/BCI2000/Home.html



(a) (b)

(c)

Fig. 5. Error performance for general classification to different strategies of feature selection for subjects a) J, b) L e c) P

each subject to reach an error rate of 10%. The differences in the number of attributes, in some cases, are not particularly significant, but it is clear that the use of feature selection based on the DBinv index leads to the best results (Method 2). In particular, this method seems to point to a exponential decrease to the mean classification error when information of new attributes were added (see, for instance, subject P in Fig. 5).

TABLE I. COMPARISON OF THE NUMBER THE ATTRIBUTES REQUIRED FOR EACH OF THE PARTICIPATING SUBJECTS FOR A MISCLASSIFICATION OF 10%.

	Method		
Subjects	1	2	3
J	13	12	16
L	13	13	17
P	30	10	30

Table II presents the correct classification results for each class separately. A comparison between the three methods used for the attribute selection when feature vector has complete 16 features was performed for each of the subjects. The variability in the performance of classification for each particular class was exposed in these results. E.g. in the case of the results for method 1 is observed that for class 1, the subject J achieved a classification performance of 91.6%, while the P subject to this same class achieved 85.3%. In the case of method 2 about 7% improvement was obtained on the results compared to method 1. This positive increase shows the effectiveness obtained by using methods based on DB index for attribute selection.

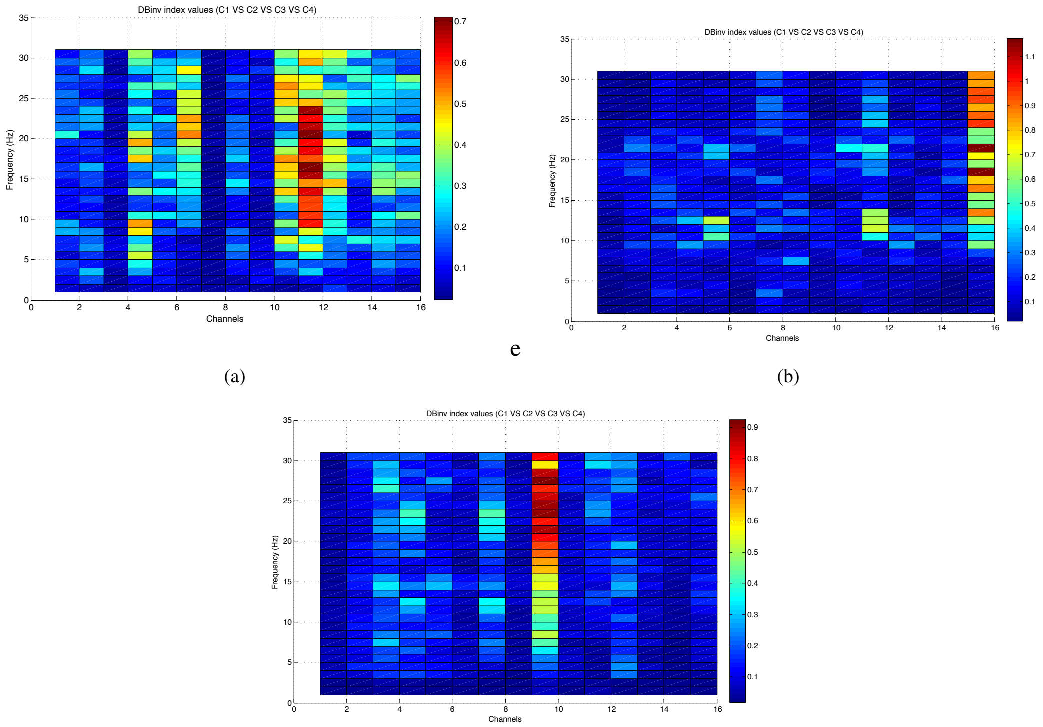
Interestingly, as Figs. 2, 3 and 4 have also revealed, the SSVEP response was more strong at the frequencies of visual stimulation in some channels, particularly those in the occipital region. This was subsequently confirmed by making the analysis of the attributes with the DBinv index. The map of DBinv coefficients is presented in Figure 6 for each subject, in which values of the DBinv index are shown so that the points of the map (electrode x frequency) with the highest values (red) indicate a coefficient with good characteristics in terms of class separation. The values shown on the map were obtained considering 32 coefficients of the frequency spectrum of the 16 electrodes. Discrimination between the 5 classes (stimulation frequencies and rest) was evaluated.

It is possible to notice on the maps that, for each different subject, there was a significant discriminant activity for different electrodes. Stronger discrimination (in terms of the measure used here) was defined by channel 11 (Iz) for subject J, by channel 16 (O10) for subject L, and by channel 9 (P3) for subject P. Furthermore, we observe that better discrimination was attained around the frequencies of stimulation, but also in other parts of the spectrum.

VI. DISCUSSION

According to the results presented in context of the SSVEP-BCI system designed here, we have obtained a high classification performance concerning classical preprocessing, feature extraction and classification techniques, which were also supported by different feature selection heuristics proposed in this work. In this case, all methods were chosen due to their efficiency and low computational.

In particular, the feature selection heuristics that used



e

(a) (b)

(c)

Fig. 6. Map of DB coefficients for the subjects. a) J, b) L e c) P

TABLE II. DETECTION ACCURACY (IN %) ACROSS 3 SUBJECTS FOR THE THREE METHODS USED WITH A 16 VECTOR ATTRIBUTE.

	Method 1				Method 2				Method 3			
	Class				Class				Class			
Subject	1 (13Hz)	2 (18Hz)	3 (21Hz)	4 (25Hz)	1 (13Hz)	2 (18Hz)	3 (21Hz)	4 (25Hz)	1 (13Hz)	2 (18Hz)	3 (21Hz)	4 (25Hz)
J	91.6	99.9	90.9	84.7	98.4	100	98.9	92.1	99.6	86	91.1	92
L	90.8	100	100	86.4	93.8	100	80.9	84.9	94.6	99.8	78.7	87.8
P	85.3	91.9	98.9	71.7	87.7	99.9	99.9	95.6	92.2	99.8	93.6	78.2
mean	89.3	97.26	96.6	80.93	93.9	99.9	93.23	90.83	95.46	95.2	87.8	86

some ranking strategy (based on a cluster measure) for the attributes also helped to precisely locate descriptors with higher discrimination performance and defined a more efficient approach, which was attested by the progressively decrease in the classification error observed in the experiments for all subjects. It is important to note that feature selection is usually performed in BCIs systems using the r2 index to assess the separability of attributes of two classes of data [12], being the DBinv index a more practical approach, since it allowed evaluating the discrimination capability for all the five classes simultaneously without using any assumption about the probability density function underlying the data as required by other feature selection techniques as the t-test and the ROC curve [7]. The comparison between the performance obtained here using the DBinv measure and these alternatives techniques outlines a natural perspective for future work.

VII. CONCLUSIONS

As the main conclusions of this work, we can mention: a) the frequency analysis performed here revealed that there is a great variability between the most promise PSD coefficients and electrodes for separating different classes for different subjects; b) usually the visual stimulation frequencies define promising features for discrimination, but other attributes can also carry relevant information and be preferred for a fast convergence towards a minimal classification error; c) ranking the electrodes before choosing the best attributes for each was, in general, a well-succeed approach; d) despite the variability of the features selected by the strategies adopted here, the information obtained from the analysis were sufficient for reaching more than 90% accuracy for each class.

Finally, by the results obtained in this work, it is possible to mention that the set of signal processing techniques used here defines an interesting paradigm for real time applications due to their simplicity and relatively low computational cost, being, therefore, also a natural perspective for a future work.

ACKNOWLEDGMENT

This research was supported by Finep and Capes, two Brazilian research funding agency.

REFERENCES

[1] C. Guger, B. Allison, and G. Edlinger, "State of the art in bci research: Bci award 2011," in *Brain-Computer Interface Research*, ser. SpringerBriefs in Electrical and Computer Engineering, C. Guger, B. Z. Allison, and G. Edlinger, Eds. Springer Berlin Heidelberg, 2013, pp. 1–5. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-36083-1_1

[2] J. d. R. Millán, R. Rupp, G. Müller-Putz, R. Murray-Smith, C. Giugliemma, M. Tangermann *et al.*, "Combining Brain-Computer Interfaces and Assistive Technologies: State-of-the-Art and Challenges." *Frontiers in neuroscience*, vol. 4, pp. 1–15, September 2010. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2944670\&tool=pmcentrez\&rendertype=abstract

[3] A. Bashashati, M. Fatourechi, R. K. Ward, and G. E. Birch, "A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals." *Journal of neural engineering*, vol. 4, no. 2, pp. R32–57, Jun. 2007. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/17409474

[4] G. Bin, X. Gao, and Y. Wang, "VEP-based brain-computer interfaces: time, frequency, and code modulations," *Computational Intelligence Magazine, IEEE*, vol. 4 , 4, pp. 22–26, November 2009. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5294934

[5] S. M. T. MLLER, B. Salezze, T. F. BASTOS FILHO, and M. SARCINELLI FILHO, "Estimativa de picos espectrais para composio de vetor de caractersticas de uma interface crebro-computador," in *XVIII Congresso Brasileiro de Automtica - CBA2010*, vol. nico, 2010, pp. 3794–3799.

[6] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. PAMI-1, no. 2, pp. 224–227, 1979.

[7] S. Theodoridis and K. Koutroumbas, *Pattern Recognition, Fourth Edition*, 4th ed. Academic Press, 2008.

[8] K. B. Ng, A. P. Bradley, and R. Cunnington, "Stimulus specificity of a steady-state visual-evoked potential-based brain-computer interface." *Journal of neural engineering*, vol. 9, no. 3, p. 036008, Jun. 2012. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/22589242

[9] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.

[10] G. E. Fabiani, D. J. McFarland, J. R. Wolpaw, and G. Pfurtscheller, "Conversion of EEG activity into cursor movement by a brain-computer interface (BCI)." *IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, vol. 12, no. 3, pp. 331–8, Sep. 2004. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/15473195

[11] C.-h. Wu and H. Lakany, "Impact of Stimulus Configuration on Steady State Visual Evoked Potentials (SSVEP) Response," in *COGNITIVE 2012, The Fourth International Conference on Advanced Cognitive Technologies and Applications*, no. 3, Nice, France, 2012, pp. 77–82. [Online]. Available: http://www.thinkmind.org/index.php?view=article\&articleid=cognitive_2012_4_20_40040

[12] J. Mnguez Zafra, L. Montesano del Campo, and E. M. Horna Prat, "Decodificacin de los objetivos finales (3d) del movimiento del brazo en tareas de alcance a partir de potenciales de movimiento anticipatorio para eeg brain computer interface," 2011.